Boston College

Department of Mathematics
Professor William Keane

MATH2211 Honors Linear Algebra
Course Notes

Spring 2019

Contents

CHAPTER I

Preliminaries

1. Sets

Definition 1.1. A set is a collection of objects, called the elements of the set. If x is an element of the set A, we write $x \in A$. The cardinality of the set A, denoted $|A|$, is the number of elements of A, which may be finite or infinite.

Notation 1.2. We describe sets in three different ways.

(1) List the elements. For example, If $A = \{-3, 2, 4\}$, then $-3 \in A$ and $7 \notin A$. For large or infinite sets, we may resort to indicating the list, as with $B = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$, the set of even integers.

(2) Use a predicate, or a condition for membership. For example, if B is as above, then

$$C = \{x \in B : x > 0\} = \{2, 4, 6, \ldots\}$$

and

$$D = \{x \in B : x \text{ is prime}\} = \{2\}.$$

(3) Denote an important and commonly used set with a special symbol. Some examples are:
   - $\emptyset$ is the empty set, the set with no elements.
   - $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the set of integers.
   - $\mathbb{Z}^+ = \{1, 2, \ldots\}$, the set of positive integers.
   - $\mathbb{Q} = \{\frac{n}{m} : n, m \in \mathbb{Z}, m \neq 0\}$, the set of rational numbers.
   - $\mathbb{R} = \{x : x = \lim_{n\to\infty} r_n, \text{ for some sequence } r_n \in \mathbb{Q}\}$, the set of real numbers.
   - $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}, i^2 = -1\}$, the set of complex numbers.

Definition 1.3. Let A and B be sets. Then the union of A and B is

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

and the intersection of A and B is

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

More generally, if $A_1, A_2, \ldots, A_n$ are sets,

$$\bigcup_{i=1}^{n} A_i = \{x : x \in A_i, \text{ for some } i = 1, 2, \ldots n\}$$

and

$$\bigcap_{i=1}^{n} A_i = \{x : x \in A_i, \text{ for all } i = 1, 2, \dots n\}.$$

Even more generally, if A_i is a set for each $i \in I$,

$$\bigcup_{i \in I} A_i = \{x : x \in A_i, \text{ for some } i \in I\}$$

and

$$\bigcap_{i \in I} A_i = \{x : x \in A_i, \text{ for all } i \in I\}.$$

If in this last case $I = \mathbb{Z}^+$, we write

$$\bigcup_{i=1}^{\infty} A_i \text{ and } \bigcap_{i=1}^{\infty} A_i.$$

Example 1.4. Let $A_n = \{1, 2, \dots, n\}$, for $n \in \mathbb{Z}^+$. Then

$$\bigcup_{i=1}^{\infty} A_i = \mathbb{Z}^+ \text{ and } \bigcap_{i=1}^{\infty} A_i = \{1\}.$$

Example 1.5. Let $B_n = [-\frac{1}{n}, 1 - \frac{1}{n}]$, for $n \in \mathbb{Z}^+$. (These are closed intervals in $\mathbb{R}$.) Then

$$\bigcup_{i=1}^{\infty} B_i = [-1, 1) \text{ and } \bigcap_{i=1}^{\infty} B_i = \{0\}.$$

Definition 1.6. A set A is a subset of the set B if $x \in A \implies x \in B$, written $A \subseteq B$. If in addition $A \neq B$, we say that A is a proper subset of B and write $A \subset B$. The set of subsets of A is called the power set of A, denoted $\mathcal{P}(A)$.

Theorem 1.7. *Let A be a finite set with n elements. Then A has 2^n subsets; that is, $|\mathcal{P}(A)| = 2^n$.*

Proof. We use induction on n. If $n = 0$, then $A = \emptyset$, which has exactly 1 subset: itself. Thus $|\mathcal{P}(A)| = 1 = 2^0$.

Now suppose that any set with k elements has 2^k subsets, and let $A = \{x_1, \dots, x_k, x_{k+1}\}$ be a set with $k+1$ elements. We divide the subsets X of A into 2 categories.

Type 1: $x_{k+1} \notin X$. In this case, $X \subseteq \{x_1, \dots, x_k\}$, a set with k elements. By our industion assumption, there are 2^k subsets of Type 1.

Type 2: $x_{k+1} \in X$. Here, we can write $X = \{x_{k+1}\} \cup Y$, where $Y \subseteq \{x_1, \dots, x_k\}$; that is, Y is of Type 1. Hence there are 2^k subsets of Type 2.

Therefore, we have a total of $2^k + 2^k = 2 \cdot 2^k = 2^{k+1}$ subsets of A, completing the proof. $\square$

2. Fields

Definition 2.1. A field is a set F with two operations, denoted $+, \cdot$, satisfying:

(1) Commutativity: $\alpha + \beta = \beta + \alpha$ and $\alpha \cdot \beta = \beta \cdot \alpha$, for all $\alpha, \beta \in F$.
(2) Associativity: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ and $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$, for all $\alpha, \beta, \gamma \in F$.

(3) Distributivity: $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$, for all $\alpha, \beta, \gamma \in F$.
(4) Identities: there exist $0, 1 \in F$ such that $0 + \alpha = \alpha$ and $1 \cdot \alpha = \alpha$, for all $\alpha \in F$.
(5) Inverses: For all $\alpha \in F$ there exists $-\alpha \in F$ such that $\alpha + (-\alpha) = 0$, and for all $0 \neq \beta \in F$ there exists $\beta^{-1} \in F$ such that $\beta \cdot \beta^{-1} = 1$.

Proposition 2.2. *In a field, the identities 0 and 1 are unique.*

Proof. Suppose that $0'$ is another identity for the operation $+$. Then

$$\begin{aligned} 0 + 0' &= 0, \text{ since } 0' \text{ is an identity;} \\ &= 0', \text{ since } 0 \text{ is an identity.} \end{aligned}$$

The proof that 1 is unique is similar. □

Example 2.3. $\mathbb{Z}$ is not a field under the usual $+$ and $\cdot$, since for example 2 has no multiplicative inverse.

Example 2.4. $\mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$ are fields under the usual $+$ and $\cdot$. The only condition that may not be immediately clear is that every $0 \neq z = a + bi \in \mathbb{C}$ has a multiplicative inverse. But in this case, $a^2 + b^2 \neq 0$, and

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

Example 2.5. There are finite fields as well. One is $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$, where the operations are to add and multiply as usual (that is, in $\mathbb{Z}$), and then take the remainder after dividing by 5. More explicitly, look at these tables:

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

It's easy (but pretty tedious!) to check that all the conditions for a field are met. In fact, it's possible to construct a field $\mathbb{Z}_p = \{0, 1, \ldots, p - 1\}$ for any prime number p, but would require some topics not relevant to our course.

3. Exercises

Exercise 3.1. For each of the following collections of sets $A_n, n \in \mathbb{Z}^+$, determine

$$\bigcup_{n=1}^{\infty} A_n \text{ and } \bigcap_{n=1}^{\infty} A_n.$$

(a) $A_n = (-n, n)$
(b) $A_n = [-n, n + 1)$

(c) $A_n = (\frac{1}{n}, n]$
(d) $A_n = [1, 1 + \frac{1}{n})$
(e) $A_n = (1, 1 + \frac{1}{n})$

Exercise 3.2. Let A and B be sets. Prove or disprove each of the following.

(a) $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$
(b) $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$

Exercise 3.3. Use induction to prove that

$$2^n > n^2, \;\; \forall n \geq 5.$$

Exercise 3.4. Let F be a field. Prove each of the following.

(a) For each $a \in F, -a$ is unique.
(b) For each $0 \neq a \in F, a^{-1}$ is unique.

Exercise 3.5. Show that the set of real numbers of the form $a + b\sqrt{2}$, where $a, b \in \mathbb{Q}$, with addition and multiplication as in $\mathbb{R}$, is a field.

CHAPTER II

Vector Spaces

1. Definition, Examples, and Elementary Properties

Definition 1.1. Let F be a field. A vector space V over F is a set with two operations:

- vector addition: $v + w \in V$, for all $v, w \in V$;
- scalar multiplication: $\alpha \cdot v \in V$, for all $\alpha \in F, v \in V$.

These operations satisfy:

(1) $v + w = w + v$, for all $v, w \in V$;
(2) $(v + w) + u = v + (w + u)$, for all $v, w, u \in V$;
(3) There exists $0 \in V$ such that $0 + v = v$, for all $v \in V$;
(4) For all $v \in V$, there exists $-v \in V$ such that $v + (-v) = 0$;
(5) $\alpha \cdot (v + w) = \alpha \cdot v + \alpha \cdot w$, for all $\alpha \in F, v, w \in V$;
(6) $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$, for all $\alpha, \beta \in F, v \in V$;
(7) $(\alpha \cdot \beta) \cdot v = \alpha \cdot (\beta \cdot v)$, for all $\alpha, \beta \in F, v \in V$;
(8) $1 \cdot v = v$, for all $v \in V$.

Example 1.2. The Euclidean spaces $\mathbb{R}^n$ of multivariable calculus and elementary geometry are, of course, vector spaces over $\mathbb{R}$ with the familiar vector addition and scalar multiplication.

Example 1.3. It's easy to generalize the previous example to the vector space

$$F^n = \{(\alpha_1, \cdots, \alpha_n) : \alpha_i \in F\}$$

over any field F: simply define addition and scalar multiplication in the same way, in each coordinate.

