

University of Houston – Fall 2017 – Dr. G. Heier
Advanced Linear Algebra I (Math 4377/6308)

Contents

0	Foundational material: The appendices	2
0.1	Appendix A: Sets	2
0.2	Appendix B: Functions	4
0.3	Appendix C: Fields	4
0.4	Appendix D: Complex Numbers	5
1	Vector spaces	7
1.1	Introduction	7
1.2	Vector Spaces	8
1.3	Subspaces	10
1.4	Linear Combinations and systems of linear equations	13
1.5	Linear dependence and linear independence	15
1.6	Bases and dimension	16
2	Linear transformations and matrices	21
2.1	Linear transformations, null spaces, and ranges	21
2.2	The matrix representation of a linear transformation	25
2.3	Composition of linear transformations and matrix multiplication	27

2.4	Invertibility and Isomorphisms	29
3	Elementary Matrix Operations and Systems of Linear Equations	31
3.1	Elementary matrix operations and elementary matrices . . .	31
3.2	The rank of a matrix and matrix inverses	32
3.3	Systems of linear equations – theoretical aspects	35
4	Determinants	38
4.1	Determinants of 2×2 matrices	38
4.2	Determinants of order n	40
4.3	Properties of determinants	48
5	Diagonalization	50
5.1	Eigenvalues and eigenvectors	50
5.2	Diagonalizability	53

These lecture notes are based on the textbook Linear Algebra, 4th edition, by Friedberg, Insel, and Spence, ISBN 0-13-008451-4. They are provided “as is” and as a courtesy only. They do not replace use of the textbook or attending class.

0 Foundational material: The appendices

0.1 Appendix A: Sets

Definition 0.1. A *set* is a collection of objects, called elements.

Example 0.2.

- $\{1, 2, 3\} = \{2, 1, 1, 1, 2, 3\}$ (no notion of “multiplicity”)
- $[1, 2] =$ the interval of reals between 1 and 2, including 1 and 2.

- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ (later)
- $\left\{ \left(\begin{array}{c} 8 \\ 0 \\ -1 \\ 2 \end{array} \right), \left(\begin{array}{c} 1 \\ 2 \\ 2 \\ 2 \end{array} \right) \right\} = \text{set of two vectors}$
- \emptyset : the empty set

Given two sets A, B , there are several operations that yield new sets from these. Most important are the following:

- $A \cup B$ (union of A and B)
- $A \cap B$ (intersection of A and B)
- $A \times B = \{(a, b) : a \in A, b \in B\}$ (product of A and B)

Definition 0.3. Let A be a set. A *relation* on A is a subset S of $A \times A$. Write $x \sim y$ if and only if $(x, y) \in S$.

Example 0.4. • $A = \{1, 2, 3\}, S = \{(1, 2), (1, 3), (2, 3)\}$. This relation is “ $<$ ”.

- $A = \{1, 2, 3\}, S = \{(1, 2), (1, 3), (2, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$. This relation is “ \neq ”.
- $A = \{1, 2, 3\}, S = \{(1, 1), (2, 2), (3, 3)\}$. This relation is “ $=$ ”.

Recall the following symbols. \forall : “for all”, \exists : “there exists”.

Definition 0.5. Let A be a set with a relation S . Then S is called an *equivalence relation* if and only if

- $\forall x \in A : x \sim x$ (reflexive)
- $\forall x, y \in A : x \sim y \Leftrightarrow y \sim x$ (symmetric)
- $\forall x, y, z \in A : (x \sim y \text{ and } y \sim z) \Rightarrow x \sim z$ (transitive)

Example 0.6. Let $A = \mathbb{Z}$. Let $x \sim y \Leftrightarrow \exists k \in \mathbb{Z} : x - y = 5k$. This defines an equivalence relation.

- reflexive: Let $x \in \mathbb{Z}$. Then $x - x = 0 = 5 \cdot 0$. Done.

- ii. symmetric: Let $x, y \in \mathbb{Z}$ with $x \sim y$. Then $x - y = 5k$ implies $y - x = 5 \cdot (-k)$. Done.
- iii. transitive: Let $x, y, z \in \mathbb{Z}$ with $x \sim y$ and $y \sim z$. Then $x - y = 5k_1$ and $y - z = 5k_2$ implies (by adding the two equalities) $x - z = 5 \cdot (k_1 + k_2)$. Done.

0.2 Appendix B: Functions

Definition 0.7. Let A, B be sets. A *function* $f : A \rightarrow B$ is a rule that associates to each element $x \in A$ a unique element of B , denoted $f(x)$. The set A is called the *domain*, the set B is called the *codomain*.

Definition 0.8. • For $S \subseteq A$, $f(S) = \{f(x) : x \in S\}$ (*image of S under f*). $f(A)$ is called the *range*.

- For $T \subseteq B$, $f^{-1}(T) = \{x \in A : f(x) \in T\}$ (*pre-image of T under f*)
- $f : A \rightarrow B = g : A \rightarrow B \Leftrightarrow \forall x \in A : f(x) = g(x)$

Definition 0.9. • $f : A \rightarrow B$ is *injective* if and only if $f(x) = f(y) \Rightarrow x = y$.

- $f : A \rightarrow B$ is *surjective* if and only if $\forall b \in B \exists a \in A : f(a) = b$.
- For $S \subseteq A$, the restriction of f to S is $f|_S : S \rightarrow B, x \mapsto f(x)$.

0.3 Appendix C: Fields

Definition 0.10. Let A be a set. A binary operation is any map $A \times A \rightarrow A$. We are very familiar with \mathbb{Q} and \mathbb{R} and the properties that the two binary operations $+$ and \cdot have.

Definition 0.11. A field F is a set with two binary operations labelled $+$ and \cdot such that

- i. $a + b = b + a$, $a \cdot b = b \cdot a$ (commutativity)
- ii. $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (associativity)
- iii. $\exists 0 \in F : a + 0 = a \ \forall a$
 $\exists 1 \in F : 1 \cdot a = a \ \forall a$ (neutral elements)

- iv. $\forall a \in A : \exists b \in A : a + c = 0$
 $\forall a \in A \setminus \{0\} : \exists b \in A : a \cdot b = 1$ (inverse elements)
- v. $a \cdot (b + c) = a \cdot b + a \cdot c$ (distributive law)

Theorem 0.12 (Cancellation Laws). *Let F be a field and $a, b, c \in F$.*

- i. $a + b = c + b \Rightarrow a = c$
- ii. $a \cdot b = c \cdot b$ and $b \neq 0 \Rightarrow a = c$

Proof. Part i. Let d be an additive inverse of b . Now, observe that $(a + b) + d = a$ and $(c + b) + d = c$. Done.

Part ii is done in detail in the textbook. □

Proposition 0.13. *The neutral element of addition is unique.*

Proof. Let 0 and $0'$ be two neutral elements of addition. Then

$$0 = 0 + 0' = 0'.$$

□

Example 0.14. Some examples of fields.

- $\mathbb{Q}, \mathbb{R}, \mathbb{C}$
- $\mathbb{Q}[\sqrt{3}] = \{a + b\sqrt{3} \mid a, b \in \mathbb{Q}\}$ (on Homework 1).
- $\mathbb{Z}/p\mathbb{Z}$ when p is a prime.

0.4 Appendix D: Complex Numbers

Motivation: In \mathbb{R} , $x^2 - 1 = 0$ has two solutions, namely $-1, 1$. However, the almost identical equation $x^2 + 1 = 0$ has no solutions. This means that the reals “leave something to be desired.” In response, we introduce the imaginary unit i , which has the property $i^2 = -1$.

Definition 0.15. A *complex number* is an expression of the form $z = a + bi$ with $a, b \in \mathbb{R}$. Sum and product are defined by

$$z + w = (a + bi) + (c + di) = a + c + (b + d)i$$

and

$$zw = (a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

Remark 0.16. Memorize the multiplication by multiplying out as one would do naively, and then use $i^2 = -1$. Do some examples!

Theorem 0.17. *The complex numbers with sum and multiplication as above form a field.*

Proof. This just involves tedious checking of all the properties—you should try a few yourself at home. \square

Remark 0.18. The multiplicative inverse of $z = a + bi$ is

$$\frac{1}{a + bi} = \frac{a - bi}{(a + bi)(a - bi)} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} + i \frac{-b}{a^2 + b^2}.$$

Definition 0.19. The *complex conjugate* of $z = a + bi$ is $\bar{z} = a - bi$.

Proposition 0.20. i. $\overline{\bar{z}} = z$

ii. $\overline{z + w} = \bar{z} + \bar{w}$

iii. $\overline{z\bar{w}} = \bar{z} \cdot w$

iv. $\overline{\frac{z}{w}} = \frac{\bar{z}}{\bar{w}}$

Remark 0.21. It is now clear that there is a bijection $\mathbb{C} \rightarrow \mathbb{R}^2$ via $a + bi \mapsto (a, b)$. By Pythagoras' Theorem, the length of a straight line from the origin to the point (a, b) is $\sqrt{a^2 + b^2}$.

Definition 0.22. The *absolute value* (or *modulus*) of $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$.

Remark 0.23. We have

$$z\bar{z} = (a + bi)(a - bi) = a^2 + b^2.$$

Thus,

$$|z| = \sqrt{z\bar{z}}.$$

Properties 0.24. i. $|zw| = |z||w|$

ii. $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$

iii. $|z + w| \leq |z| + |w|$

Theorem 0.25 (Fundamental Theorem of Algebra). *Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0$ be a complex polynomial (i.e., $a_i \in \mathbb{C}$). Then $\exists z_0 \in \mathbb{C} : p(z_0) = 0$.*

Proof. No proof is given here. This is a comparatively hard theorem to prove. □

1 Vector spaces

1.1 Introduction

Geometrically, a vector in, say, \mathbb{R}^2 , is the datum of a direction and a magnitude. Thus, it can be represented by an arrow which points in the given direction and has the given length. Two vectors can be added using the parallelogram rule (see the textbook for some nice pictures explaining this).

Physically, the vectors may, e.g., represent forces that are exerted on an object. The result of the addition is the resulting net force that the object experiences when the original two forces are applied.

Algebraically, when $v = (a_1, a_2)$ and $w = (b_1, b_2)$, then $v + w = (a_1 + b_1, a_2 + b_2)$. Scalar multiplication is defined via $t(a_1, a_2) = (ta_1, ta_2)$.

Definition 1.1. The vectors v and w are parallel if and only if $\exists t \in \mathbb{R} : tv = w$.

A vector can be interpreted as the *displacement vector* between its start and end point. If the start point is (x_1, x_2) and the end point is (y_1, y_2) , then the displacement vector is $(y_1 - x_1, y_2 - x_2)$.

Definition 1.2. The *line* through the points $A = (x_1, x_2)$ and $B = (y_1, y_2)$ is

$$\{(x_1, x_2) + t(y_1 - x_1, y_2 - x_2) : t \in \mathbb{R}\}.$$

Definition 1.3. The *line* through the points $A = (x_1, x_2, x_3)$ and $B = (y_1, y_2, y_3)$ is

$$\{(x_1, x_2, x_3) + t(y_1 - x_1, y_2 - x_2, y_3 - x_3) : t \in \mathbb{R}\}.$$

Definition 1.4. The *plane* through the points $A = (x_1, x_2, x_3)$, $B = (y_1, y_2, y_3)$ and $C = (z_1, z_2, z_3)$ (not all three on a line) is

$$\{(x_1, x_2, x_3) + s(y_1 - x_1, y_2 - x_2, y_3 - x_3) + t(z_1 - x_1, z_2 - x_2, z_3 - x_3) : s, t \in \mathbb{R}\}.$$

Example 1.5. i. The line through $(1, 1, 2)$ and $(0, 3, -1)$ is

$$\{(1, 1, 2) + t(-1, 2 - 3) : t \in \mathbb{R}\}.$$

ii. The plane through the points $A = (1, 0, -1)$, $B = (0, 1, 2)$ and $C = (1, 1, 0)$ is

$$\{(1, 0, -1) + s(-1, 1, 3) + t(0, 1, 1) : s, t \in \mathbb{R}\}.$$

Now, observe that vector addition and scalar multiplication satisfy certain laws, e.g., $v + w = w + v$, $1 \cdot v = v$, $(ab)v = a(bv)$. Next, we will distill these obvious properties into an abstract definition.