Example 1.4. The set of sequences

$$F^\infty = \{(\alpha_1, \alpha_2, \cdots) : \alpha_i \in F\},$$

with coordinate addition and scalar multiplication, is a vector space over F.

Example 1.5. The set of polynomials

$$\mathcal{P}_\infty(F) = \{\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n : \alpha_i \in F, n \in \mathbb{Z}, n \geq 0\},$$

with the usual polynomial addition and scalar multiplication, is a vector space over F. Notice that we are "forgetting" that we know how to multiply polynomials. *We don't multiply vectors!*

Example 1.6. For a fixed $n \geq 0$, the set of polynomials

$$\mathcal{P}_n(F) = \{\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n : \alpha_i \in F\},$$

is a vector space over F.

Example 1.7. For fixed $m, n \geq 0$, the set $M_{m\times n}(F)$ of $m \times n$ matrices

$$(\alpha_{ij}) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ & & \vdots & \\ & & \vdots & \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}$$

with entries in F is a vector space over F. If $m = n$, that is if the matrices are square, we simply write $M_n(F)$.

Example 1.8. If F is a subfield of K, then K is a vector space over F, where scalar multiplication is just multiplication in K.

Notation 1.9. From now on, we'll suppress the "$\cdot$" when referring to either multiplication in the field F or scalar multiplication of F on a vector space V.

Proposition 1.10. *Let V be a vector space over the field F, $\alpha \in F, v \in V$.*

(1) $0v = 0$;
(2) $\alpha 0 = 0$;
(3) $(-\alpha)v = -(\alpha v) = \alpha(-v)$;
(4) $\alpha v = 0 \implies \alpha = 0$ *or* $v = 0$.

Proof.

(1) $0v = (0+0)v = 0v+0v \implies 0v+(-(0v)) = 0v+0v+(-(0v)) \implies 0 = 0v+0 = 0v$.
(2) $\alpha 0 = \alpha(0 + 0) = \alpha 0 + \alpha 0 \implies 0 = \alpha 0$.
(3) $\alpha v + (-\alpha)v = (\alpha + (-\alpha))v = 0v = 0 \implies (-\alpha)v = -(\alpha v)$. (Here we use the fact that the additive inverse in V is unique; this is proven just as in a field, which was a homework problem.) The other result can be proven similarly.
(4) If $\alpha \neq 0$, then $v = 1v = (\alpha^{-1}\alpha)v = \alpha^{-1}(\alpha v) = \alpha^{-1}0 = 0$.

□

2. Subspaces

Definition 2.1. Let V be a vector space over F. A subset W of F is a subspace if W is itself a vector space under the addition and scalar multiplication of V. We write $W \leq V$.

Theorem 2.2. *A subset W of the vector space V is a subspace if and only if:*

(1) $0_V \in W$;

(2) $w, w' \in W \implies w + w' \in W$ *(closure under +);*
(3) $w \in W, \alpha \in F \implies \alpha w \in W$ *(closure under ·).*

Proof. ($\implies$) Suppose $W \leq V$. Conditions (2) and (3) follow immediately, since we are assuming that addition and scalar multiplication are operations on W. Suppose then that 0_W is additive identity in W. Then

$$\begin{aligned} 0_W + 0_W = 0_W &\implies 0_W + 0_W + (-(0_W)) = 0_W + (-(0_W)) \\ &\implies 0_W + 0_V = 0_V \\ &\implies 0_W = 0_V \in W. \end{aligned}$$

($\impliedby$) Suppose that the three conditions hold. The only requirement for a vector space that is not given or does not follow from the same requirement for V is the existence of additive inverses in W. But if $w \in W$, then $-w = -(1w) = (-1)w \in W$ by condition (3). ☐

Example 2.3. $\{(x, 0) : x \in \mathbb{R}\} \leq \mathbb{R}^2$.

Example 2.4. Fix $v \in \mathbb{R}^n$. Then $\{\alpha v : \alpha \in \mathbb{R}\} \leq \mathbb{R}^n$.

Example 2.5. Let $n \in \mathbb{Z}^+$ and F a field. Then $\{\alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_n x^n : \alpha_i \in F\} \leq \mathcal{P}_n(F)$.

Example 2.6. For any field F, the set of diagonal matrices

$$D = \left\{ \begin{pmatrix} \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_2 & 0 & \cdots & 0 & 0 \\ & & & \vdots & & \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & \alpha_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_n \end{pmatrix} : \alpha_i \in F \right\}$$

is a subspace of $M_n(F)$.

Proposition 2.7. *Let* $\{W_i : i \in I\}$ *be a collection of subspaces of* V*. Then*

$$W = \bigcap_{i \in I} W_i \leq V.$$

Proof. $0 \in W_i$ for all i, so $0 \in W$. If $w, w' \in W$, then $w, w' \in W_i$ for all i, so $w + w' \in W_i$ for all i, meaning that $w + w' \in W$. Closure under scalar multiplication is proven similarly and left as an exercise. ☐

Example 2.8. The corresponding statement for unions is not true. Consider that

$$(1, 0) \in W_x = \{(x, 0) : x \in \mathbb{R}\} \leq \mathbb{R}^2$$

and

$$(0, 1) \in W_y = \{(0, y) : y \in \mathbb{R}\} \leq \mathbb{R}^2$$

but

$$(1, 0) + (1, 0) = (1, 1) \notin W_x \cup W_y.$$

3. Linear Combinations and Spans

Definition 3.1. Let V be a vector space over F, and let $v_1, \ldots, v_n \in V, \alpha_1, \ldots, \alpha_n \in F$. The vector

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n = \sum_{i=1}^{n} \alpha_i v_i$$

is a linear combination of the vectors $\{v_i\}$.

Example 3.2. In $\mathbb{R}^2$, $v = (-6, 12)$ is a linear combination of $\{(1,4), (-6,0)\}$ since

$$(-6, 12) = 3 \cdot (1,4) + 2 \cdot (-6, 0).$$

Example 3.3. In $\mathcal{P}_3(\mathbb{C})$, $v = 3-(4+4i)x-x^2$ is a linear combination of $\{1-(1+i)x, i+ix^2\}$ since

$$3 - (4+4i)x - x^2 = 4 \cdot (1 - (1+i)x) + i \cdot (i + ix^2).$$

Remark 3.4. How do we tell if a given vector v is a linear combination of $\{v_i\}$? We must find the scalars $\{\alpha_i\}$.

Example 3.5. In $\mathbb{R}^2$, to see if $(1,1)$ a linear combination of $\{1,2),(1,3)\}$, we set

$$(1,1) = \alpha_1(1,2) + \alpha_2(1,3) = (\alpha_1 + \alpha_2, 2\alpha_1 + 3\alpha_3).$$

So we must solve

$$\begin{array}{rcl} \alpha_1 + \alpha_2 & = & 1 \\ 2\alpha_1 + 3\alpha_2 & = & 1 \end{array} \iff \begin{array}{rcl} \alpha_1 + \alpha_2 & = & 1 \\ \alpha_2 & = & -1 \end{array} \iff \begin{array}{rcl} \alpha_1 & = & 2 \\ \alpha_2 & = & -1 \end{array}$$

Example 3.6. In $\mathcal{P}_3(\mathbb{R})$, to see if $1+x^3$ a linear combination of $\{1+x+x^2+x^3, x^2-2x^3\}$, we set

$$1 + x^3 = \alpha_1(1 + x + x^2 + x^3) + \alpha_2(x^2 - 2x^3) = \alpha_1 + \alpha_1 x + (\alpha_1 + \alpha_2)x^2 + (\alpha_1 - 2\alpha_2)x^3.$$

So we equate coefficients and try to solve

$$\begin{array}{rcl} \alpha_1 & = & 1 \\ \alpha_1 & = & 0 \\ \alpha_1 + \alpha_2 & = & 0 \\ \alpha_1 - 1\alpha_2 & = & 1 \end{array}$$

But there is clearly no solution.

Definition 3.7. Let V be a vector space over F, and let $X \subseteq V$. The span of X is the set of linear combinations of the elements of X. That is,

$$\operatorname{Span}(X) = \{\alpha_1 v_1 + \ldots + \alpha_n v_n : n \in \mathbb{Z}^+, \alpha_i \in F, v_i \in X\}.$$

For convenience, we'll take $\operatorname{Span}(\emptyset) = \{0\}$.

Example 3.8. In $\mathbb{R}^2$, $(1,1) \in \operatorname{Span}(\{1,2),(1,3)\})$.

Example 3.9. In $\mathcal{P}_3(\mathbb{R})$, $1 + x^3 \notin \operatorname{Span}(\{1 + x + x^2 + x^3, x^2 - 2x^3\})$.

Proposition 3.10. *Let V be a vector space over F, and let $X \subseteq V$. Then* $\operatorname{Span}(X) \leq V$.

Proof. If $X = \emptyset$, then $\operatorname{Span}(X) = \{0\}$ is trivially a subspace. So suppose $v, w \in X$ and $\alpha \in F$.

(1) $0 = 0 \cdot v \in \operatorname{Span}(X)$.

(2) By padding with 0 coefficients if necessary, we may write

$$v = \sum_{i=1}^{n} \alpha_i v_i, \quad w = \sum_{i=1}^{n} \beta_i v_i,$$

where $v_i \in X$ and $\alpha_i, \beta_i \in F$. Then

$$v + w = \sum_{i=1}^{n} \alpha_i v_i + \sum_{i=1}^{n} \beta_i v_i = \sum_{i=1}^{n} (\alpha_i + \beta_i) v_i \in \operatorname{Span}(X).$$

(3)

$$\alpha v = \alpha \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} (\alpha \alpha_i) v_i \in \operatorname{Span}(X).$$

□

Example 3.11. In $\mathbb{R}^2$, $\operatorname{Span}(\{1,2),(1,3)\}) = \mathbb{R}^2$. To see this, set

$$(x, y) = \alpha_1(1,2) + \alpha_2(1,3) = (\alpha_1 + \alpha_2, 2\alpha_1 + 3\alpha_2).$$

and solve

$$\begin{array}{rcl} \alpha_1 + \alpha_2 & = & x \\ 2\alpha_1 + 3\alpha_2 & = & y \end{array} \iff \begin{array}{rcl} \alpha_1 + \alpha_2 & = & x \\ \alpha_2 & = & y - 2x \end{array} \iff \begin{array}{rcl} \alpha_1 & = & 3x - y \\ \alpha_2 & = & y - 2x \end{array}$$

Example 3.12. In $\mathbb{R}^2$, $\operatorname{Span}(\{1,2),(2,4)\}) \neq \mathbb{R}^2$, because if we set

$$(x, y) = \alpha_1(1,2) + \alpha_2(2,4) = (\alpha_1 + 2\alpha_2, 2\alpha_1 + 4\alpha_2),$$

and try to solve

$$\begin{array}{rcl} \alpha_1 + 2\alpha_2 & = & x \\ 2\alpha_1 + 4\alpha_2 & = & y \end{array} \iff \begin{array}{rcl} \alpha_1 + 2\alpha_2 & = & x \\ 0 & = & y - 2x \end{array}$$

we see that only vectors of the form $(x, 2x)$ are in $\operatorname{Span}(\{1,2),(2,4)\}) \neq \mathbb{R}^2$. This is obviously because $(2,4) \in \operatorname{Span}(\{(1,2)\})$ (or equivalently, $(1,2) \in \operatorname{Span}(\{(2,4)\})$).

4. Linear Independence and Bases

Definition 4.1. A nonempty subset X of a vector space V over F is linearly independent if for all scalars $\alpha_1, \ldots, \alpha_n \in F$ and vectors $x_1, \ldots, x_n \in X$,

$$\sum_{i=1}^{n} \alpha_i x_i = 0 \implies \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0.$$

If X is not linearly independent, or in short is linearly dependent, then there must be a dependence relation among the vectors in X; that is, there exist scalars $\alpha_1, \ldots, \alpha_n \in F$ that are not all 0, and vectors $x_1, \ldots, x_n \in X$, which together satisfy

$$\sum_{i=1}^{n} \alpha_i x_i = 0.$$

Example 4.2. Let $X = \{(1,1),(1,2)\} \subseteq \mathbb{R}^2$. Suppose that

$$\alpha_1(1,1) + \alpha_2(1,2) = (\alpha_1 + \alpha_2, \alpha_1 + 2\alpha_2) = (0,0).$$

Then

$$\begin{array}{rcl} \alpha_1 + \alpha_2 &=& 0 \\ \alpha_1 + 2\alpha_2 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 + \alpha_2 &=& 0 \\ \alpha_2 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 &=& 0 \\ \alpha_2 &=& 0 \end{array}$$

Thus X is linearly independent.

Example 4.3. Let $X = \{1 + 2x + x^3, 3 - x^2, 2 - x + x^3\} \subseteq \mathcal{P}_3(\mathbb{R})$. Suppose that

$$\alpha_1(1+2x+x^3)+\alpha_2(3-x^2)+\alpha_3(2-x+x^3) = (\alpha_1+3\alpha_2+2\alpha_3)+(2\alpha_1-\alpha_3)x-\alpha_2x^2+(\alpha_1+\alpha_3)x^3 = 0.$$

Then equating coefficients, we have

$$\begin{array}{rcl} \alpha_1 + 3\alpha_2 + 2\alpha_3 &=& 0 \\ 2\alpha_1 - \alpha_3 &=& 0 \\ -\alpha_2 &=& 0 \\ \alpha_1 + \alpha_3 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 + 2\alpha_3 &=& 0 \\ 2\alpha_1 - \alpha_3 &=& 0 \\ \alpha_2 &=& 0 \\ 3\alpha_1 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 &=& 0 \\ \alpha_2 &=& 0 \\ \alpha_3 &=& 0 \end{array}$$

Thus X is linearly independent.

Example 4.4. Let $X = \{(1,1,1),(1,2,3),(-1,-3,-5)\} \subseteq \mathbb{R}^3$. Suppose that

$$\alpha_1(1,1,1)+\alpha_2(1,2,3)+\alpha_3(-1,-3,-5) = (\alpha_1+\alpha_2-\alpha_3, \alpha_1+2\alpha_2-3\alpha_3, \alpha_1+3\alpha_2-5\alpha_3) = (0,0,0).$$

Then

$$\begin{array}{rcl} \alpha_1 + \alpha_2 - \alpha_3 &=& 0 \\ \alpha_1 + 2\alpha_2 - 3\alpha_3 &=& 0 \\ \alpha_1 + 3\alpha_2 - 5\alpha_3 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 + \alpha_2 - \alpha_3 &=& 0 \\ \alpha_2 - 2\alpha_3 &=& 0 \\ 2\alpha_2 - 4\alpha_3 &=& 0 \end{array} \implies \begin{array}{rcl} \alpha_1 + \alpha_2 - \alpha_3 &=& 0 \\ \alpha_2 - 2\alpha_3 &=& 0 \\ 0 &=& 0 \end{array}$$

Hence choosing $\alpha_3 = 1$ leads to $\alpha_2 = 2, \alpha_1 = -1$ and the dependence relation

$$-(1,1,1) + 2(1,2,3) + (-1,-3,-5) = (0,0,0).$$

Thus X is linearly dependent.

Remark 4.5. It's easy to see (*check!*) that if X is dependent and $X \subseteq Y$, then Y is dependent; and also that if X is independent and $\emptyset \neq Y \subseteq X$, then Y is independent.

Proposition 4.6. *Let X be a linearly independent set in V, and $x \in V$. Then $X \cup \{x\}$ is linearly independent if and only if $x \notin \operatorname{Span}(X)$.*

Proof. ($\implies$) Suppose that $x \in \operatorname{Span}(X)$, so that we can write

$$x = \sum_{i=1}^{n} \alpha_i x_i, \text{ for some } \alpha_i \in F, x_i \in X.$$

But then

$$1 \cdot x + \sum_{i=1}^{n} (-\alpha_i) x_i = 0$$

is a dependence relation among the vectors in $X \cup \{x\}$.

($\Longleftarrow$) Suppose that $X \cup \{x\}$ is linearly dependent. Then we can write

$$\alpha x + \sum_{i=1}^{n} \alpha_i x_i = 0,$$

for some $\alpha, \alpha_i \in F$ not all 0 and some $x_i \in X$. Then $\alpha \neq 0$ since otherwise we would have a dependence relation among the vectors of X. So α has a multiplicative inverse in F, and therefore

$$x = \sum_{i=1}^{n} (-\alpha^{-1} \alpha_i) x_i \in \mathrm{Span}(X).$$

$\square$

Definition 4.7. A subset $\mathcal{B}$ of a vector space V over the field F is a basis of V if:

(1) $\mathcal{B}$ is linearly independent;
(2) $\mathcal{B}$ spans V; that is, $\mathrm{Span}(\mathcal{B}) = V$.

Example 4.8. $\mathcal{B} = \{(1,0),(0,1)\}$ is a basis of $\mathbb{R}^2$:

- $\mathcal{B}$ is independent:

$$\alpha(1,0) + \beta(0,1) = (\alpha, \beta) = (0,0) \implies \alpha = \beta = 0.$$

- $\mathcal{B}$ spans $\mathbb{R}^2$:

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

Example 4.9. $\mathcal{B} = \{(1,1),(1,2)\}$ is another basis of $\mathbb{R}^2$:

- $\mathcal{B}$ is independent:

$$\alpha(1,1) + \beta(1,2) = (\alpha + \beta, \alpha + 2\beta) = (0,0) \implies \begin{array}{rcl} \alpha + \beta & = & 0 \\ \alpha + 2\beta & = & 0 \end{array}$$

$$\implies \begin{array}{rcl} \alpha + \beta & = & 0 \\ \beta & = & 0 \end{array}$$

$$\implies \begin{array}{rcl} \alpha & = & 0 \\ \beta & = & 0 \end{array}$$

- $\mathcal{B}$ spans $\mathbb{R}^2$:

$$(x,y) = \alpha(1,1) + \beta(1,2) = (\alpha + \beta, \alpha + 2\beta) \implies \begin{array}{rcl} \alpha + \beta & = & x \\ \alpha + 2\beta & = & y \end{array}$$

$$\implies \begin{array}{rcl} \alpha + \beta & = & x \\ \beta & = & y - x \end{array}$$

$$\implies \begin{array}{rcl} \alpha & = & 2x - y \\ \beta & = & y - x \end{array}$$

Thus $(x,y) = (2x-y)(1,1) + (y-x)(1,2)$.

Example 4.10. Let F be a field, and let $e_1 = (1,0,\dots,0), e_2 = (0,1,\dots,0), \dots, e_n = (0,0,\dots,1) \in F^n$. Then $\mathcal{B} = \{e_1,\dots,e_n\}$ is a basis of F^n.