1.2 Vector Spaces

Definition 1.6. A *vector space* (or *linear space*) V over a field F (think $F = \mathbb{R}$, or \mathbb{C}) is a set with a binary operation denoted “+” and a second map $\cdot : F \times V \rightarrow V$ such that

- i. $\forall x, y \in V : x + y = y + x$
- ii. $\forall x, y, z \in V : (x + y) + z = x + (y + z)$
- iii. $\exists 0 \in V : \forall x \in V : x + 0 = x$
- iv. $\forall x \in V \exists y \in V : x + y = 0$
- v. $\forall x \in V : 1x = x$, where 1 is the neutral element of multiplication in F
- vi. $\forall a, b \in F \forall x \in V : (ab)x = a(bx)$
- vii. $\forall a \in F \forall x, y \in V : a(x + y) = ax + ay$
- viii. $\forall a, b \in F \forall x \in V : (a + b)x = ax + bx$

Definition 1.7. The elements of F are called *scalars*. The elements of V are called *vectors*. Because of item ii above, sums like $x + y + z + w$ are well-defined.

Remark 1.8. To simplify typing, we will usually not adorn vectors with an arrow, i.e., we will write x instead of \vec{x} and 0 instead of $\vec{0}$. Note that the neutral element of addition in the field is also denoted with 0 , but it should always be clear from the context what is meant.

Quiz 1.

1. (5 points) Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be two functions. Assume that the composition $g \circ f$ is injective. Does this necessarily imply that f is injective? Prove your answer.

2. (5 points) Let $h : \mathbb{Z} \rightarrow \mathbb{Z}$ be defined by $x \mapsto 2x + 4$. Is h injective? Is h surjective? Prove your answers.

Answer to 1. Yes. Let $x, y \in A$ with $f(x) = f(y)$. Apply g to both sides of the equality. Then $g(f(x)) = g(f(y))$. Since $g \circ f$ is injective, we can conclude $x = y$, q.e.d.

Answer to 2. Injective: Yes. Let $x, y \in \mathbb{Z}$ with $h(x) = h(y)$. Then $2x + 4 = 2y + 4$. Subtracting 4 on both sides and then dividing by 2 yields $x = y$, q.e.d. Surjective: No. For an arbitrary $x \in \mathbb{Z}$, $h(x) = 2x + 4$ is clearly an even integer. Therefore the range of h cannot be \mathbb{Z} , q.e.d.

Example 1.9. i. (THE example, see later section on isomorphisms) Take a field F . (We will mostly just take \mathbb{R} , or perhaps \mathbb{C} .) An n -tuple is (a_1, \dots, a_n) , where $a_1, \dots, a_n \in F$. Note that $\{n\text{-tuples}\} \cong F^n$ naturally. Define $(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$. Also, $c(a_1, \dots, a_n) = (ca_1, \dots, ca_n)$ for $c \in F$.

ii. $\text{Mat}_{m,n}(F)$ is a vector space with componentwise addition and scalar multiplication

The following examples of vector spaces are substantially different from the examples above. They are “infinite dimensional,” more about that later.

Example 1.10. i. Let S be a set of real numbers Let \mathcal{F} be the set of all real-valued functions on S . Then \mathcal{F} is a vector space (over \mathbb{R}) with the usual addition and scalar multiplication of real-valued functions.

ii. Let S now be an interval of reals. Consider in \mathcal{F} only those functions that are continuous. This is also a vector space (use the summation theorem for continuous functions from calculus)

iii. Consider in \mathcal{F} only those functions that are differentiable. This is also a vector space (use the summation theorem for differentiable functions from calculus)

- iv. Assume that $x_0 \in S$. Then $\{f : S \rightarrow \mathbb{R} \mid f(x_0) = 0\}$ is a vector space. (check it!). $\{f : S \rightarrow \mathbb{R} \mid f(x_0) = 1\}$ is not!

Again, we would like to infer more properties of vector spaces from the original list of 8 properties. To start, we observe that the zero vector is unique, with the same proof as in the case of fields in the Introduction. Moreover, we also have a cancellation law:

Theorem 1.11 (Cancellation law for vector spaces). *Let V be a vector space and $x, y, z \in V$. If $x + z = y + z$, then $x = y$.*

Proof. Let v be such that $z + v = 0$ (condition iv). Then

$$\begin{aligned} x &= x + 0 = x + (z + v) = (x + z) + v \\ &= (y + z) + v = y + (z + v) = y + 0 = y \end{aligned}$$

due to conditions ii and iii. □

Corollary 1.12. *The additive inverse is unique.*

Theorem 1.13. *Let V be a vector space. Then the following statements are true.*

i. $\forall x \in V : 0x = \vec{0}$

ii. $\forall x \in V \forall a \in F : (-a)x = -(ax) = a(-x)$

iii. $\forall a \in F : a\vec{0} = \vec{0}$

Proof. The textbook has detailed proofs of i and ii. The item iii is left to the reader. □

1.3 Subspaces

Definition 1.14. A subset W of a vector space V over the field F is called a *subspace of V* if W is a vector space with $+$ and scalar multiplication from V .

Example 1.15. • $\{\vec{0}\}, V$

• $\mathbb{R}^2 \cong \{(a, b, 0) \mid a, b \in \mathbb{R}\} \subset \mathbb{R}^3$

- The above examples 1.10ii and 1.10iii in 1.10i.

In order to verify that W is a subspace of V , it is not necessary to check all the vector space axioms in the definition of a vector space. For example, the restricted addition is clearly commutative since it was already commutative before the restriction.

Theorem 1.16. *A nonempty subset W of the vector space V is a subspace of V if and only if*

- i. $\forall x, y \in W : x + y \in W$ (closedness under $+$)*
- ii. $\forall c \in F \forall x \in W : c \cdot x \in W$ (closedness under scalar multiplication)*

Proof. First, observe that the implication \Rightarrow is trivial. The proof of the other direction consists of some easy verifications. For example, let's see why $\vec{0} \in W$: Take an arbitrary element x of W . Since W is nonempty, such an element exists. Now, simply observe that $0 \cdot x = \vec{0}$, which is an element of W by ii. The remaining details are left to the reader. \square

More examples:

Example 1.17. • Let $W = \{(a, b) | a + b = 0\} \subset \mathbb{R}^2$. Closedness under $+$ is checked as follows. Let $(a, b), (c, d) \in W$. Then the result of the addition is $(a+c, b+d)$, which satisfies $(a+c) + (b+d) = (a+b) + (c+d) = 0 + 0 = 0$. Closedness under scalar multiplication is seen as follows. Let $(a, b) \in W$ and c a scalar. Then the result of the scalar multiplication is (ca, cb) , which satisfies $ac + cb = c(a + b) = c \cdot 0 = 0$.

- Let $W = \{(a, b, c) | 3a - b + 2c = 0\} \subset \mathbb{R}^3$. Check it as in i.
- What about $W = \{(a, b, c) | 3a - b + 2c = 1\} \subset \mathbb{R}^3$? Let $(a, b, c), (d, e, f) \in W$. Then the result of their addition is $(a+d, b+e, c+f)$, which satisfies $3(a+d) - b - e + 2(c+f) = 3a - b + 2c + 3d - e + 2f = 1 + 1 = 2 \neq 1$. Thus, W is not closed under addition and not a subspace.
- Let $W = \{(a, b) | a^2 - b^2 = 0\} \subset \mathbb{R}^2$. The subset W is not a subspace because, for example, $(1, 1)$ and $(2, -2)$ are elements of W , but $(1, 1) + (2, -2) = (3, -1)$ is not, due to $3^2 - (-1)^2 = 8 \neq 0$.
- Any intersection of subspaces in a vector space is itself a subspace.

A major class of examples is given by sums and direct sums of subspaces.

Definition 1.18. Let V be a vector space. Let S, T be nonempty subsets of V . Then let $S + T = \{x + y | x \in S, y \in T\}$. We call $S + T$ the *sum* of S and T .

Definition 1.19. Let V be a vector space. Let W, U be subspaces of V . Then we call V the *direct sum* of W, U if $W + U = V$ and $W \cap U = \{0\}$. Write $V = W \oplus U$.

Proposition 1.20. Let V be a vector space. Let W, U be subspaces of V . Then the sum $W + U$ is a subspace of V (containing both W and U).

Proof. $(w_1 + u_1) + (w_2 + u_2) = (w_1 + w_2) + (u_1 + u_2)$, which is the sum of a vector in W , namely $w_1 + w_2$, and a vector in U , namely $u_1 + u_2$. Thus, $W + U$ is closed under addition. The closedness under scalar multiplication is completely analogous. \square

Example 1.21. • $\{(a, b, 0, c) | a, b, c \in \mathbb{R}\} + \{(d, 0, e, f) | d, e, f \in \mathbb{R}\} = \mathbb{R}^4$. But this is not a direct sum.

• $\{(a, 0, 0, b) | a, b \in \mathbb{R}\} \oplus \{(0, c, d, 0) | c, d \in \mathbb{R}\} = \mathbb{R}^4$. This is a direct sum.

• $\{(a, 0, 0) | a \in \mathbb{R}\} \oplus \{(0, b, 0) | b \in \mathbb{R}\} = \{(a, b, 0) | a, b \in \mathbb{R}\}$

Quiz 2:

1. (5 points) Let V be a vector space and let W_1, W_2 be subspaces of V . Prove that the intersection $W_1 \cap W_2$ is a subspace of V .

2. (5 points) Let V be the vector space of 2×2 matrices with real entries. (You may assume without proof that V is a vector space with the usual componentwise addition and scalar multiplication.) Recall that the determinant of such a matrix is defined to be

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc.$$

Let W be the subset of V consisting of those matrices whose determinant is equal to 0. Is W a subspace of V ? Prove your answer.

Solution:

1. We need to prove that $W_1 \cap W_2$ is non-empty, closed under addition and closed under scalar multiplication.

- Being subspaces, both W_1 and W_2 contain the zero-vector $\vec{0}$. So $\vec{0} \in W_1 \cap W_2 \neq \emptyset$.
- Let $v, w \in W_1 \cap W_2$. In particular, $v, w \in W_1$ and by the closedness of W_1 under addition, $v + w \in W_1$. Also, $v, w \in W_2$ and by the closedness of W_2 under addition, $v + w \in W_2$. Altogether, $v + w \in W_1 \cap W_2$.
- Let $v \in W_1 \cap W_2$ and $c \in F$. In particular, $v \in W_1$ and by the closedness of W_1 under scalar multiplication, $cv \in W_1$. Also, $v \in W_2$ and by the closedness of W_2 under scalar multiplication, $cv \in W_2$. Altogether, $cv \in W_1 \cap W_2$.

2. No! Simply observe that $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in W$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in W$, but

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \notin W.$$

This counterexample shows that W is not closed under $+$.

1.4 Linear Combinations and systems of linear equations

Definition 1.22. Let V be a vector space and S a nonempty subset of V . We call $v \in V$ a *linear combination* of vectors in S if there exist vectors $u_1, \dots, u_n \in S$ and scalars $a_1, \dots, a_n \in F$ such that $v = a_1u_1 + \dots + a_nu_n$.

Example 1.23. • $(3, 4, 1) = 3(1, 0, 0) + 4(0, 1, 0) + 1(0, 0, 1)$.

- If we want to write $(3, 1, 2)$ as a linear combination of $(1, 0, 1)$, $(0, 1, 1)$, $(1, 2, 1)$, how do we find the coefficients a_1, a_2, a_3 ? Answer: Make the Ansatz

$$(3, 1, 2) = a_1(1, 0, 1) + a_2(0, 1, 1) + a_3(1, 2, 1)$$

and solve the system of linear equations

$$a_1 + a_3 = 3, \quad a_2 + 2a_3 = 1, \quad a_1 + a_2 + a_3 = 2.$$

Solution: $a_1 = 2, a_2 = -1, a_3 = 1$.