Example 4.11. $\{1,x,x^2,\dots,x^n\}$ is a basis of $\mathcal{P}_n(F)$ and $\{1,x,x^2,\dots\}$ is a basis of $\mathcal{P}_\infty(F)$.

Example 4.12. $\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ is a basis of $M_2(F)$.

Example 4.13. More generally, ket $A_{ij} = (\alpha_{kl}) \in M_{m\times n}(F)$, where

$$\alpha_{kl} = \begin{cases} 1, & \text{if } k = i \text{ and } l = j \\ 0, & \text{otherwise} \end{cases}$$

Then $\mathcal{B} = \{A_{ij}\}$ is a basis of $M_{m\times n}(F)$.

Theorem 4.14. *A subset $\mathcal{B}$ is a basis of the vector space V if and only every vector in V is uniquely a linear combination of the elements of V.*

Proof. ($\Longrightarrow$) If $\mathcal{B}$ is a basis, then $\mathrm{Span}(\mathcal{B}) = V$ means that every $v \in V$ is a linear combination of the elements of $\mathcal{B}$. But also, by independence, if $b_1,\dots,b_n \in \mathcal{B}$,

$$v = \sum_{i=1}^n \alpha_i b_i = \sum_{i=1}^n \beta_i b_i \implies 0 = \sum_{i=1}^n (\alpha_i - \beta_i) b_i \implies \alpha_i - \beta_i = 0 \implies \alpha_i = \beta_i, \forall i.$$

($\Longleftarrow$) It's immediate that $V = \mathrm{Span}(\mathcal{B})$, and

$$\sum_{i=1}^n \alpha_i b_i = 0 = \sum_{i=1}^n 0 \cdot b_i \implies \alpha_i = 0, \forall i.$$

Thus $\mathcal{B}$ is also independent. $\square$

Lemma 4.15. *Let F be a field, and let*

$$\begin{array}{rcl} \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n & = & 0 \\ \alpha_{21}x_1 + \alpha_{22}x_2 + \ldots + \alpha_{1n}x_n & = & 0 \\ & \vdots & \\ \alpha_{m1}x_1 + \alpha_{m2}x_2 + \ldots + \alpha_{mn}x_n & = & 0 \end{array} \tag{1}$$

be a system of m linear equations in the n unknowns x_j with coefficients $\alpha_{ij} \in F$, where $m < n$. Then there exists a nontrivial (that is, not all 0) solution

$$x_j = \alpha_j \in F, j = 1,2,\dots,n.$$

Proof. We use induction on m.

If $m = 1$, we have simply

$$\alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n = 0,$$

where $n \geq 2$. If $\alpha_{11} = 0$, then $x_1 = 1, x_2 = x_3 = \ldots = x_n = 0$ is a nontrivial solution. If $\alpha_{11} \neq 0$, then

$$x_j = \begin{cases} 1, & \text{if } j > 1 \\ -\alpha_{11}^{-1}(\alpha_{12} + \cdots + \alpha_{1n}), & \text{if } j = 1 \end{cases}$$

is a nontrivial solution.

Suppose now that any homogeneous system of $m-1$ equations in more than $m-1$ unknowns has a nontrivial solution, and consider the system in the statement of the Lemma. If all $\alpha_{ij} = 0$, then $x_1 = x_2 = \ldots = x_n = 1$ is a nontrivial solution. Thus we can assume that at least one coefficient is nonzero, so by reordering the equations and renumbering the unknowns (if necessary), we can assume $\alpha_{11} \neq 0$. By adding appropriate multiples of the first equation to the others, we get the equivalent system

$$\begin{aligned} \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n &= 0 \\ \beta_{22}x_2 + \ldots + \beta_{1n}x_n &= 0 \\ &\vdots \\ \beta_{m2}x_2 + \ldots + \beta_{mn}x_n &= 0 \end{aligned} \tag{2}$$

where $\beta ij \in F$. By induction, the last $m-1$ equations have a nontrivial solution

$$x_2 = \alpha_2, \ldots, x_n = \alpha_n, \alpha_j \in F.$$

Now choose

$$x_1 = -\alpha_{11}^{-1}(\alpha_{12}\alpha_2 + \cdots + \alpha_{1n}\alpha_n)$$

to arrive at a nontrivial solution of the original system (1). $\square$

Lemma 4.16. *Let X be a spanning set in the vector space V containing n elements, for some $n \in \mathbb{Z}^+$. Then any set of $n+1$ or more vectors in V is dependent.*

Proof. Let $X = \{x_1, \ldots, x_n\}$ and let $Y = \{v_1, \ldots, v_n, v_{n+1}\} \subseteq V$. Then for each $i = 1, \ldots, n+1$ we can write

$$v_i = \sum_{j=1}^{n} \alpha_{ij}x_j, \text{ for some } \alpha_{ij} \in F.$$

Now take an arbitrary linear combination

$$v = \sum_{i=1}^{n+1} \alpha_i v_i, \alpha_i \in F.$$

Setting $v = 0$ and equating coefficients gives rise to the following system of linear equations:

$$\begin{aligned} \alpha_{11}\alpha_1 + \alpha_{21}\alpha_2 + \ldots + \alpha_{n+1\,1}\alpha_{n+1} &= 0 \\ \alpha_{12}\alpha_1 + \alpha_{22}\alpha_2 + \ldots + \alpha_{n+1\,2}\alpha_{n+1} &= 0 \\ &\vdots \\ \alpha_{1n}\alpha_1 + \alpha_{2n}\alpha_2 + \ldots + \alpha_{n+1\,n}\alpha_{n+1} &= 0 \end{aligned}$$

But this is a homogeneous system with fewer equations than unknowns, so must have a nontrivial solution. Thus Y is dependent. ☐

Theorem 4.17. *Let V be a vector space over F, and suppose that $\mathcal{B}$ is a basis of V containing $n \in \mathbb{Z}^+$ elements. Then any basis of V also contains n elements.*

Proof. If $\mathcal{B}'$ were another basis containing more than n elements (including possible infinitely many), it would be a dependent set, since $\mathcal{B}$ spans V. Thus $\mathcal{B}'$ must contain $m \leq n$ elements. But if $m < n$, then $\mathcal{B}$ would be dependent. Thus $m = n$. ☐

Definition 4.18. Let V be a vector space over F. If V has a finite basis containing n elements, we call V finite dimensional over F, of dimension n, and write $n = \dim_F V$. Otherwise, V is infinite dimensional over F.

Example 4.19. $\dim_F F^n = n$

Example 4.20. $\dim_F \mathcal{P}_n(F) = n + 1$

Example 4.21. $\dim_F M_{m\times n}(F) = mn$

Example 4.22. F^∞ and $\mathcal{P}_\infty(F)$ are infinite dimensional over F.

Example 4.23. $\dim_{\mathbb{R}} \mathbb{C} = 2$

Proposition 4.24. *Let $\dim_F V = n$. Then any independent subset $X = \{x_1, \ldots, x_m\}$ of V is contained in a basis.*

Proof. We must have $m \leq n$ by Lemma 4.16. If $\operatorname{Span}(X) = V$, then X is itself a basis (and $m = n$). If on the other hand $x \in \operatorname{Span}(X) \setminus V$, then $X \cup \{x\}$ is independent by Proposition 4.6.

Now repeat the process; since we cannot have $n+1$ independent vectors, it must stop when we reach a total of n and have a basis. ☐

Proposition 4.25. *Let $\dim_F V$. Then any spanning set $X = \{x_1, \ldots, x_m\}$ in V contains a basis.*

Proof. First notice that this time we must have $m \geq n$ (again by Lemma 4.16), since otherwise a basis would be dependent. If X is linearly independent, then X is a basis (and again $m = n$). Otherwise, we can write

$$\alpha_1 x_1 + \ldots + \alpha_m x_m = 0,$$

where, without loss of generality, $\alpha_1 \neq 0$. Then

$$x_1 = -\alpha_1^{-1}(\alpha_2 x_2 + \ldots + \alpha_m x_m) \subseteq \operatorname{Span}(X).$$

We claim that $X' = \{x_2, \ldots, x_m\}$ now spans V. To see this, let $v \in V$, and write

$$v = \beta_1 x_1 + \ldots + \beta_m x_m = \beta_1(-\alpha_1^{-1}(\alpha_2 x_2 + \ldots + \alpha_m x_m)) + \beta_2 x_2 + \ldots + \beta_m x_m \in \operatorname{Span}(X').$$

Now repeat this process. It must stop with an independent set, since at worst, we reach the independent set $\{x_m\}$. $\square$

Corollary 4.26. *If* $\dim_F V = n$ *and* $\mathcal{B}$ *is an* $n-$ *element subset of* V*, then*

$$\mathcal{B} \text{ is a basis} \iff \operatorname{Span}(\mathcal{B}) = V \iff \mathcal{B} \text{ is independent.}$$

$\square$

5. Exercises

Exercise 5.1.

(a) Is $\mathbb{C}^n$ (with coordinate addition and scalar multiplication) a vector space over $\mathbb{R}$? Justify your answer.
(b) Is $\mathbb{R}^n$ (with coordinate addition and scalar multiplication) a vector space over $\mathbb{C}$? Justify your answer.

Exercise 5.2. Determine, with proof, whether each of the following subsets W is a subspace of the given vector space V.

(a) $W = \{(x_1, x_2, \ldots, x_n) : x_1 = 0\}; V = F^n$
(b) $W = \{(x_1, x_2, \ldots, x_n) : x_1^2 = x_2\}; V = F^n$
(c) $W = \{f : f(0) = 0\}; V = \mathcal{P}_\infty(\mathcal{F})$
(d) $W = \{f : f(0) = 1\}; V = \mathcal{P}_\infty(\mathcal{F})$

Exercise 5.3. Let W_1 and W_2 be subspaces of V. Show that $W_1 \cup W_2$ is a subspace of V if and only if $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.

Exercise 5.4. Determine, with proof, whether each of the following vectors v is a linear combination of the set X in the given vector space V.

(a) $v = (1, -2); X = \{(1, 1), (1, 2)\}; V = \mathbb{R}^2$
(b) $v = x; X = \{1 + x + x^2, 1 + x - x^2, x^2\}; V = \mathcal{P}_2(\mathbb{R})$
(c) $v = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; X = \left\{ \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right\}; V = M_2(F)$

Exercise 5.5. Show that a subset W of a vector space is a subspace if and only if $\operatorname{Span}(W) = W$.

Exercise 5.6. Determine, with proof, whether each of the following sets of vectors X is linearly independent in the given vector space V.

(a) $X = \{(1, 0), (1, 1), (1, -1)\}; V = \mathbb{R}^2$
(b) $X = \{1, 1 + x, 1 - x^2\}; V = \mathcal{P}_2(\mathbb{R})$
(c) $X = \{(1, \frac{1}{2}, \frac{1}{3}, \ldots), (\sin 1, \sin 2, \sin 3, \ldots)\}; V = \mathbb{R}^\infty$

Exercise 5.7. Suppose that $\{v_1, v_2, \ldots, v_n\}$ is linearly independent in V. Show that $\{v_1, v_1 + v_2, \ldots, v_1 + v_2 + \ldots + v_n\}$ is linearly independent as well.

Exercise 5.8. Determine if the following sets are bases of the indicated vector space.

(a) $\{(-1,3,1),(2,-4,-3),(-3,8,2)\} \subseteq \mathbb{R}^3$
(b) $\{(-1,-3,-2),(-3,1,3),(-2,-10,-2)\} \subseteq \mathbb{R}^3$
(c) $\{1+2x-x^2, 4-2x+x^2, -1+18x-9x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$
(d) $\{-1+2x+4x^2, 3-4x-10x^2, -2-5x-6x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$

Exercise 5.9. Consider the following system of linear equations:

$$x_1 - 2x_2 + x_3 = 0$$
$$2x_1 - 3x_2 + x_3 = 0$$

(a) Show that the set of solutions is a subspace of $\mathbb{R}^3$.
(b) Find a basis for that subspace.

Exercise 5.10. The *trace* of a matrix $A = (\alpha_{ij}) \in M_n(F)$ is defined to be

$$tr(A) = \sum_{i=1}^{n} \alpha_{ii}.$$

(a) Show that the set of matrices with trace 0 is a subspace of $M_n(F)$.
(b) Find a basis for that subspace.

Exercise 5.11. A matrix $A = (\alpha_{ij}) \in M_n(F)$ is *symmetric* if $\alpha_{ij} = \alpha_{ji}$, for all i and j.

(a) Show that the set of symmetric matrices is a subspace of $M_n(F)$.
(b) Find a basis for that subspace.

CHAPTER III

Linear Transformations

1. Functions

Definition 1.1. A function $f : A \to B$ is a rule that assigns to each element a of the set A a unique element $f(a)$ of the set B. A is the domain of f, B the codomain, and the image of f is the set

$$f(A) = \{b \in B : \exists a \in A \text{ such that } b = f(a)\}.$$

Definition 1.2. Let $f : A \to B$ be a function.

- f is injective (or $1-1$) if $f(a_1) = f(a_2) \implies a_1 = a_2, \forall a_1, a_2 \in A$.
- f is surjective (or onto) if $\forall b \in B\ \exists a \in A$ such that $b = f(a)$.
- f is bijective if f is both injective and surjective.

Example 1.3. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is bijective.

Example 1.4. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is neither injective nor surjective.

Example 1.5. $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (x^2 + y, z - y)$ is surjective but not injective.

Example 1.6. $f : M_2(\mathbb{R}) \to \mathbb{R}$ defined by

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

is surjective but not injective.

Example 1.7. $f : \mathbb{R} \to \mathbb{Z}$ defined by $f(x) = \lfloor x \rfloor$ (the greatest integer less than or equal to x) is surjective but not injective.

Example 1.8. Let $S = \{s : s \text{ is one of the 50 states}\}$ and C the set of US citizens. Define $g : S \to C$ by $g(s)$ is the governor of s. Then g is injective but not surjective.

Example 1.9. For any set A, the identity function $i_A : A \to A$, defined by $i_A(a) = a$, is bijective.

Definition 1.10. Let $f : A \to B$ and $g : B \to C$. The composition $g \circ f : A \to C$ is defined by $(g \circ f)(a) = g(f(a)), \forall a \in A$.

Definition 1.11. The function $f : A \to B$ is invertible if there exists a function $f^{-1} : B \to A$ such that $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.

Theorem 1.12. *Let $f : A \to B$. Then f is invertible if and only if f is bijective.*

Proof. ($\Longrightarrow$) Suppose f is invertible. If $a_1, a_2 \in A$, then

$$f(a_1) = f(a_2) \implies f^{-1}(f(a_1)) = f^{-1}(f(a_2)) \implies i_A(a_1) = i_A(a_2) \implies a_1 = a_2,$$

so f is injective. If $b \in B$, then $f^{-1}(b) \in A$, and thus $f(f^{-1}(b)) = i_B(b) = b$, so f is also surjective.
($\Longleftarrow$) Suppose that f is bijective. If $b \in B$, then there exists $a \in A$ such that $f(a) = b$ since f is surjective, and moreover a is unique since f is injective. So define $f^{-1}(b) = a$. Then $f(f^{-1}(b)) = f(a) = b$ and $f^{-1}(f(a)) = f^{-1}(b) = a$. $\square$

2. Linear Transformations

Definition 2.1. Let V and W be vector spaces over the field F. A function $T : V \to W$ is a linear transformation if:

(1) $T(u + v) = T(u) + T(v), \forall u, v \in V$;
(2) $T(\alpha v) = \alpha T(v), \forall \alpha \in F, v \in V$.

Example 2.2. $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x, y) = (x + 2y, 3x, x - y)$ is a linear transformation, since

(1)

$$\begin{aligned} T((x, y) + (x', y')) &= T(x + x', y + y') \\ &= ((x + x') + 2(y + y'), 3(x + x'), (x + x') - (y + y')) \\ &= (x + 2y, 3x, x - y) + (x' + 2y', 3x', x' - y') \\ &= T(x, y) + T(x', y'); \end{aligned}$$

(2)

$$\begin{aligned} T(\alpha(x, y)) &= T(\alpha x, \alpha y) \\ &= (\alpha x + 2\alpha y, 3\alpha x, \alpha x - \alpha y) \\ &= \alpha(x + 2y, 3x, x - y) \\ &= \alpha T(x, y). \end{aligned}$$

Example 2.3. $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x, y) = (x^2 + y, y, x)$ is not a linear transformation, since for example: $T(1, 0) + T(1, 0) = (2, 0, 2)$ but $T(2, 0) = (4, 0, 2)$.

Example 2.4. Define $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by $R_\theta(v)$ is the vector obtained by rotating v counterclockwise by θ radians. Thus if we write $v = (x, y) = (r \cos \phi, r \sin \phi)$, we see that

$$\begin{aligned} R_\theta(v) &= (r \cos(\phi + \theta), r \sin(\phi + \theta)) \\ &= (r[\cos \phi \cos \theta - \sin \phi \sin \theta], r[\sin \phi \cos \theta + \cos \phi \sin \theta]) \\ &= (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta). \end{aligned}$$

Thus, as in the first example, R_θ is a linear transformation.

Example 2.5. By elementary calculus, the function $D : \mathcal{P}_n(\mathbb{R}) \to \mathcal{P}_{n-1}(\mathbb{R})$ defined by $D(f(x)) = f'(x)$ is a linear transformation.

Example 2.6. By elementary calculus, the function $I : \mathcal{P}_n(\mathbb{R}) \to \mathbb{R}$ defined by$I(f(x)) = \int_0^1 f(x)dx$ is a linear transformation.

Example 2.7. Clearly $i_V : V \to V$ is a linear transformation.

Proposition 2.8. *Let $T : V \to W$ be a linear transformation. Then $T(0_V) = 0_W$.*

Proof. $T(0_V) = T(0 \cdot 0_V) = 0 \cdot T(0_V) = 0_W$. □

Proposition 2.9. *Let $T : V \to W$ be a linear transformation. If $v \in V$, then $T(-v) = -T(v)$.*

Proof. $T(-v) = T((-1) \cdot v) = (-1) \cdot T(v) = -T(v)$. □

Definition 2.10. Let $T : V \to W$ be a linear transformation. The kernel or nullspace of T is $\operatorname{Ker} T = \{v \in V : T(v) = 0_W\}$.