- Find the a_i in a given situation may or may not be possible. E.g., writing

$$(3, 1, 2) = a_1(1, 0, 0) + a_2(0, 1, 0) + a_3(1, 2, 0)$$

is clearly impossible.

- There may be many choices for the a_i :

$$(2, 6, 8) = a_1(1, 2, 1) + a_2(-2, -4, -2) + a_3(0, 2, 3) + a_4(2, 0, -3) + a_5(-3, 8, 16)$$

is equivalent to

$$(a_1, a_2, a_3, a_4, a_5) \in \{(-4 + 2s - t, s, 7 - 3t, 3 + 2t, t) \mid s, t \in \mathbb{R}\}.$$

(There are two “free variables”.)

There are three types of operations that we used to solve the above systems of linear equations:

- Interchange the order of any two equations.
- Multiply an equation by a nonzero scalar.
- Add one equation to another.

Key point: These operations do not change the set of solutions.

Definition 1.24. Let V be a vector space. Let S be a nonempty subset of V . We call $\text{span}(S)$ the set of all vectors in V that can be written as a linear combination of vectors in S .

Example 1.25. Let $S = \{(1, 0, 0), (0, 1, 0), (2, 1, 0)\}$. Then $\text{span}(S) = \{(s, t, 0) \mid s, t \in \mathbb{R}\}$.

Theorem 1.26. *The span of any subset S of a vector space V is a subspace of V .*

Proof. Let $v = a_1u_1 + \dots + a_nu_n \in \text{span}(S)$. Then $cv = (ca_1)u_1 + \dots + (ca_n)u_n \in \text{span}(S)$. Thus, closedness under scalar multiplication is ok.

Let $v = a_1v_1 + \dots + a_nv_n \in \text{span}(S)$ and let $w = b_1w_1 + \dots + b_mw_m \in \text{span}(S)$. Then

$$v + w = a_1v_1 + \dots + a_nv_n + b_1w_1 + \dots + b_mw_m.$$

Thus, closedness under addition is ok. □

Example 1.27. • $S = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ spans \mathbb{R}^3 .

- $S = \{(1, 2), (2, 1)\}$ spans \mathbb{R}^2 .

- $S = \{(1, 1), (2, 2)\}$. $\text{span}(S) = \{(s, s) | s \in \mathbb{R}\} \neq \mathbb{R}^2$.
- $S = \{(1, 2)\}$ does not span \mathbb{R}^2 .
- Which (a, b, c) are in $\text{span}(\{(1, 1, 2), (0, 1, 1), (2, 1, 3)\})$? Answer: Those that satisfy $a + b = c$.

1.5 Linear dependence and linear independence

Motivation: Let W be a subspace of V . We are interested in a set $S \subset W$ such that $\text{span}(S) = W$ and S is “as small as possible”.

Definition 1.28. A subset S of a vector space V is called *linearly dependent* if there exist a finite number of vectors $u_1, \dots, u_n \in S$ and scalars a_1, \dots, a_n , not all equal to zero, such that

$$a_1u_1 + \dots + a_nu_n = 0.$$

We also say that the vectors in S are linearly dependent.

Quiz 3: (10 points) Find the condition(s) for $(a, b, c, d) \in \mathbb{R}^4$ to be in

$$\text{span}\{(1, 0, 1, -1), (-1, -2, -1, -1), (0, 1, 0, 1), (1, 3, 1, 2)\}.$$

Solution:

Make the Ansatz

$$a_1(1, 0, -1, 1) + a_2(-1, -2, -1, -1) + a_3(0, 1, 0, 1) + a_4(1, 3, 1, 2) = (a, b, c, d).$$

We have to determine for which values of a, b, c, d the above system is consistent, i.e., solvable. Reducing this system to echelon form yields:

$$\begin{aligned} a_1 - a_2 \quad \quad + a_4 &= a \\ -2a_2 + a_3 + 3a_4 &= b \\ 0 &= c - a \\ 0 &= d + a - b \end{aligned}$$

It is now clear that the system is consistent if and only if $c = a$ and $b = a + d$. Done.

Example 1.29. Let $S = \{(1, 3, -4, 2), (2, 2, -4, 0), (1, -3, 2-4), (-1, 0, 1, 0)\}$. Then

$$4(1, 3, -4, 2) - 3(2, 2, -4, 0) + 2(1, -3, 2-4) + 0(-1, 0, 1, 0) = (0, 0, 0, 0).$$

Thus, S is linearly dependent.

Definition 1.30. If S is not linearly dependent, we say S is *linearly independent*.

Remark 1.31. Linear independence is equivalent to: “ $\sum a_i v_i = \vec{0} \Rightarrow$ all $a_i = 0$ ”.

Remark 1.32. The empty set \emptyset is linearly independent. The singleton set $\{v\}$ is linearly independent if and only if $v \neq \vec{0}$.

Theorem 1.33. Let V be a vector space. If $S_1 \subseteq S_2$ and S_1 is linearly dependent, then S_2 is linearly dependent.

Proof. This is immediate from the definition. □

Theorem 1.34. Let S be a linearly independent subset of V . Let $v \in V \setminus S$. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \text{span}(S)$.

Proof. “ \Rightarrow ”. Write $a_1 u_1 + \dots + a_n u_n + a_{n+1} v = 0$ with not all a_i equal to zero and $u_i \in S$.

Claim: $a_{n+1} \neq 0$.

Proof of claim: If $a_{n+1} = 0$, then at least one of a_1, \dots, a_n is not equal to zero and $a_1 u_1 + \dots + a_n u_n = 0$. Contradiction to linear independence of S .

So, $a_{n+1} \neq 0$, and we can write $v = \frac{-a_1}{a_{n+1}} u_1 + \dots + \frac{-a_n}{a_{n+1}} u_n$, qed.

“ \Leftarrow ”. Write $v = a_1 u_1 + \dots + a_n u_n$. Then $a_1 u_1 + \dots + a_n u_n + (-1)v = 0$. Thus, $S \cup \{v\}$ is linearly dependent, qed. □

1.6 Bases and dimension

Definition 1.35. Let V be a vector space. A *basis* β is a linearly independent subset of V which satisfies $\text{span}(\beta) = V$.

Theorem 1.36. *Let V be a vector space. Let $\beta = \{u_1, \dots, u_n\}$ be a subset of V . Then*

$$\beta \text{ is a basis} \Leftrightarrow \forall v \in V : \exists! a_1, \dots, a_n \in F : v = a_1 u_1 + \dots + a_n u_n.$$

(Recall that $\exists!$ means **unique existence**.)

Proof. “ \Rightarrow ”. Spanning property is already known. We just have to prove uniqueness.

Let

$$a_1 u_1 + \dots + a_n u_n = v = b_1 u_1 + \dots + b_n u_n.$$

This implies

$$(a_1 - b_1)u_1 + \dots + (a_n - b_n)u_n = 0.$$

Linear independence implies $a_1 - b_1 = 0, \dots, a_n - b_n = 0$. Done.

“ \Leftarrow ”. Spanning property is already known. To show lin. indep., just observe that

$$a_1 u_1 + \dots + a_n u_n = 0$$

is solved by the trivial solution $a_1 = \dots = a_n = 0$. However, by assumption, this is the only solution. Thus, we have established linear independence. Done. \square

Theorem 1.37. *Let V be a vector space. Let S be a finite subset of V with $\text{span}(S) = V$. Then there exists a subset of S which is a basis for V . In particular, V has a finite basis.*

Proof. We conduct this proof by induction over the cardinality of S .

If $\#S = 1$, then $S = \{v\}$, and S is clearly linearly independent (unless we are in a trivial cases).

Now, assume that we know the theorem for $\#S = n$. We have to prove it for $\#S = n + 1$.

If S is not a basis, then S is lin. dep. Claim: $\exists v \in S : V = \text{span}(S) = \text{span}(S \setminus \{v\})$. Proof of Claim: lin. dep. means that there is a linear combination

$$a_1 u_1 + \dots + a_n u_n = 0$$

with some $a_{i_0} \neq 0$. We can solve the above equation for u_{i_0} . It is now clear that a linear combination of the vectors u_1, \dots, u_n can be expressed as a

linear combination of the vectors $u_1, \dots, u_{i_0-1}, u_{i_0+1}, \dots, u_n$. Thus, letting $v = u_{i_0}$ establishes the Claim.

If we let $\tilde{S} = S \setminus \{u_{i_0}\}$, then we can apply the induction hypothesis to obtain that \tilde{S} contains a basis. Since $\tilde{S} \subset S$, we can conclude that S contains a basis. Done. \square

Example 1.38. • Let $S = \{(1, 0), (1, 1), (2, 3)\}$. Observe that S spans \mathbb{R}^2 , but S is lin. dep. After removing any one of the three vectors from S , we obtain a basis.

- Let $S = \{(1, 0), (0, 1), (0, 2)\}$. Observe that S spans \mathbb{R}^2 , but S is lin. dep. After removing the second or third vector, we obtain a basis. However, removing the first vector does not yield a basis.
- Let $S = \{(2, -3, 5), (8, -12, 20), (1, 0, -2), (0, 2, -1), (7, 2, 0)\}$. Observe that S spans \mathbb{R}^3 , but S is lin. dep.. Consider the span of the first vector. Obviously, the span remains unchanged after adding the second vector (which is 4 times the first), so the second vector should be removed. The third vector is not a multiple of the first, so we keep it. A direct computation shows that the first, third and fourth vector are lin. indep. and span \mathbb{R}^3 . The fifth can be disregarded.

Theorem 1.39 (Replacement Theorem). *Let V be a vector space. Let $V = \text{span}(G)$, where G is a subset of V of cardinality n . Let L be a linearly independent subset of V of cardinality m . Then the following holds.*

- $m \leq n$
- there exists a subset $H \subseteq G$ of cardinality $n - m$ such that $\text{span}(L \cup H) = V$

Remark 1.40. A typical situation is for example $m = 2$ and $n = 5$, i.e., $L = \{v_1, v_2\}$ and $G = \{w_1, w_2, w_3, w_4, w_5\}$. The replacement theorem now says that there are two vectors in G that can be **replaced** with the two vectors from L such that the set obtained by the replacement still spans V . In other words, L can be injected into G and the result still spans V .

Corollary 1.41. *Let V be a vector space with a finite basis. Then all bases contain the same number of elements.*

Proof. Let β basis of cardinality m and γ basis of cardinality n . Since β lin. indep. and γ spans, we have $m \leq n$. By symmetry, we have $m = n$. \square

Definition 1.42. A vector space is called *finite dimensional* if there exists a basis consisting of finitely many vectors. The unique cardinality of a basis of a finite dimensional vector space is called the *dimension* of V , denoted $\dim(V)$.

Example 1.43. $\dim(\mathbb{R}^n) = n, \dim(\text{Mat}_{m \times n}) = mn$. (Consider the standard bases.)

Here are some more Corollaries.

Corollary 1.44. *Let V be a vector space of dimension n . Then any generating set S of V contains at least n elements.*

Proof. By Theorem 1.37, S contains a basis. By Corollary 1.41, that basis has n elements. So S contains at least n elements. \square

Corollary 1.45. *Let V be a vector space and $S \subset V$ a subset. If $V = \text{span}(S)$ and $\#S = \dim(V)$, then S is a basis.*

Proof. By Theorem 1.37, S contains a basis. This basis must have $\dim V = \#S$ elements. Thus, this basis is S itself. \square

Corollary 1.46. *Let V be a vector space and $S \subset V$ a subset. If S is lin. indep. and $\#S = \dim(V)$, then S is a basis.*

Proof. Take any basis G . Apply the Replacement Theorem with G and $L = S$. Since $\#G = \#S = \dim V$, we have $H = \emptyset$ and $V = \text{span } G = \text{span } S$. \square

Corollary 1.47. *Let V be a vector space. Every lin. indep. subset S of V can be extended to a basis.*

Proof. Take any basis G of V . Apply the Replacement Theorem with $S = L$ and G . \square

Finally, let us prove the Replacement Theorem.

Proof of Replacement Theorem. For a fixed $n = \#G$, we do induction over $\#L = m$.

For $m = 0$, we have $L = \emptyset$. Take $H = G$. Done.

Induction step: “ $m \rightarrow m + 1$ ”.

Let $L = \{v_1, \dots, v_{m+1}\}$, let $\tilde{L} = \{v_1, \dots, v_m\}$.

Induction hypothesis $\Rightarrow \exists \tilde{H} = \{u_1, \dots, u_{n-m}\}$ such that $V = \text{span}(\tilde{L} \cup \tilde{H})$.

Write

$$v_{m+1} = a_1v_1 + \dots + a_mv_m + b_1u_1 + \dots + b_{n-m}u_{n-m}. \quad (1)$$

Since L is lin. indep. we know that there exists i such that $b_i \neq 0$. Thus $n - m > 0$, i.e., $n \geq m + 1$. This proves the first part of the claim for $m + 1$.

It remains to show that if, w.l.o.g., $b_1 \neq 0$, then $H = \{u_2, \dots, u_{n-m}\}$ works, i.e., $V = \text{span}(L \cup H)$. Let $v \in V$ be arbitrary. We know we can write

$$v = \alpha_1v_1 + \dots + \alpha_mv_m + \gamma_1u_1 + \dots + \gamma_{n-m}u_{n-m}. \quad (2)$$

If we solve (1) for u_1 and substitute into (2), we see that v can be written as a linear combination of $v_1, \dots, v_{m+1}, u_2, \dots, u_{n-m}$. Done. \square

Now, let us discuss the dimension of subspaces.

Theorem 1.48. *Let V be a vector space. Let W be a subspace of V . Assume $\dim V$ is finite. Then $\dim W \leq \dim V$ and equality holds if and only if $V = W$.*

Proof. This is immediate from the Replacement Theorem. \square

In the following examples, the task is to find a basis for (and the dimension of) the subspace W .

Example 1.49.

- Let $V = \mathbb{R}^3$. Let $W = \{(a_1, a_2, a_3) \mid a_1 + a_3 = 0 \text{ and } a_1 + a_2 - a_3 = 0\}$. Solving the system

$$a_1 + a_3 = 0 \text{ and } a_1 + a_2 - a_3 = 0$$

yields $W = \{(-t, 2t, t) \mid t \in \mathbb{R}\}$. Thus $\{(-1, 2, 1)\}$ is a basis, and the dimension of W is one.

- Let $V = \mathbb{R}^5$. Let $W = \{(a_1, a_2, a_3, a_4, a_5) \mid a_1 + a_3 + a_5 = 0 \text{ and } a_2 = a_4\}$. Solving the system

$$a_1 + a_3 + a_5 = 0 \text{ and } a_2 = a_4$$

yields

$$\begin{aligned} W &= \{(-a_3 - a_5, a_4, a_3, a_4, a_5) \mid a_3, a_4, a_5 \in \mathbb{R}\} \\ &= \{a_3(-1, 0, 1, 0, 0) + a_4(0, 1, 0, 1, 0) + a_5(-1, 0, 0, 0, 1) \mid a_3, a_4, a_5 \in \mathbb{R}\} \end{aligned}$$

Thus $\{(-1, 0, 1, 0, 0), (0, 1, 0, 1, 0), (-1, 0, 0, 0, 1)\}$ is a basis for W , and the dimension of W is three.

Quiz 4:

(10 points) Let u, v, w be pairwise distinct vectors in a vector space V . Prove that if $\{u, v, w\}$ is a basis for V , then $\{u + v + w, v + w, w\}$ is also a basis for V .

Solution:

Since $\{u, v, w\}$ has 3 elements, we know that $\dim V = 3$. By a corollary to the Replacement Theorem, a set of 3 vectors is a basis of V if and only if the 3 vectors are linearly independent. So let's check this:

Ansatz:

$$a(u + v + w) + b(v + w) + cw = \vec{0}.$$

This is equivalent to

$$au + (a + b)v + (a + b + c)w = \vec{0}.$$

Since u, v, w are linearly independent, we know that the coefficients $a = a + b = a + b + c = 0$. This clearly implies $a = b = c = 0$, qed.

2 Linear transformations and matrices

2.1 Linear transformations, null spaces, and ranges

Definition 2.1. Let V, W be vector spaces over the same field F . We call a function $T : V \rightarrow W$ a *linear transformation* from V to W if

i. $\forall x, y \in V : T(x + y) = T(x) + T(y)$

ii. $\forall c \in F \forall x \in V : T(cx) = cT(x)$

Remark 2.2. We say T is *linear* for short.

Properties 2.3. • $T(\vec{0}) = \vec{0}$

• $T(x - y) = T(x) - T(y)$

• $T(a_1v_1 + \dots + a_nv_n) = a_1T(v_1) + \dots + a_nT(v_n)$

Example 2.4. • $T(a_1, a_2) = (2a_1 + a_2, a_1)$ (Check it!)

• $T : \mathbb{R}^5 \rightarrow \mathbb{R}^7, T(a_1, \dots, a_5) = (a_1, a_2, 0, a_7, 0, 0, a_1)$

• $T : C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R}), T(f) = \frac{df}{dx}$

Definition 2.5. Let V, W be vector spaces. Let $T : V \rightarrow W$ linear. We define the *null space* (aka *kernel*) of T to be

$$N(T) = \{x \in V : T(x) = \vec{0}\}.$$

Remark 2.6. Recall that the range of T is

$$R(T) = \{T(x) : x \in V\}.$$

Example 2.7. Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2, T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3)$. To find the null space, set

$$T(a_1, a_2, a_3) = (0, 0) \Leftrightarrow a_1 - a_2 = 0 \text{ and } 2a_3 = 0.$$

The solution of the above system of two equations in three variables is

$$N(T) = \{(t, t, 0) | t \in \mathbb{R}\}.$$

Moreover, it is clear that T is onto, so $R(T) = \mathbb{R}^2$.

Theorem 2.8. Let V, W be vector spaces and $T : V \rightarrow W$ linear. Then

i. $N(T)$ is a subspace of V

ii. $R(T)$ is a subspace of W

Proof. i. $N(T)$ is non-empty because $\vec{0} \in N(T)$. Now, we just have to check closedness. If $T(x) = 0$ and $T(y) = 0$, then $T(ax + by) = aT(x) + bT(y) = 0 + 0 = 0$. Done.

ii. Again, just closedness. If $T(x) = v_1$ and $T(y) = v_2$, then $av_1 + bv_2 = aT(x) + bT(y) = T(ax + by)$. Done. \square

Theorem 2.9. *Let V, W be vector spaces and $T : V \rightarrow W$ linear. Let $\{v_1, \dots, v_n\}$ be a basis for V . Then $R(T) = \text{span}\{T(v_1), \dots, T(v_n)\}$.*

Proof. Let $v = a_1v_1 + \dots + a_nv_n$. Then $T(v) = T(a_1v_1 + \dots + a_nv_n) = a_1T(v_1) + \dots + a_nT(v_n) \in \text{span}\{T(v_1), \dots, T(v_n)\}$. Done. \square

Example 2.10. Problem: Find (a basis for) $R(T)$ when $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3, T(a_1, a_2, a_3) = (a_1 - 2a_2, a_2 + a_3, 2a_1 + a_2 + 5a_3)$.

First, we note $T(1, 0, 0) = (1, 0, 2)$, $T(0, 1, 0) = (-2, 1, 1)$, $T(0, 0, 1) = (0, 1, 5)$. Thus, according to Theorem 2.9, $R(T) = \text{span}\{(1, 0, 2), (-2, 1, 1), (0, 1, 5)\}$. Now, note that twice the first vector plus the second equals the third, so $R(T) = \text{span}\{(1, 0, 2), (-2, 1, 1)\}$. The set $\{(1, 0, 2), (-2, 1, 1)\}$ is clearly a basis for $R(T)$, and $\dim R(T) = 2$.

Note that $N(T)$ is easily computed to be one-dimensional, and $\dim N(T) + \dim R(T) = 3$.

Definition 2.11. Let V, W be vector spaces and $T : V \rightarrow W$ linear. If $N(T), R(T)$ are finite dimensional, then let

$$\text{nullity}(T) = \dim N(T), \quad \text{rank}(T) = \dim R(T).$$

Theorem 2.12 (Dimension Theorem). *Let V, W be vector spaces and $T : V \rightarrow W$ linear. If V is finite-dimensional, then*

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

Proof. Let $\{v_1, \dots, v_k\}$ be a basis for $N(T)$. In particular, $k = \text{nullity}(T)$. Let $n = \dim V$. The Replacement Theorem implies that there are vectors $v_{k+1}, \dots, v_n \in V$ such that $\{v_1, \dots, v_n\}$ is a basis for V .

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

Spanning: Let $v = a_1v_1 + \dots + a_nv_n \in V$ arbitrary. Then

$$T(v) = a_1T(v_1) + \dots + a_kT(v_k) + a_{k+1}T(v_{k+1}) + \dots + a_nT(v_n).$$

However, the first k summands are zero due to $v_1, \dots, v_k \in N(T)$.

Lin. indep.: Write

$$b_{k+1}T(v_{k+1}) + \dots + b_nT(v_n) = 0.$$

We have to conclude $b_{k+1} = \dots = b_n = 0$. To do this, note that $b_{k+1}T(v_{k+1}) + \dots + b_nT(v_n) = T(b_{k+1}v_{k+1} + \dots + b_nv_n)$, i.e., $b_{k+1}v_{k+1} + \dots + b_nv_n \in N(T)$. Thus, there exist a_1, \dots, a_k :

$$a_1v_1 + \dots + a_kv_k = b_{k+1}v_{k+1} + \dots + b_nv_n.$$

Since $\{v_1, \dots, v_n\}$ is a basis for V and thus lin. indep., this is only possible if

$$a_1 = \dots = a_k = b_{k+1} = \dots = b_n = 0.$$

□

Theorem 2.13. *Let V, W vector spaces. Let $T : V \rightarrow W$ linear. Then T is one-to-one if and only if $N(T) = \{\vec{0}\}$.*

Proof. \Rightarrow . Saw: $T(\vec{0}) = \vec{0}$. Since T is one-to-one, this implies $N(T) = \{\vec{0}\}$.

\Leftarrow . Assume $T(x) = T(y)$. Then $T(x) - T(y) = 0$. By linearity of T , $T(x - y) = \vec{0}$. By assumption, $x - y = \vec{0}$. Done. □

Quiz 5:

- i. (5 points) Carefully state the Dimension Theorem, as discussed in class. Be sure to define in detail the terms appearing in the formula. Do NOT provide a proof.
- ii. (2 points) Let V, W be vector spaces and $T : V \rightarrow W$ a linear transformation. Prove that T is injective if and only if the null space of T is the zero vector space. (This was discussed as Theorem 2.4 in class. I am asking you to reproduce the proof.)
- iii. (3 points) Let V, W be vector spaces and $T : V \rightarrow W$ a linear transformation. Let $\dim V = \dim W < \infty$. Prove that T is injective if and only if T is surjective.

Solution:

(a) See above

(b) See above

(c) “ \Rightarrow ” T injective implies $\text{nullity}(T) = 0$. The formula in the Dimension Theorem thus becomes $\text{rank}(T) = \dim V = \dim W$. Therefore, $\text{Range}(T)$ is a subspace of W of the same dimension as W . This implies $\text{Range}(T) = W$, i.e., T is surjective.

“ \Leftarrow ” The formula in the Dimension Theorem yields $\text{nullity}(T) = 0$. Therefore, the null space of T is the zero vector space. By (b), T is injective. \square

Theorem 2.14. *Let V, W vector spaces. Let $\{v_1, \dots, v_n\}$ be a basis for V . Let w_1, \dots, w_n be a list of arbitrary vectors in W . Then there exists a unique $T : V \rightarrow W$ linear such that $T(v_i) = w_i$ for all $i = 1, \dots, n$.*

Proof. Recall that an arbitrary $v \in V$ can be written as $v = \sum_i a_i v_i$ with unique coefficients a_i . Then set $T(v) = \sum_i a_i w_i$. It is easy to check that this defines a well-defined linear map as required in the Theorem. Uniqueness is also clear. \square

Corollary 2.15. *Let V, W vector spaces. Let $U, T : V \rightarrow W$ linear with $U(v_i) = T(v_i)$ on a basis $\{v_1, \dots, v_n\}$ for V . Then $U = T$.*

2.2 The matrix representation of a linear transformation

Definition 2.16. Let V be a finite dimensional vector space. An *ordered basis* for V is a basis endowed with a specific order.