Proposition 2.11. *Let $T : V \to W$ be a linear transformation. Then $\operatorname{Ker} T \leq V$.*

Proof. $0_V \in \operatorname{Ker} T$ by Proposition 2.8. If $u, v \in \operatorname{Ker} T$, then

$$T(u + v) = T(u) + T(v) = 0_W + 0_W = 0_W \implies u + v \in \operatorname{Ker} T,$$

and if $\alpha \in F, v \in V$, then

$$T(\alpha v) = \alpha T(v) = \alpha \cdot 0_W = 0_W \implies \alpha v \in \operatorname{Ker} T.$$

□

Example 2.12. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x, y, z) = (x + y, x - 2z)$. Then

$$(x, y, z) \in \operatorname{Ker} T \iff \begin{array}{rcl} x + y & = & 0 \\ x - 2z & = & 0 \end{array} \iff (x, y, z) = (2\alpha, -2\alpha, \alpha), \text{ for some } \alpha \in \mathbb{R}.$$

Thus $\operatorname{Ker} T = \operatorname{Span}(\{(2, -2, 1)\})$.

Terminology 2.13. When $T : V \to W$ is a linear transformation, we denote the image of T (as a function) by

$$\operatorname{Im}(T) = T(V) = \{w \in W : \exists v \in V \text{ such that } w = T(v)\}.$$

Proposition 2.14. *Let $T : V \to W$ be a linear transformation. Then $\operatorname{Im} T \leq W$.*

Proof. We have proven that $T(0_V) = 0_W$, so $0_V \in \operatorname{Im} T$. If $w, w' \in \operatorname{Im} T$, then we have $T(v) = w$ and $T(v') = w'$ for some $v, v' \in V$. Then

$$w + w' = T(v) + T(v') = T(v + v') \implies w + w' \in \operatorname{Im} T.$$

Similarly, if $w \in \operatorname{Im} T$ so that $T(v) = w$ for some $v \in V$, and if $\alpha \in F$, then

$$\alpha w = \alpha T(v) = T(\alpha v) \implies \alpha w \in \operatorname{Im} T.$$

$\square$

Example 2.15. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $T(x, y) = (x + y, x - y, y)$, which is easily seen to be a linear transformation. Then

$$(a, b, c) \in \operatorname{Im} T \iff \begin{array}{rcl} x + y &=& a \\ x - y &=& b \\ y &=& c \end{array} \iff \begin{array}{rcl} x + y &=& a \\ y &=& \frac{a-b}{2} \\ y &=& c \end{array} \iff \begin{array}{rcl} x &=& \frac{a+b}{2} \\ y &=& \frac{a-b}{2} \\ y &=& c \end{array}$$

Thus

$$\operatorname{Im} T = \left\{ \left(a, b, \frac{a-b}{2} \right) \right\} = \left\{ a\left(1, 0, \frac{1}{2}\right) + b\left(0, 1, -\frac{1}{2}\right) \right\} = \operatorname{Span}(\{(2, 0, 1), (0, 2, -1)\}).$$

Proposition 2.16. *Let $T : V \to W$ be a linear transformation. Then*

(1) T is surjective $\iff \operatorname{Im} T = W$.
(2) T is injective $\iff \operatorname{Ker} T = \{0_V\}$.

Proof.

(1) This is immediate from the definitions.
(2) If T is injective and $v \in \operatorname{Ker} T$, then $T(v) = 0_W = T(0_V) \implies v = 0_V$. Conversely, if $\operatorname{Ker} T = \{0_V\}$, then

$$T(v) = T(v') \implies T(v - v') = 0_W \implies v - v' = 0_V \implies v = v'.$$

$\square$

Definition 2.17. Let $T : V \to W$ be a linear transformation. The nuliity of T is $n(T) = \dim(\operatorname{Ker} T)$ and the rank of T is $r(T) = \dim(\operatorname{Im} T)$.

Theorem 2.18 (***Rank-Nullity Theorem***)**.** *Let $T : V \to W$ be a linear transformation, and let $\dim V = n$. Then*

$$n(T) + r(T) = n.$$

Proof. Let $m = n(T)$ and take a basis $X = \{x_1, \ldots, x_m\}$ of $\operatorname{Ker} T$. Expand X to a basis $\{x_1, \ldots, x_n\}$ of V. It will suffice to show that $Y = \{T(x_{m+1}), \ldots, T(x_n)\}$ is a basis of $\operatorname{Im} T$. Suppose then that

$$0 = \alpha_{m+1}T(x_{m+1}) + \ldots + \alpha_n T(x_n) = T(\alpha_{m+1}x_{m+1} + \ldots + \alpha_n x_n).$$

Then $\alpha_{m+1}x_{m+1} + \ldots + \alpha_n x_n \in \operatorname{Ker} T$, so we can express it as a linear combination of X:

$$\alpha_{m+1}x_{m+1} + \ldots + \alpha_n x_n = \alpha_1 x_1 + \ldots + \alpha_m x_m.$$

But the independence of X then implies, in particular, $\alpha_{m+1} = \ldots = \alpha_n = 0$. Hence Y is independent.

Now let $w \in \operatorname{Im} T$, so that $w = T(v)$, for some $v \in V$. We can then express v as a linear combination of $X : v = \beta_1 x_1 + \dots \beta_n x_n$. Then

$$w = T(v) = T(\beta_1 x_1 + \dots \beta_n x_n) = \beta_1 T(x_1) + \dots \beta_n T(x_n) = \beta_{m+1} T(x_{m+1}) + \dots \beta_n T(x_n),$$

since $x_1, \dots, x_m \in \operatorname{Ker} T$. Thus Y spans $\operatorname{Im} T$ as well. □

Corollary 2.19. *Let $T : V \to W$ be a linear transformation, and let $\dim V = n = \dim W$. Then*

$$T \text{ is injective} \iff n(T) = 0 \iff r(T) = n \iff T \text{ is surjective.}$$

□

3. The Matrix of a Linear Transformation

Remark 3.1. Throughout this section, all vector spaces will be finite dimensional. Also, we will maintain the *order* of the elements in any basis.

Notation 3.2. Let $\mathcal{B} = \{x_1, \dots, x_n\}$ be a basis of a vector space V over the field F. Then by Theorem 4.14, every $x \in V$ has a unique representation

$$x = \sum \alpha_i x_i, \text{ where } \alpha_i \in F.$$

That is, there is a one to one correspondence between V and F^n given by

$$x \longleftrightarrow \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \doteq [x]_{\mathcal{B}}.$$

Note that we are writing the elements of F^n in a column rather than a row; the reason we do this will be apparent shortly.

Example 3.3. Let $\mathcal{B} = \{e_1, e_2, \dots, e_n\}$ be the standard basis of F^n, and let $x = (\alpha_1, \alpha_2, \dots, \alpha_n) \in F^n$. Then

$$[x]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Example 3.4. Consider the basis $\mathcal{B} = \{(1,1), (1,2)\}$ of $\mathbb{R}^2$. Then

$$[(1,1)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } [(2,3)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Example 3.5. Let $\mathcal{B} = \{1, x, \ldots, x^n\}$ be the standard basis of $\mathcal{P}_n(F)$. Then

$$[\alpha_0 + \alpha_1 x + \ldots + \alpha_n x^n]_{\mathcal{B}} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in F^{n+1}.$$

Example 3.6. Consider the basis $\mathcal{B} = \{1, 1+x, 1+x+x^2\}$ of $\mathcal{P}_2(\mathbb{R})$. Then

$$[6 + 5x + 3x^2]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Definition 3.7. Let $T : V \to W$ be a linear transformation, and let $\mathcal{B} = \{x_1, \ldots, x_n\}$ and $\mathcal{C} = \{y, \ldots, y_m\}$ be bases of V and W respectively. For each $j = 1, \ldots, n$, let

$$[T(x_j)]_{\mathcal{C}} = \begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{mj} \end{pmatrix}, \text{ where } \alpha_{ij} \in F.$$

The matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$[T]_{\mathcal{B}}^{\mathcal{C}} \doteq (\alpha_{ij}) \in M_{m \times n}(F).$$

Example 3.8. Define $T : F^3 \to F^2$ by $T(x, y, z) = (x + y, y - z)$, and let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases of F^3 and F^2 respectively. Then

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in M_{2\times 3}(F).$$

Example 3.9. Define $T : \mathcal{P}_4(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ by $T(f(x)) = f'(x)$, and let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases of $\mathcal{P}_4(\mathbb{R})$ and $\mathcal{P}_3(\mathbb{R})$ respectively. Then

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \in M_{4\times 5}(F).$$

Theorem 3.10. *Let $T : V \to W$ be a linear transformation, let $\mathcal{B} = \{x_1, \ldots, x_n\}$ and $\mathcal{C} = \{y, \ldots, y_m\}$ be bases of V and W respectively, and let $x \in V$. Then*

$$[T]_{\mathcal{B}}^{\mathcal{C}} \cdot [x]_{\mathcal{B}} = [T(x)]_{\mathcal{C}}.$$

Proof. Let $[T]_{\mathcal{B}}^{\mathcal{C}} = (\alpha_{ij}), x = \sum_{j=1}^{n} \beta_j x_j$, and $T(x) = \sum_{i=1}^{m} \gamma_i y_i$. Then

$$\begin{aligned} T(x) &= T\Big(\sum_{j=1}^{n} \beta_j x_j\Big) \\ &= \sum_{j=1}^{n} \beta_j T(x_j) \\ &= \sum_{j=1}^{n} \beta_j \Big(\sum_{i=1}^{m} \alpha_{ij} y_i\Big) \\ &= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} \alpha_{ij} \beta_j\Big) y_i. \end{aligned}$$