Example 2.17. As ordered bases,

$$\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} \neq \{(0, 1, 0), (1, 0, 0), (0, 0, 1)\}.$$

Definition 2.18. Let $\beta = \{u_1, \dots, u_n\}$ ordered basis for V . We saw earlier:

$$\forall x \in V \exists! a_1, \dots, a_n : x = a_1 u_1 + \dots + a_n u_n.$$

Write

$$[x]_\beta = (a_1, \dots, a_n)$$

for the *coordinate vector* of x relative to β . In particular, $[u_i]_\beta = e_i$.

Definition 2.19. Take V with $\beta = \{v_1, \dots, v_n\}$, W with $\gamma = \{w_1, \dots, w_m\}$. Let $T : V \rightarrow W$ linear. Write

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

for $j = 1, \dots, n$. Call the matrix (a_{ij}) the *matrix representation of T* with respect to β and γ . When $V = W$ and $\beta = \gamma$, write $A = [T]_\beta$.

Remark 2.20. The key fact to remember is that the j -th **column** of the matrix representation is $[T(v_j)]_\gamma$.

Example 2.21. (a) Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ be given by $T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$. Let β and γ be the respective standard bases. Then

$$T(1, 0) = (1, 0, 2), \quad T(0, 1) = (3, 0, -4).$$

Thus,

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

(b) Same map, but with $\gamma' = \{e_2, e_1, e_3\}$:

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

(c) Same map as in (a), but with $\beta' = \{e_2, e_1\}$:

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

(d) Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ be given by $T(a_1, a_2) = (a_1 - a_2, a_1, 2a_1 + a_2)$. Let β be the standard basis and $\gamma = \{(1, 1, 0), (0, 1, 1), (2, 2, 3)\}$. By solving a system of linear equations, we find

$$T(1, 0) = (1, 1, 2) = -\frac{1}{3}(1, 1, 0) + \frac{2}{3}(2, 2, 3),$$

and

$$T(0, 1) = (-1, 0, 1) = -(1, 1, 0) + (0, 1, 1).$$

Thus,

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

Definition 2.22. Let $U, T : V \rightarrow W$ be linear. Then

$$(U + T)(x) = U(x) + T(x)$$

and

$$(cT)(x) = cT(x).$$

Theorem 2.23. Let V, W be given vector spaces. The set of all linear transformations $V \rightarrow W$ is a vector space with $+$ and \cdot defined as above. Write $\mathcal{L}(V, W)$ for this vector space. Write $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$.

Proof. Check the axioms! □

Definition 2.24. Let $U, T : V \rightarrow W$ linear. Then

i. $[U + T]_{\beta}^{\gamma} = [U]_{\beta}^{\gamma} + [T]_{\beta}^{\gamma}$

ii. $[aT]_{\beta}^{\gamma} = a[T]_{\beta}^{\gamma}$

Proof. Write $U(v_j) = \sum a_{ij}w_i$, $T(v_j) = \sum b_{ij}w_i$. Then

$$(U + T)(v_j) = U(v_j) + T(v_j) = \sum a_{ij}w_i + \sum b_{ij}w_i = \sum (a_{ij} + b_{ij})w_i.$$

Thus, the ij entry of $[U + T]_{\beta}^{\gamma}$ is $a_{ij} + b_{ij}$. □

2.3 Composition of linear transformations and matrix multiplication

Theorem 2.25. Let V, W, Z be vector spaces over the same field. Let $T : V \rightarrow W$ and $U : W \rightarrow Z$ be linear. Then $U \circ T : V \rightarrow Z$ is linear.

Proof.

$$\begin{aligned} (U \circ T)(ax + by) &= U(T(ax + by)) = U(aT(x) + bT(y)) = aU(T(x)) + bU(T(y)) \\ &= a(U \circ T)(x) + b(U \circ T)(y). \end{aligned}$$

□

Theorem 2.26. Let $U, S, T : V \rightarrow V$ linear. Then

- $U \circ (S + T) = U \circ S + U \circ T$

- $(U + S) \circ T = U \circ T + S \circ T$
- $U \circ (S \circ T) = (U \circ S) \circ T$
- $\text{id} \circ U = U \circ \text{id} = U$
- $a(U \circ S) = (aU) \circ S = U \circ (aS)$

Now, let us investigate the matrix of a composition of linear transformations. Let $T : V \rightarrow W$, $U : W \rightarrow Z$. Let $\alpha = \{v_j | j = 1, \dots, n\}$, $\beta = \{w_k | k = 1, \dots, m\}$, $\gamma = \{z_i | i = 1, \dots, p\}$ be the corresponding ordered basis, in alphabetical order. Let $[T]_{\alpha}^{\beta} = B$, $[U]_{\gamma}^{\beta} = A$. Then

$$(U \circ T)(v_j) = U(T(v_j)) = U\left(\sum_k b_{kj} w_k\right) = \sum_k b_{kj} U(w_k) =$$

$$\sum_k b_{kj} \left(\sum_i a_{ik} z_i\right) = \sum_i \left(\sum_k a_{ik} b_{kj}\right) z_i.$$

Consequently, if $C = [U \circ T]_{\alpha}^{\gamma}$, $c_{ij} = \sum_k a_{ik} b_{kj}$.

Definition 2.27. Let A be a $p \times m$ matrix and B an $m \times n$ matrix. Define the *matrix product* of A and B to be the $p \times n$ matrix given by

$$(AB)_{ij} = \sum_{k=1}^m a_{ik} b_{kj},$$

where $i = 1, \dots, p$, $j = 1, \dots, n$.

We have just established the following theorem.

Theorem 2.28. Let $T : V \rightarrow W$, $U : W \rightarrow Z$. Then $[U \circ T]_{\alpha}^{\gamma} = [U]_{\beta}^{\gamma} [T]_{\alpha}^{\beta}$.

Theorem 2.29. Let $A \in \text{Mat}_{m \times n}$, $B, C \in \text{Mat}_{n \times p}$, $D, E \in \text{Mat}_{q \times m}$. Then

- i. $A(B + C) = AB + AC$
- ii. $(D + E)A = DA + EA$
- iii. $a(AB) = A(aB) = (aA)B$
- iv. $I_m A = AI_n$

Theorem 2.30. Let $T : V \rightarrow W$ linear. Then $\forall v \in V$:

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

Proof. I find the proof in the book somewhat uninformative. Here is an alternative. By linearity, it is clear that we have to check the identity only on the elements of the basis $\beta = \{v_1, \dots, v_n\}$.

First, note that

$$[T(v_j)]_\gamma = [a_{1j}w_1 + \dots + a_{mj}w_m]_\gamma = (a_{1j}, \dots, a_{mj}).$$

On the other hand,

$$[T]_\beta^\gamma [v_j]_\beta = [T]_\beta^\gamma e_j = (a_{1j}, \dots, a_{mj}).$$

Done. □

Definition 2.31. Let $A \in \text{Mat}_{m \times n}(F)$. Define $L_A : F^n \rightarrow F^m$ by $L_A(x) = Ax$. L_A is the left-multiplication transformation given by the matrix A .

Theorem 2.32. Let β, γ be the standard ordered bases.

- i. $L_A : F^n \rightarrow F^m$ is linear.
- ii. $[L_A]_\beta^\gamma = A$
- iii. $L_A = L_B \Leftrightarrow A = B$
- iv. $L_{A+B} = L_A + L_B, L_{aA} = aL_A$
- v. For $T : F^n \rightarrow F^m, T = L_{[T]_\beta^\gamma}$

2.4 Invertibility and Isomorphisms

Definition 2.33. Let V, W be vector spaces. Let $T : V \rightarrow W$ linear. We define $U : W \rightarrow V$ to be the *inverse* of T if $T \circ U = \text{id}_W$ and $U \circ T = \text{id}_V$. If T has an inverse, T is called *invertible*.

Remark 2.34. If T has an inverse, the inverse is certainly unique. We write T^{-1} for it. Recall that in general, for sets A, B , a function $f : A \rightarrow B$ is invertible if and only if f is one-to-one and onto.

Theorem 2.35. *If $T : V \rightarrow W$ is linear and invertible, then the inverse T^{-1} is linear also.*

Proof. Let $w_1, w_2 \in W$, $a \in F$. Since T is in particular onto, for some v_1, v_2 the following holds:

$$\begin{aligned} T^{-1}(w_1 + w_2) &= T^{-1}(T(v_1) + T(v_2)) = T^{-1}(T(v_1 + v_2)) = \\ &v_1 + v_2 = T^{-1}(w_1) + T^{-1}(w_2). \end{aligned}$$

Moreover,

$$T^{-1}(aw_1) = T^{-1}(aT(v_1)) = T^{-1}(T(av_1)) = av_1 = aT^{-1}(w_1).$$

□

Lemma 2.36. *Let $T : V \rightarrow W$ linear and invertible. Let V, W be finite dimensional. Then*

$$\dim V = \dim W.$$

Proof. The dimension formula says $\text{rank}(T) = \dim V$ since $\text{nullity}(T) = 0$ due to T being one-to-one. But $\text{rank}(T) \leq \dim W$ always. Thus, $\dim V \leq \dim W$. By symmetry, we are done. □

Definition 2.37. Let A be an $n \times n$ matrix. Say A is *invertible* if there exists an $n \times n$ matrix B such that

$$AB = I_n = BA.$$

Remark 2.38. Such a B is unique, if it exists. Reason: Let B, C be two matrices with the above property. Then

$$C = CI_n = C(AB) = (CA)B = I_n B = B.$$

Thus, we can write A^{-1} for B .

Theorem 2.39. *Let V, W be finite dimensional vector spaces with ordered bases β, γ respectively. Let $T : V \rightarrow W$ linear. Then T is invertible if and only if $[T]_{\beta}^{\gamma}$ is an invertible matrix. Moreover, in this case, $[T^{-1}]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^{-1}$.*

Proof. Based on:

$$I_n = [\text{id}_V]_\beta^\beta = [T^{-1} \circ T]_\beta^\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma.$$

See the textbook for complete details. □

Definition 2.40. Let V, W be vector spaces. Say V is *isomorphic* to W if and only if there exists $T : V \rightarrow W$ linear and invertible. Such a T is called an *isomorphism* from V onto W .

Remark 2.41. Being isomorphic is an equivalence relation on the set of vector spaces (over a given field).

Theorem 2.42. *Let V, W be finite dimensional vector spaces (over the same field). Then V is isomorphic to W if and only if $\dim V = \dim W$.*

Proof. We have already seen \Rightarrow . So it remains to prove \Leftarrow . Let $\beta = \{v_1, \dots, v_n\}$ and $\gamma = \{w_1, \dots, w_n\}$. Saw earlier that $T(v_j) = w_j$ for all j defines a linear transformation $V \rightarrow W$. It is onto because $\text{Range}(T) = \text{span}\{T(v_1), \dots, T(v_n)\} = \text{span}\{w_1, \dots, w_n\} = W$. It is one-to-one because of the dimension formula. □

Corollary 2.43. *Let V be a vector space over F . Then V is isomorphic to F^n if and only if $\dim V = n$.*

3 Elementary Matrix Operations and Systems of Linear Equations

3.1 Elementary matrix operations and elementary matrices

Definition 3.1. Let A be an $m \times n$ matrix. An *elementary row operation* is any one of the following.

- i. interchanging any two rows of A
- ii. multiplying any row of A with a non-zero scalar
- iii. adding any scalar multiple of a row of A to another row.

Definition 3.2. An $n \times n$ *elementary matrix* E is obtained by performing an elementary operation on I_n . We say E is of type 1, 2, or 3 if the elementary operation was of that type according to Definition 3.1.

Theorem 3.3. *Let $A \in \text{Mat}_{m \times n}$. Let B be obtained from A by an elementary row operation corresponding to the elementary matrix E (of size $m \times m$). Then*

$$B = EA.$$

Proof. Direct verification in the three cases. □

Theorem 3.4. *Elementary matrices are invertible. The inverse of an elementary matrix is an elementary matrix of the same type.*

Proof. Explicit checking. □

3.2 The rank of a matrix and matrix inverses

Definition 3.5. Let $A \in \text{Mat}_{m \times n}(F)$. We define the *rank* of A , denoted by $\text{rank } A$, to be the rank of the linear transformation $L_A : F^n \rightarrow F^m$ given by A .

Theorem 3.6. *Let $A \in \text{Mat}_{m \times n}(F)$. If $P \in \text{Mat}_{m \times m}(F)$ and $Q \in \text{Mat}_{n \times n}(F)$ are invertible, then*

i. $\text{rank}(AQ) = \text{rank}(A)$

ii. $\text{rank}(PA) = \text{rank}(A)$

iii. $\text{rank}(PAQ) = \text{rank}(A)$

Proof.

$$\text{range}(L_{AQ}) = \text{range}(L_A \circ L_Q) = (L_A \circ L_Q)(F^n) = L_A(F^n)$$

where the last equality holds because L_Q is onto (it is actually an isomorphism). Thus,

$$\text{range}(L_{AQ}) = \text{range}(L_A).$$

In particular, the ranks agree.

The two remaining statements are left as exercises. □

Corollary 3.7. *Elementary row operations do not change the rank of a matrix.*

Theorem 3.8. *The rank of any matrix A equals the maximum number of linearly independent **columns**, i.e., the dimension of the subspace generated by the columns.*

Proof. Let β be the standard ordered basis for F^n . Then

$$\text{range}(L_A) = \text{span}(L_A(\beta)) = \text{span}\{L_A(e_1), \dots, L_A(e_n)\}.$$

Note that $L_A(e_i)$ simply is the i -th column of A . □

The following sums up the discussion of Theorem 3.6, Corollary 1 and Corollary 2 in the textbook.

By elementary row and column operations, any $m \times n$ matrix can be transformed to an $m \times n$ matrix of the form

$$D = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

where I_r is the identity matrix of size $r \times r$. More precisely, there exist elementary matrices E_1, \dots, E_p and G_1, \dots, G_q such that

$$D = E_p \dots E_1 A G_1 \dots G_q.$$

Then

$$\text{rank } A = \text{rank } D = \text{rank } D^t = \text{rank } A^t,$$

where the last inequality is due to $D^t = G_q^t \dots G_1^t A E_1^t \dots E_p^t$ and the fact that the transpose of an elementary matrix is an elementary matrix.

We have just proven that

$$\text{rank } A = \text{rank } A^t.$$

Definition 3.9. Let the *row rank* of a matrix denote the maximum number of linearly independent rows.

Since the row rank of a matrix is clearly equal to the rank of its transpose, we have proven that the row rank and the rank of any matrix agree.

Quiz 6:

(10 points) Find a 3×3 matrix E such that

$$E \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 0 & 0 \\ 4 & 4 & 4 \end{pmatrix}.$$

Explain the steps you took to find E .

Solution:

The matrix on the right hand side is obtained from the matrix on the left-hand side by successively performing the following row operations:

- i. Interchange the second and third row.
- ii. Add the first row to the third
- iii. Multiply the first row by 2.

The corresponding elementary matrices are

i. $E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

ii. $E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

iii. $E_3 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Then

$$E = E_3 E_2 E_1 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad \square$$

We continue with the discussion of the rank of a matrix. What we have discussed yields the following corollary.

Corollary 3.10. *The rank of a matrix A can be found by reducing A to echelon form via row operations and counting the number of non-zero rows.*

Here is an example. To find the rank of $A =$

$$\begin{pmatrix} 0 & 2 & 4 & 2 & 2 \\ 4 & 4 & 4 & 8 & 0 \\ 8 & 2 & 0 & 10 & 2 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix},$$

we simply reduce it, via elementary row operations to echelon form:

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 & 1 \\ 4 & 1 & 0 & 5 & 1 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & -3 & -4 & -3 & 1 \\ 0 & -3 & -4 & -3 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The number of non-zero rows is 3, which means that the rank of A is 3.

Finally, a remark on the method of computing the inverse of a matrix.

Let A be an $n \times n$ invertible matrix. If we reduce A by elementary row operations to the identity, then we are effectively executing the following multiplication of A with elementary matrices:

$$E_p \dots E_1 A = I_n.$$

Observe that obviously $E_p \dots E_1 = A^{-1}$. Writing trivially $A^{-1} = E_p \dots E_1 = E_p \dots E_1 I_n$, we conclude that the inverse of A can be obtained by applying the same sequence of elementary row operations to the identity. This method is commonly taught in an introductory course to Linear Algebra, but we now have a very nice justification for it.

3.3 Systems of linear equations – theoretical aspects

A general system of m linear equations in n variables is

$$a_{11}x_1 + \dots + a_{1n}x_n = b_1, \dots, a_{m1}x_1 + \dots + a_{mn}x_n = b_m.$$

This can be rewritten as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

or

$$Ax = b.$$

Definition 3.11. A solution of the above system is $s = (s_1, \dots, s_n) \in F^n$ such that $As = b$. The system is called *consistent* if there exists a solution. The system is called *homogeneous* if and only if $b = \vec{0}$.

Theorem 3.12. *The set of solutions of $Ax = \vec{0}$ is the null space of L_A .*

Proof. Clear! □

Corollary 3.13. *If $m < n$, the system $Ax = \vec{0}$ has a non-zero solution.*

Proof. The dimension formula yields

$$\text{nullity}(L_A) = n - \text{rank}(L_A) \geq n - m > 0.$$

□

Theorem 3.14. *Let K be the set of solutions of $Ax = b$ and K_H be the set of solutions of $Ax = 0$. Then for any s with $As = b$,*

$$K = \{s\} + K_H = \{s + k : k \in K_H\}.$$

Proof. “ \supseteq ”. Let $t \in K_H$. Then

$$A(s + t) = As + At = As + \vec{0} = As = b.$$

“ \subseteq ”. Let $t \in K$. Let $w = t - s$. Then

$$Aw = A(t - s) = At - As = b - b = 0.$$

Thus, $w \in K_H$ and $t = s + w \in \{s\} + K_H$. □

Example 3.15. Let us consider the single inhomogeneous equation in three variables $x_1 - 2x_2 + x_3 = 4$. An obvious solution is $s = (4, 0, 0)$. Moreover, the solution set of the homogeneous equation $x_1 - 2x_2 + x_3 = 0$ is

$$K_H = \{(2x_2 - x_3, x_2, x_3) : x_2, x_3 \in \mathbb{R}\},$$

which can be rewritten as

$$\{x_2(2, 1, 0) + x_3(-1, 0, 1) : x_2, x_3 \in \mathbb{R}\}.$$

Theorem 3.14 now says that

$$K = \{(4, 0, 0) + x_2(2, 1, 0) + x_3(-1, 0, 1) : x_2, x_3 \in \mathbb{R}\}$$

(which we could have found directly, of course).

Theorem 3.16. *Let $Ax = b$ be a system of n equations in n variables. Then A is invertible if and only if the system has exactly one solution.*

Proof. \Rightarrow . Let us multiply the equation from the left with A^{-1} . Then it becomes

$$A^{-1}Ax = x = A^{-1}b.$$

Thus, $x = A^{-1}b$ is the unique solution.

\Leftarrow . We saw: $K = \{s\} + K_H$, where S is a solution. Since the solution is unique, we can infer $K_H = \{\vec{0}\}$. Thus, the null space of L_A is $\{\vec{0}\}$, and A is invertible. \square

Example 3.17. Let

$$A = \begin{pmatrix} 0 & -1 \\ 2 & 4 \end{pmatrix}.$$

Then

$$A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

is equivalent to

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A^{-1} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 & \frac{1}{2} \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}.$$

Theorem 3.18. *The system $Ax = b$ is consistent if and only if $\text{rank } A = \text{rank}(A|b)$, where $(A|b)$ is the augmented matrix*

$$\begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{pmatrix}.$$

Proof. The system is consistent if and only if $b \in R(L_A)$, which means $b \in \text{span}\{L_A(e_1), \dots, L_A(e_n)\}$. This in turn is equivalent to

$$\text{span}\{L_A(e_1), \dots, L_A(e_n)\} = \text{span}\{L_A(e_1), \dots, L_A(e_n), b\},$$

which is equivalent to $\text{rank } A = \text{rank}(A|b)$. □

Example 3.19. Let's consider the system $x_1 + 2x_2 = 1$, $2x_1 + 4x_2 = 0$. It is not consistent because

$$\text{rank} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = 1,$$

but

$$\text{rank} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \end{pmatrix} = 2.$$

4 Determinants

4.1 Determinants of 2×2 matrices

Definition 4.1. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix over some field F . Then we define the *determinant* of A (also $\det A$ or $|A|$) to be the scalar $ad - bc$.

Example 4.2. $\det \begin{pmatrix} 5 & 0 \\ 1 & 2 \end{pmatrix} = 5 \cdot 2 - 1 \cdot 0 = 10$, $\det \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = 1 \cdot 4 - 2 \cdot 2 = 0$.

The following is an important observation:

$$\det \left(\begin{pmatrix} 5 & 0 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right) = \det \begin{pmatrix} 6 & 2 \\ 4 & 6 \end{pmatrix} = 36 - 8 = 28.$$

On the other hand,

$$\det \begin{pmatrix} 5 & 0 \\ 1 & 2 \end{pmatrix} + \det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 10 - 2 = 8.$$

The above clearly shows that the determinant is not linear, which we state as the following remark.

Remark 4.3. $\det : M_{n \times n}(F) \rightarrow F$ is not a linear functional.

Instead, the right kind of linearity is the following.

Theorem 4.4. $\det : M_{n \times n}(F) \rightarrow F$ is a linear function of each row when the other row is held fixed, i.e., for the first row, we have

$$\det \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} & a_{22} \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \det \begin{pmatrix} b_{11} & b_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

and

$$\det \begin{pmatrix} ca_{11} & ca_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The statements for the second row are analogous.

Proof. Expand all expressions according to the definition and compare both sides of the equations. \square

Theorem 4.5. Let $A \in M_{2 \times 2}(F)$. Then

$$\det A \neq 0 \Leftrightarrow A \text{ is invertible.}$$

Proof. “ \Rightarrow ” Since $\det A \neq 0$, the matrix $B = \frac{1}{\det A} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$ is well-defined. Since $AB = I_2 = BA$ by explicit computations, it is clear that A is invertible and its inverse is B .

“ \Leftarrow ” Since A is invertible, its rank is two. Thus, at least one of a_{11}, a_{21} must be not equal to zero. Without loss of generality, let's assume it is a_{11} . If we perform on A the elementary row operation of adding $\frac{-a_{21}}{a_{11}}$ times the first row to the second, we obtain the matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} \end{pmatrix}.$$

We know that elementary row operations do not change the rank, so this matrix still has rank two, which implies

$$a_{22} - \frac{a_{12}a_{21}}{a_{11}} \neq 0.$$

Multiplying this relation by a_{11} yields

$$\det A = a_{11}a_{22} - a_{12}a_{21} \neq 0.$$

□

4.2 Determinants of order n

We now define the determinant of a matrix $A \in M_{n \times n}$ for $n \geq 3$.

Remark 4.6. In class and in these notes so far, our convention was to refer to the entries of matrices with lower case letters, e.g., as a_{ij} . The textbook however uses upper case letters, e.g., A_{ij} . We follow the convention of the textbook in this section to make going back and forth between these notes and the textbook easier. Be sure not to confuse A_{ij} with \tilde{A}_{ij} as defined below.

Definition 4.7. Let $A \in M_{n \times n}$. For fixed $i, j \in \{1, \dots, n\}$ denote by \tilde{A}_{ij} the $(n-1) \times (n-1)$ matrix obtained by deleting the i -th row and j -th column of A .