Therefore, by the uniqueness of representation as a linear combination of $\mathcal{C}$, we conclude that

$$\gamma_i = \sum_{j=1}^{n} \alpha_{ij} \beta_j, \text{ for } i = 1, \ldots, m.$$

□

Example 3.11. Referring to Example 3.8, we see that $T(-1, 2, 4) = (1, -2)$, and

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

Example 3.12. Referring to Example 3.9, we see that $T(-3 - x^2 + 4x^4) = -2x + 16x^3$, and

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \cdot \begin{pmatrix} -3 \\ 0 \\ -1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 16 \end{pmatrix}.$$

Remark 3.13. It's easy, but a bit tedious, to check that composition of linear transformations corresponds to matrix multiplication. To be precise: let $T : V \to W$ and $U : W \to X$ be linear transformations, where $\dim_F V = n, \dim_F W = m$, and $\dim_F X = p$. Let $\mathcal{B}, \mathcal{C}$, and $\mathcal{D}$ be bases of V, W, and X respectively. Then

$$[UT]_{\mathcal{B}}^{\mathcal{D}} = [U]_{\mathcal{C}}^{\mathcal{D}} \cdot [T]_{\mathcal{B}}^{\mathcal{C}}.$$

Example 3.14. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x, y, z) = (x + y, y, x - y)$ and $U : \mathbb{R}^3 \to \mathbb{R}^2$ by $U(x, y, z) = (2x + y, x - 3y)$; then $UT(x, y) = (2x + 3y, x - 2y)$. Let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively. We see that

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } [U]_{\mathcal{C}}^{\mathcal{B}} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -3 & 0 \end{pmatrix}$$

so that

$$[U]_{\mathcal{C}}^{\mathcal{B}} \cdot [T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 2 & 3 \\ 3 & -2 \end{pmatrix} = [UT]_{\mathcal{B}}.$$

4. Isomorphisms

Definition 4.1. A bijective linear transformation $T : V \to W$ is called an isomorphism. We say that V and W are isomorphic, and write $V \cong W$.

Theorem 4.2. *Let V and W be finite dimensional vector spaces over F. Then*

$$V \cong W \iff \dim_F V = \dim_F W.$$

Proof. ($\implies$) Let $T : V \to W$ be an isomorphism. then

$$\begin{aligned} \dim_F V &= n(T) + r(T) \\ &= r(T), \text{ since } T \text{ is injective} \\ &= \dim_F W, \text{ since } T \text{ is surjective.} \end{aligned}$$

($\impliedby$) Let $\{x_1, \ldots, x_n\}$ be a basis of V and $\{y_1, \ldots, y_n\}$ a basis of W. We must construct an isomorphism $T : V \to W$. Begin by defining

$$T(x_i) = y_i, \text{ for } i = 1, \ldots, n.$$

Now extend linearly: if $x \in V$, write $x = \sum_{i=1}^n \alpha_i x_i$, and define

$$T(x) = \sum_{i=1}^n \alpha_i T(x_i) = \sum_{i=1}^n \alpha_i y_i.$$

Clearly, T is a linear transformation (essentially, by the way we defined it). Also, T is surjective, since

$$y \in W \implies y = \sum_{i=1}^n \beta_i y_i = T\Big(\sum_{i=1}^n \beta_i x_i\Big).$$

Finally, T is injective since

$$T\Big(\sum_{i=1}^n \alpha_i x_i\Big) = 0 \implies \sum_{i=1}^n \alpha_i y_i = 0 \implies \alpha_i = 0, \forall i.$$

$\square$

Corollary 4.3. *Any vector space over F of dimension n is isomorphic to F^n.*

Proof. $\dim_F F^n = n$. $\square$

Remark 4.4. For vector spaces V of dimension n and W of dimension m, let

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

We can easily make $\mathcal{L}(V, W)$ into a vector space over F itself, by defining

$$(T + U)(x) = T(x) + U(x), \forall x \in V$$

$$(\alpha T)(x) = \alpha(T(x)), \forall \alpha \in F, \forall x \in V.$$

Then if we fix bases $\mathcal{B}$ of V and $\mathcal{C}$ of W, we get an isomorphism

$$\mathcal{T} : \mathcal{L}(V, W) \to M_{m\times n}(F) \text{ defined by } \mathcal{T}(T) = [T]_{\mathcal{B}}^{\mathcal{C}}.$$

Thus $\dim_F \mathcal{L}(V, W) = mn$. *You should check all the missing details!)*

Proposition 4.5. *If $T : V \to W$ is an isomorphism, then $T^{-1} : W \to V$ is also an isomorphism.*

Proof. We know that the inverse function T^{-1} exists by Theorem 1.12. The issue is whether the inverse is linear. So let $w, w' \in W$. Then since T is surjective, there exist $v, v' \in V$ with $T(v) = w, T(v') = w'$. Then

$$T(v + v') = T(v) + T(v') = w + w' \implies T^{-1}(w + w') = v + v' = T^{-1}(w) + T^{-1}(w').$$

Similarly, if $\alpha \in F$, then

$$T(\alpha v) = \alpha T(v) = \alpha w \implies T^{-1}(\alpha w) = \alpha v = \alpha T^{-1}(w).$$

□

Remark 4.6. If $T : V \to W$ is an isomorphism, and $\mathcal{B}, \mathcal{C}$ are bases of V, W, then

$$[T^{-1}]_{\mathcal{C}}^{\mathcal{B}} \cdot [T]_{\mathcal{B}}^{\mathcal{C}} = [i_V]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

and

$$[T]_{\mathcal{B}}^{\mathcal{C}} \cdot [T^{-1}]_{\mathcal{C}}^{\mathcal{B}} = [i_W]_{\mathcal{C}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

These identity matrices are of course actually the same. They are square: $n \times n$, where $n = \dim_F V = \dim_F W$.

5. The Change of Basis Matrix

Discussion 5.1. Suppose we have a vector space V over F of dimension n, and two bases

$$\mathcal{B} = \{x_1, \ldots, x_n\} \text{ and } \mathcal{B}' = \{x'_1, \ldots, x'_n\}.$$

Then if $x \in V$, we get two expressions

$$x = \alpha_1 x_1 + \ldots + \alpha_n x_n = \alpha'_1 x'_1 + \ldots + \alpha'_n x'_n.$$

What's the connection between

$$[x]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in F^n \text{ and } [x]_{\mathcal{B}'} = \begin{pmatrix} \alpha'_1 \\ \vdots \\ \alpha'_n \end{pmatrix} \in F^n?$$

Consider the identity linear transformation $I : V \to V$ defined by $I(x) = x, \forall x \in V$. We can construct

$$[I]_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ & \vdots & \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} \in M_n(F),$$

and conclude that

$$[I]_{\mathcal{B}'}^{\mathcal{B}}[x]_{\mathcal{B}'} = [x]_{\mathcal{B}}.$$

Moreover, we could similarly construct $[I]_{\mathcal{B}}^{\mathcal{B}'} \in M_n(F)$, and then

$$\left([I]_{\mathcal{B}}^{\mathcal{B}'}[I]_{\mathcal{B}'}^{\mathcal{B}}\right)[x]_{\mathcal{B}'} = [I]_{\mathcal{B}}^{\mathcal{B}'}[x]_{\mathcal{B}} = [x]_{\mathcal{B}'}$$

and

$$\left([I]_{\mathcal{B}'}^{\mathcal{B}}[I]_{\mathcal{B}}^{\mathcal{B}'}\right)[x]_{\mathcal{B}} = [I]_{\mathcal{B}'}^{\mathcal{B}}[x]_{\mathcal{B}'} = [x]_{\mathcal{B}}.$$

Thus these two matrices serve to change the expression of x in one basis into the corresponding expression of x in the other basis.

Example 5.2. Let $V = \mathbb{R}^2$, $\mathcal{B} = \{(1,0),(0,1)\}$, and $\mathcal{B}' = \{(1,1),(1,2)\}$. Then

$$[I]_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \text{ and } [I]_{\mathcal{B}}^{\mathcal{B}'} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

As an example, consider $x = (-1,0) \in \mathbb{R}^2$. Then

$$[(-1,0)]_{\mathcal{B}} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \text{ and } [(-1,0)]_{\mathcal{B}'} = \begin{pmatrix} -2 \\ 1 \end{pmatrix},$$

so

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Notice that

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

Discussion 5.3. Noe suppose we have a linear transformation $T : V \to V$. (T is called a linear operator.) What's the connection between the two matrices

$$[T]_{\mathcal{B}'}^{\mathcal{B}} \in M_n(F) \text{ and } [T]_{\mathcal{B}}^{\mathcal{B}'} \in M_n(F)?$$

Well, remember that

$$[T]_{\mathcal{B}}[x]_{\mathcal{B}} = [T(x)]_{\mathcal{B}} \text{ and } [T]_{\mathcal{B}'}[x]_{\mathcal{B}'} = [T(x)]_{\mathcal{B}'},$$

so

$$\left([I]_{\mathcal{B}}^{\mathcal{B}'}[T]_{\mathcal{B}}[I]_{\mathcal{B}'}^{\mathcal{B}}\right)[x]_{\mathcal{B}'} = \left([I]_{\mathcal{B}}^{\mathcal{B}'}[T]_{\mathcal{B}}\right)[x]_{\mathcal{B}} = [I]_{\mathcal{B}}^{\mathcal{B}'}[T(x)]_{\mathcal{B}} = [T(x)]_{\mathcal{B}'}.$$

But this means that

$$[T]_{\mathcal{B}'} = [I]_{\mathcal{B}}^{\mathcal{B}'}[T]_{\mathcal{B}}[I]_{\mathcal{B}'}^{\mathcal{B}}.$$

A similar argument (or just multiplying both sides by the change of basis matrices appropriately) shows that

$$[T]_{\mathcal{B}} = [I]_{\mathcal{B}'}^{\mathcal{B}}[T]_{\mathcal{B}'}[I]_{\mathcal{B}}^{\mathcal{B}'}.$$

We say that the two different matrices of T are similar.