Example 4.8. Let

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & \pi \\ \sqrt{2} & 1 & -1 \end{pmatrix}.$$

Then

$$\tilde{A}_{11} = \begin{pmatrix} 3 & \pi \\ 1 & -1 \end{pmatrix}, \quad \tilde{A}_{12} = \begin{pmatrix} 1 & \pi \\ \sqrt{2} & -1 \end{pmatrix}, \quad \tilde{A}_{32} = \begin{pmatrix} 2 & 0 \\ 1 & \pi \end{pmatrix}.$$

Definition 4.9. Let $A \in M_{n \times n}$. If $n = 1$, then let $\det A = A_{11}$.

If $n \geq 2$, define $\det A$ recursively as

$$\det A = \sum_{j=1}^n (-1)^{1+j} A_{1j} \det(\tilde{A}_{1j}).$$

Definition 4.10. The determinant $\det \tilde{A}_{ij}$ is called the (i, j) -minor. The scalar $c_{ij} = (-1)^{i+j} \det(\tilde{A}_{ij})$ is called the (i, j) -cofactor.

With the above definitions the definition of the determinant can be restated as

$$\det A = \sum_{j=1}^n c_{1j} A_{1j}.$$

For obvious reasons, this is also referred to as the cofactor expansion of the determinant along the first row. Here are some examples:

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $\det A = (-1)^2 a \det(d) + (-1)^3 b \det(c) = ad - bc$, as defined in the previous subsection.

Let $A = \begin{pmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{pmatrix}$. Then $\det A =$

$$(-1)^2 \cdot 1 \cdot \det \begin{pmatrix} -5 & 2 \\ 4 & -6 \end{pmatrix} + (-1)^3 \cdot 3 \cdot \det \begin{pmatrix} -3 & 2 \\ -4 & -6 \end{pmatrix} + (-1)^4 \cdot (-3) \cdot \det \begin{pmatrix} -3 & -5 \\ -4 & 4 \end{pmatrix} = 40.$$

Let $A = \begin{pmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 4 & -4 & 4 \end{pmatrix}$. Then $\det A =$

$$0 + (-1)^3 \cdot 1 \cdot \det \begin{pmatrix} -2 & -5 \\ 4 & 4 \end{pmatrix} + (-1)^4 \cdot 3 \cdot \det \begin{pmatrix} -2 & -3 \\ 4 & -4 \end{pmatrix} = 48.$$

Let $A = \begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & -3 \\ -2 & -3 & -5 & 2 \\ 4 & -4 & 4 & -6 \end{pmatrix}$. Then $\det A =$

$$(-1)^2 \cdot 2 \cdot \det \begin{pmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{pmatrix} + (-1)^5 \cdot 1 \cdot \det \begin{pmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 4 & -4 & 4 \end{pmatrix} = 32.$$

Theorem 4.11. *The determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed.*

Proof. Write the matrix in question as

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{r1} + kb_{r1} & \cdots & a_{rn} + kb_{rn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Also, let

$$B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{r1} & \cdots & a_{rn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

and

$$C = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ b_{r1} & \cdots & b_{rn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

The theorem is immediate from the definition of the determinant if the row in question is the first row, i.e., $r = 1$. So we may assume that the row in question is not the first row, i.e., $r \geq 2$. By induction, we know that

$$\det \tilde{A}_{1j} = \det \tilde{B}_{1j} + k \det \tilde{C}_{1j},$$

because $\tilde{A}_{1j}, \tilde{B}_{1j}, \tilde{C}_{1j}$ are matrices of size $(n-1) \times (n-1)$ and have the exact form necessary to apply the theorem. Thus,

$$\det A = \sum_{j=1}^n (-1)^{1+j} A_{1j} \det(\tilde{A}_{1j}) = \sum_{j=1}^n (-1)^{1+j} A_{1j} (\det(\tilde{B}_{1j}) + k \det(\tilde{C}_{1j})).$$

Since clearly

$$A_{1j} = B_{1j} = C_{1j},$$

the above equals

$$\sum_{j=1}^n (-1)^{1+j} B_{1j} \det(\tilde{B}_{1j}) + \sum_{j=1}^n (-1)^{1+j} C_{1j} k \det(\tilde{C}_{1j}) = \det B + k \det C.$$

□

Corollary 4.12. *If A has a row of zeroes, then its determinant is zero.*

Proof. Let it be the r -th row of A which is zero. Let B be a matrix which agrees with A outside of the r -th row. Then by the preceding theorem,

$$\det B = \det B + \det A.$$

Thus, $\det A = 0$. □

Next, we establish that it is not necessary to use the first row for the expansion to compute the determinant: any row will work.

Theorem 4.13. *The determinant of a square matrix $A \in Mat_{n \times n}$ can be found by cofactor expansion along any row, i.e., for all $i = 1, \dots, n$,*

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

Proof. The case $i = 1$ is the definition, so we may assume $i \geq 2$. By linearity with respect to the i -th row, we obtain

$$\det A = a_{i1} \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \dots + a_{in} \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

The next step in the proof uses the Lemma on pages 213-214 in the textbook. We do not give its proof here, because a detailed rigorous proof (filling the entire page 214) is given in the textbook. In essence, the lemma says that the above is equal to

$$a_{i1}(-1)^{i+1} \det \tilde{A}_{i1} + \dots + a_{in}(-1)^{i+n} \det \tilde{A}_{in},$$

which is what we intended to prove. □

Quiz 7:

- (5 points) For an unknown real parameter t , let

$$A = \begin{pmatrix} 1 & t & t^2 \\ t & 1 & t \\ t^2 & t & 1 \end{pmatrix}.$$

Compute $\det A$. For which values of t is $\det A = 0$?

2. (5 points) Let $B \in \text{Mat}_{n \times n}(\mathbb{R})$. What is the condition for $\det(B) = \det(-B)$? Justify your answer carefully. Hint: Be sure to consider all possible cases.

Solution:

1. An expansion along the first row yields

$$\det A = 1 \cdot \begin{pmatrix} 1 & t \\ t & 1 \end{pmatrix} - t \begin{pmatrix} t & t \\ t^2 & 1 \end{pmatrix} + t^2 \begin{pmatrix} t & 1 \\ t^2 & t \end{pmatrix} = \dots = (1 - t^2)^2.$$

Thus, $\det A = 0$ if and only if $t = \pm 1$.

2. The key observation here is that $\det(cA) = c^n \det A$ by an n -fold application of the linearity property of determinants with respect to rows of matrices. Therefore, $\det(-B) = (-1)^n \det B$ and the problem asks when this is equal to $\det B$. In case $\det B = 0$, the equality always holds, i.e., there is no condition on n . If $\det B \neq 0$, then equality holds if and only if n is even.

Corollary 4.14. *If $A \in \text{Mat}_{n \times n}$ has two identical rows, then $\det A = 0$.*

Proof. We proceed by induction. Let $n = 2$. Then clearly

$$\det A = \det \begin{pmatrix} a & b \\ a & b \end{pmatrix} = 0.$$

In the case of general $n \geq 3$, let the two identical rows be rows r and s . Let $i \neq r$, $i \neq s$. Then we compute $\det A$ by an expansion along the i -th row:

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

However, the matrices \tilde{A}_{ij} also have two identical rows (since we are deleting a third row), so by induction their determinants are 0. Thus, $\det A$ is also zero. \square

Next, we investigate the influence of elementary row operations on the determinant.

Theorem 4.15. *Let $A \in \text{Mat}_{n \times n}$. If B is obtained from A by interchanging two rows, then $\det B = -\det A$.*

Proof. Write

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix}, \quad B = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix},$$

with the a_i representing rows. Next, observe that

$$0 = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r + a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix}$$

due to Corollary 4.14 and linearity. We can expand this further to

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix}.$$

Due to Corollary 4.14, the two outer determinants are zero, which proves our claim. \square

Theorem 4.16. *Let $A \in \text{Mat}_{n \times n}$ and let B be obtained from A by adding a multiple of one row to another row of A . Then $\det B = \det A$.*

Proof.

$$\det B = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ ka_r + a_s \\ \vdots \\ a_n \end{pmatrix} = k \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix}.$$

Due to Corollary 4.14, the determinant of the matrix with the two identical rows is zero, which proves our claim. \square

Corollary 4.17. *Let $A \in \text{Mat}_{n \times n}$. If A has rank less than n then $\det A = 0$.*

Proof. We have seen that performing an elementary row operation on a matrix may change the value of its determinant. However, it will not change whether the determinant is zero or not. Now, if we transform A into echelon form, the fact that A has rank less than n means that the echelon form will contain a row of zeroes. Thus, the determinant of the matrix in echelon form is equal to zero due to Corollary 4.12. This means that the determinant of A was zero to begin with. \square

Quiz 8:

1. (5 points) Over the real numbers, find the value of k that satisfies the following equation:

$$\det \begin{pmatrix} 2a_1 & 2a_2 & 2a_3 \\ 3b_1 + 5c_1 & 3b_2 + 5c_2 & 3b_3 + 5c_3 \\ 4c_1 & 4c_2 & 4c_3 \end{pmatrix} = k \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

Justify your answer carefully.

2. (5 points) Compute

$$\det \begin{pmatrix} 3 & 3 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & -1 & 1 & 1 \\ 0 & 2 & -2 & 2 \end{pmatrix}.$$

You **MUST** use the method of row reductions.

Solution:

1.

$$\det \begin{pmatrix} 2a_1 & 2a_2 & 2a_3 \\ 3b_1 + 5c_1 & 3b_2 + 5c_2 & 3b_3 + 5c_3 \\ 4c_1 & 4c_2 & 4c_3 \end{pmatrix} = \det \begin{pmatrix} 2a_1 & 2a_2 & 2a_3 \\ 3b_1 & 3b_2 & 3b_3 \\ 4c_1 & 4c_2 & 4c_3 \end{pmatrix}$$

because adding $-\frac{5}{4}$ times the third row to the second one does not change the determinant. Moreover,

$$\det \begin{pmatrix} 2a_1 & 2a_2 & 2a_3 \\ 3b_1 & 3b_2 & 3b_3 \\ 4c_1 & 4c_2 & 4c_3 \end{pmatrix} = 2 \cdot 3 \cdot 4 \cdot \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

Thus, $k = 24$.

2. We use elementary row operations to reduce the matrix to upper triangular form. In each step, note the effect on the factor in front of the determinant. At the end, we use that the determinant of a triangular matrix is simply the product of the diagonal entries.

$$\begin{aligned} & \det \begin{pmatrix} 3 & 3 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & -1 & 1 & 1 \\ 0 & 2 & -2 & 2 \end{pmatrix} \\ &= -\det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 3 & 3 & 2 & 1 \\ 1 & -1 & 1 & 1 \\ 0 & 2 & -2 & 2 \end{pmatrix} \\ &= -\det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & -3 & 2 & -2 \\ 0 & -3 & 1 & 0 \\ 0 & 2 & -2 & 2 \end{pmatrix} \\ &= 2 \det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & -3 & 1 & 0 \\ 0 & -3 & 2 & -2 \end{pmatrix} \\ &= 2 \det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & -2 & 3 \\ 0 & 0 & -1 & 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= -2 \det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -2 & 3 \end{pmatrix} \\
&= -2 \det \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= (-2)(-1) = 2.
\end{aligned}$$

4.3 Properties of determinants

It is clear from our previous discussion that

$$\det(\text{elementary matrix which interchanges two rows}) = -1,$$

$$\det(\text{elementary matrix which adds a multiple of one row to another}) = 1,$$

$$\det(\text{elementary matrix which multiplies a given row by } k) = k.$$

While we saw earlier that matrix addition is not compatible with taking the determinant, we now show that matrix multiplication is.

Theorem 4.18. *Let $A, B \in \text{Mat}_{n \times n}$. Then*

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

Proof. We saw earlier that $\det(EC) = \det(E)\det(C)$ for any $C \in \text{Mat}_{n \times n}$ when E is an elementary matrix. We may assume that $\text{rank } A = \text{rank } B = n$, because otherwise both sides are clearly equal to zero. As we did earlier, we write A as a product of elementary matrices $A = E_1 \dots E_p$. Now

$$\begin{aligned}
\det(AB) &= \det(E_1 \dots E_p B) \\
&= \det(E_1) \det(E_2 \dots E_p B) \\
&= \det(E_1) \dots \det(E_p) \det(B) \\
&= \det(E_1) \dots \det(E_{p-1} E_p) \det(B) \\
&= \det(E_1 \dots E_p) \det(B) \\
&= \det A \cdot \det B
\end{aligned}$$

□

Corollary 4.19. *Let $A \in \text{Mat}_{n \times n}$. Then A is invertible if and only if $\det A \neq 0$. Moreover, if $\det A \neq 0$, then $\det A^{-1} = \frac{1}{\det A}$.*

Proof. We already gave an argument for “ \Leftarrow ” in the previous section. We now prove “ \Rightarrow ”, for which we just observe that

$$1 = \det I_n = \det(AA^{-1}) = \det A \cdot \det A^{-1},$$

which implies that $\det A$ cannot be zero. Moreover, $\det A^{-1} = \frac{1}{\det A}$ is also clear now. \square

Theorem 4.20. *Let $A \in \text{Mat}_{n \times n}$. Then $\det A = \det A^T$.*

Proof.

$$\begin{aligned} \det(A^T) &= \det((E_1 \dots E_p)^T) \\ &= \det(E_p^T \dots E_1^T) \\ &= \det(E_p^T) \dots \det(E_1^T) \\ &= \det(E_1^T) \dots \det(E_p^T) \\ &= \det(E_1) \dots \det(E_p) \\ &= \det(E_1 \dots E_p) \\ &= \det(A). \end{aligned}$$

Note that the fourth equal sign is simply due to the commutativity of the reals. The equality $\det E_i^T = \det E_i$ is obvious for all three types of elementary matrices. \square

Theorem 4.21 (Cramer’s Rule). *Let $A \in \text{Mat}_{n \times n}$ be an invertible matrix. Let $x = (x_1, \dots, x_n)$ be the unique solution of $Ax = b$. Let M_k be obtained from A by replacing the k -th column of A by b . Then*

$$x_k = \frac{\det M_k}{\det A}.$$

Proof. Let X_k be obtained from I_n by replacing the k -th column by the vector x . The expansion of X_k along the k -th row shows that $\det X_k = x_k$. Moreover, it is straightforward to check that

$$A \cdot X_k = M_k.$$

Thus,

$$\det M_k = \det(A \cdot X_k) = \det A \cdot \det X_k = (\det A) \cdot x_k,$$

which yields

$$x_k = \frac{\det M_k}{\det A}.$$

□

5 Diagonalization

5.1 Eigenvalues and eigenvectors

Motivation: Let $T : V \rightarrow V$ be a linear transformation. It may happen that a certain nonzero vector v gets mapped to a scalar multiple of itself, i.e., $T(v) = \lambda v$ for some $\lambda \in F$. Such vectors are clearly special, and we now study these types of pairs of scalars and vectors (v, λ) and related questions.

Definition 5.1. A linear operator T on a finite dimensional vector space V is called diagonalizable if there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. Moreover, we call a square matrix diagonalizable if $L_A : F^n \rightarrow F^n, v \mapsto Av$ is diagonalizable.

Definition 5.2. Let T be a linear operator on the vector space V . A nonzero $v \in V$ is called *eigenvector* of T if

$$T(v) = \lambda v$$

for some scalar λ in F . The scalar λ is called the corresponding *eigenvalue*. Moreover, for $A \in \text{Mat}_{n \times n}(F)$, the nonzero vector $v \in F^n$ is an *eigenvector* of A if and only if $L_A(v) = \lambda v$, i.e., $Av = \lambda v$, for some scalar $\lambda \in F$, which we again call the corresponding *eigenvalue*.

Theorem 5.3. Let T be a linear operator on V . Then T is diagonalizable if and only if there exists an ordered basis $\beta = \{v_1, \dots, v_n\}$ of eigenvectors.

Proof. “ \Rightarrow ” It is clear that

$$[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & 0 \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix},$$

which is a diagonal matrix, so T is diagonalizable.

“ \Leftarrow ” Let $\beta = \{v_1, \dots, v_n\}$ be the ordered basis for which $[T]_\beta$ is diagonal. Then

$$[T]_\beta[v_i]_\beta = \lambda_i e_i = \lambda_i [v_i]_\beta.$$

Thus, $T(v_i) = \lambda_i v_i$, which implies that v_i is an eigenvector. \square

Example 5.4. Let

$$A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Then $Av_1 = (-2, 2) = (-2)(1, -1)$ and $Av_2 = (15, 20) = 5(3, 4)$. Thus, $\beta = \{(1, -1), (3, 4)\}$ and

$$[L_A]_\beta = \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix}.$$

Remark 5.5. Linear operators and matrices do not necessarily have eigenvectors. For example, the rotation by 90 degrees in the plane clearly has none.

Example 5.6. Let $T : C^\infty(\mathbb{R}, \mathbb{R}) \rightarrow C^\infty(\mathbb{R}, \mathbb{R})$ be given by $f \mapsto \frac{df}{dx}$. Then $T(f) = \lambda f$ is equivalent to $\frac{df}{dx} = \lambda f$, which is solved by $f(x) = ce^{\lambda x}$.

Theorem 5.7. Let $A \in \text{Mat}_{n \times n}(F)$. Then $\lambda \in F$ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. For a nonzero vector $v \in V$, the following are equivalent:

$$\begin{aligned} Av &= \lambda v \\ \Leftrightarrow Av - \lambda v &= \vec{0} \\ \Leftrightarrow Av - \lambda I_n v &= \vec{0} \\ \Leftrightarrow (A - \lambda I_n)v &= \vec{0} \\ \Leftrightarrow \vec{0} \neq v &\in \ker(A - \lambda I_n) \end{aligned}$$

Thus, $\det(A - \lambda I_n) = 0$, and conversely, $\det(A - \lambda I_n) = 0$ implies the existence of a vector v with $Av = \lambda v$. \square

Definition 5.8. Let $A \in \text{Mat}_{n \times n}(F)$. The polynomial $f(t) = \det(A - tI_n)$ is called the characteristic polynomial of A .

Remark 5.9. Theorem 5.7 says that the set of eigenvalues of A equals the set of roots of the characteristic polynomial.

Let us find the eigenvalues of $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$.

$$\det(A - tI_2) = \det \begin{pmatrix} 1-t & 1 \\ 4 & 1-t \end{pmatrix} = (1-t)^2 - 4 = t^2 - 2t - 3.$$

The roots of this polynomial are $\lambda = 3$ or $\lambda = -1$.

Let us find the eigenvalues of $A = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}$.

$$\det(A - tI_3) = \det \begin{pmatrix} 5-t & 8 & 16 \\ 4 & 1-t & 8 \\ -4 & -4 & -11-t \end{pmatrix} = \det \begin{pmatrix} 5-t & 8 & 16 \\ 4 & 1-t & 8 \\ 0 & -3-t & -3-t \end{pmatrix}.$$

Extracting the common factor from the third row and then doing an expansion along the same row yields:

$$(-3-t) \left(-\det \begin{pmatrix} 5-t & 16 \\ 4 & 8 \end{pmatrix} + \det \begin{pmatrix} 5-t & 8 \\ 4 & 1-t \end{pmatrix} \right) = -(t+3)(t^2 + 2t - 3).$$

This is a third degree polynomial which can be factored into $-(t+3)^2(t-1)$, which means that the eigenvalues of the matrix A are $\lambda = 1$ and $\lambda = -3$, the latter with multiplicity two.

Quiz 8:

1. (5 points) Find the eigenvalues of $A = \begin{pmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{pmatrix}$.

2. (5 points) Let V be a vector space and $T : V \rightarrow V$ a linear transformation with pairwise distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Let v_1 be an eigenvector for λ_1 , v_2 an eigenvector for λ_2 and v_3 an eigenvector for λ_3 . Prove that v_1, v_2, v_3 are linearly independent.

Hint: Make the Ansatz $a_1v_1 + a_2v_2 + a_3v_3 = 0$ and apply the linear transformations $T - \lambda_1 \cdot \text{Id}, T - \lambda_2 \cdot \text{Id}, T - \lambda_3 \cdot \text{Id}$.

Solution:

1. Let us find the eigenvalues of $A = \begin{pmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{pmatrix}$ as the roots of the characteristic polynomial:

$$\det(A - tI_3) = \det \begin{pmatrix} -t & -2 & -3 \\ -1 & 1-t & -1 \\ 2 & 2 & 5-t \end{pmatrix} = -\det \begin{pmatrix} -1 & 1-t & -1 \\ -t & -2 & -3 \\ 2 & 2 & 5-t \end{pmatrix}.$$

In the above, as a first step, we have moved $-t$ out of the top left hand corner by interchanging rows 1 and 2. Under obvious row operations, this becomes

$$\det \begin{pmatrix} 1 & t-1 & 1 \\ 0 & t^2-t-2 & t-3 \\ 0 & -2t+4 & 3-t \end{pmatrix}.$$

This determinant can be expanded along the first row to obtain

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

Thus, the eigenvalues of A are 1, 2, 3.

2. Start with the Ansatz $a_1v_1 + a_2v_2 + a_3v_3 = \vec{0}$. Apply the linear transformations $T - \lambda_1 \cdot \text{Id}$ to get

$$a_1(\lambda_1 - \lambda_1)v_1 + a_2(\lambda_2 - \lambda_1)v_2 + a_3(\lambda_3 - \lambda_1)v_3 = a_2(\lambda_2 - \lambda_1)v_2 + a_3(\lambda_3 - \lambda_1)v_3 = \vec{0}.$$

Note that $\lambda_2 - \lambda_1 \neq 0 \neq \lambda_3 - \lambda_1$, as $\lambda_1, \lambda_2, \lambda_3$ are pairwise distinct. Now, apply $T - \lambda_2 \cdot \text{Id}$ to get

$$a_2(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_2)v_2 + a_3(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)v_3 = a_3(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)v_3 = \vec{0}.$$

Since $\lambda_3 - \lambda_1 \neq 0 \neq \lambda_3 - \lambda_2$, we must have $a_3 = 0$. By symmetry, we also get $a_1 = a_2 = 0$, q.e.d.

5.2 Diagonalizability

A slight generalization of Quiz 9, Problem 2 is the following:

Theorem 5.10. *Let $T : V \rightarrow V$ linear operator, let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of T . If v_1, \dots, v_k are eigenvectors corresponding to $\lambda_1, \dots, \lambda_k$ respectively, then v_1, \dots, v_k are linearly independent.*

Proof. We proceed by induction on k . The case $k = 1$ is trivial. Now, we assume that the statement is true for $k - 1$ and show that it also holds for k . To this end, we write

$$a_1 v_1 + \dots + a_k v_k = \vec{0}$$

and have to conclude that $a_1 = \dots = a_k = 0$. We apply the linear operator $T - \lambda_k \text{id}$ to this equation, which yields

$$a_1(\lambda_1 - \lambda_k)v_1 + \dots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} + a_k(0)v_k = \vec{0},$$

i.e.,

$$a_1(\lambda_1 - \lambda_k)v_1 + \dots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = \vec{0}.$$

Due to the induction assumption, v_1, \dots, v_{k-1} are linearly independent, so

$$a_i(\lambda_i - \lambda_k) = 0$$

for $i = 1, \dots, k - 1$. Since all eigenvalues are distinct, $\lambda_i - \lambda_k \neq 0$ when $i \neq k$. Thus, $a_i = 0$ for $i = 1, \dots, k - 1$. Going back to the first displayed equation in this proof, we finally obtain $a_k = 0$. \square

Corollary 5.11. *Let $T : V \rightarrow V$ linear. If T has $n = \dim V$ distinct eigenvalues, then T is diagonalizable.*

Proof. Let v_1, \dots, v_n be eigenvectors corresponding to the distinct eigenvalues, respectively. The previous theorem says that v_1, \dots, v_n are linearly independent. Since $n = \dim V$, this is actually a basis for V . Thus, we have found a basis of eigenvectors of T for V , which means T is diagonalizable. \square

Good luck for the Final Exam!

THE END