Example 5.4. Let T be the linear operator on $\mathbb{R}^2$ defined by $T(x,y) = (x+y, x-y)$. Then with $\mathcal{B}$ and $\mathcal{B}'$ as in the previous example,

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

so

$$[T]_{\mathcal{B}'} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ -2 & -4 \end{pmatrix}.$$

To verify that this is correct, observe that

$$T(1,1) = (2,0) = 4(1,1) + (-2)(1,2)$$

$$T(1,2) = (3,-1) = 7(1,1) + (-4)(1,2).$$

6. Exercises

Exercise 6.1. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x,y) = (x+y, 0, 2x-y)$. Show that T is a linear transformation and find bases for the kernel $\operatorname{Ker} T$ and the image $\operatorname{Im} T = T(\mathbb{R}^2)$.

Exercise 6.2. Define $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ by $T(f(x)) = xf(x) + f'(x)$. Show that T is a linear transformation and find bases for the kernel $\operatorname{Ker} T$ and the image $\operatorname{Im} T = T(\mathcal{P}_2(\mathbb{R}))$.

Exercise 6.3. Let $T : V \to W$ be an injective linear transformation, and let $X \subseteq V$ be linearly independent. Show that $T(X) = \{T(v) : v \in X\}$ is a linearly independent subset of W.

Exercise 6.4. Let $T : V \to V$ be a linear transformation. Show that the following are equivalent:

(1) $\operatorname{Ker} T \cap \operatorname{Im} T = \{0\}$.
(2) if $T(T(v)) = 0$, then $T(v) = 0$.

Exercise 6.5. Let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. For each linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, compute $[T]_{\mathcal{B}}^{\mathcal{C}}$.

(a) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x,y,z) = (2x+3y-z, x+z)$.
(b) $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(x_1, x_2, \ldots, x_n) = (x_n, x_{n-1}, \ldots, x_1)$.

Exercise 6.6. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $T(x, y) = (x - y, x, 2x + y)$. Let $\mathcal{B}$ be the standard basis of $\mathbb{R}^2, \mathcal{C} = \{(1, 1, 0), (0, 1, 1), (2, 2, 3)\}$, and $\mathcal{D} = \{(1, 2), (2, 3)\}$.

(a) Compute $[T]_{\mathcal{B}}^{\mathcal{C}}$.
(b) Compute $[T]_{\mathcal{D}}^{\mathcal{C}}$.

Exercise 6.7. Define $T : M_{2\times 2}(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ by $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a + b) + 2dx + bx^2$. Let $\mathcal{B}$ and $\mathcal{C}$ be the standard bases of $M_{2\times 2}$ and $\mathcal{P}_2(\mathbb{R})$ respectively. Compute $[T]_{\mathcal{B}}^{\mathcal{C}}$.

Exercise 6.8. Let V be an n-dimensional vector space over F with basis $\mathcal{B}$. Show that $T : V \to F^n$ defined by $T(x) = [x]_{\mathcal{B}}$ is an isomorphism.

Exercise 6.9. Let $\mathcal{B}$ and $\mathcal{C}$ be bases of the vector spaces V and W over the field F respectively. Suppose that $\dim_F V = n$ and $\dim_F W = m$. Show that $\mathcal{T} : \mathcal{L}(V, W) \to M_{m\times n}(F)$ defined by $\mathcal{T}(T) = [T]_{\mathcal{B}}^{\mathcal{C}}$ is an isomorphism.

Exercise 6.10. A square matrix $(\alpha_{ij}) \in M_n(F)$ is *diagonal* if $\alpha_{ij} = 0$ unless $i = j$. Let V and W be vector spaces with $\dim_F V = \dim_F W$, and let $T : V \to W$ be a linear transformation. Show that there exist bases $\mathcal{B}$ and $\mathcal{C}$ of V and W respectively such that $[T]_{\mathcal{B}}^{\mathcal{C}}$ is diagonal.

Exercise 6.11. For each of the following pairs of bases $\mathcal{B}$ and $\mathcal{B}'$ of the indicated vector space V, find the change of basis matrix $[I_V]_{\mathcal{B}}^{\mathcal{B}'}$.

(a) $\mathcal{B} = \{(-4, 3), (2, -1)\}, \mathcal{B}' = \{(2, 1), (-4, 1)\}, V = \mathbb{R}^2$
(b) $\mathcal{B} = \{x^2 - x + 1, x + 1, x^2 + 1\}, \mathcal{B}' = \{x^2 + x + 4, 4x^2 - 3x + 2, 2x^2 + 3\}, V = \mathcal{P}_2(\mathbb{R})$

CHAPTER IV

Matrices

1. Elementary Operations

Definition 1.1. Let $A \in M_{m\times n}(F)$. We define three elementary row operations as follows:

Type 1: for $i \neq j$, switch each element in row i with the element in row j in the same column;
Type 2: multiply each element in row i by $0 \neq \alpha \in F$;
Type 3: for $i \neq j$, add each element in row i to the element in row j in the same column.

Remark 1.2. Informally, the three operations are to switch two rows, to multiply a row by a nonzero scalar, and to add one row to another. We can and do define corresponding operations on the columns of A, but these are less natural, as we will now see.

Motivation 1.3. Suppose we have a system of linear equations:

$$\alpha_{11}x_1 + \ldots + \alpha_{1n}x_n = \beta_1$$
$$\vdots$$
$$\alpha_{m1}x_1 + \ldots + \alpha_{mn}x_n = \beta_m$$

If we define

$$A = (\alpha_{ij}) \in M_{m\times n}(F), X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n, \text{ and } B = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \in F^m,$$

then the entire system becomes one matrix equation $AX = B$. If we think of X as an element in the set of solutions to the system, then the elementary row operations preserve that set.

Example 1.4. Here is an illustration of how we can modify $A \in M_{m\times n}(F)$ with row operations so that the solutions of the associated system can be easily read. For simplicity, we will often do several operations at once.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 6 \\ -1 & 0 & -2 \\ 1 & 5 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 3 & 3 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3/2 \\ 0 & 2 & 1 \\ 0 & 3 & 3 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3/2 \\ 0 & 0 & -2 \\ 0 & 0 & -3/2 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3/2 \\ 0 & 0 & 1 \\ 0 & 0 & -3/2 \end{pmatrix}$$

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

A system of equations with this matrix A would also have a matrix B. Suppose for example we want to solve

$$AX = A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 11 \\ -3 \\ 12 \end{pmatrix}.$$

We would use an augmented matrix and perform the same sequence of row operations:

$$(A|B) = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 1 & 4 & 6 & 11 \\ -1 & 0 & -2 & -3 \\ 1 & 5 & 6 & 12 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we may just read off the solution $x_1 = x_2 = x_3 = 1$.

Exercise 1.5. What happens if

$$B = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}?$$

Definition 1.6. An elementary matrix $E \in M_n(F)$ is one that is obtained by performing a single elementary row operation on the identity matrix

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Example 1.7. Here are three elementary 3×3 matrices, corresponding to the three elementary row operations:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Remark 1.8. Notice that an elementary row matrix can be thought of as having been obtained by performing the corresponding *column* operation on I_n. We are thus justified into referring to them simply as elementary matrices.

Theorem 1.9. *Let $E \in M_n(F)$ be elementary, and let $A \in M_n(F)$. Then EA is the matrix obtained by performing the row operation of E on A.*

Proof. We prove the assertion for a switch of rows $i < j$; the other two operations can be verified similarly.

$$EA = \begin{array}{c} \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & 0 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \\ \cdots & 1 & \cdots & 0 & \cdots \\ & \vdots & & \vdots & \end{pmatrix} \end{array} \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & \alpha_{ii} & \cdots & \alpha_{ij} & \cdots \\ & \vdots & & \vdots & \\ \cdots & \alpha_{ji} & \cdots & \alpha_{jj} & \cdots \\ & \vdots & & \vdots & \end{pmatrix} \end{array} \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \end{array} = \begin{array}{c} \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & \alpha_{ji} & \cdots & \alpha_{jj} & \cdots \\ & \vdots & & \vdots & \\ \cdots & \alpha_{ii} & \cdots & \alpha_{ij} & \cdots \\ & \vdots & & \vdots & \end{pmatrix} \end{array} \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \end{array}$$

$\square$

Exercise 1.10. Show that AE is the matrix obtained by performing the *column* operation of E on A.

Definition 1.11. A matrix $A \in M_n(F)$ is invertible if there is a matrix $B \in M_n(F)$ such that $AB = BA = I_n$. The matrix B is the inverse of A, and is denoted A^{-1}.

Example 1.12. The matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if $ad - bc \neq 0$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}.$$

But A is not invertible if $ad - bc = 0$. For example,

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \implies x + z = 1 \text{ yet } 2x + 2z = 0.$$

Proposition 1.13. *Elementary matrices are invertible.*

Proof. We treat the three types separately.

1. If E switches rows i and j, then clearly $EE = I_n$.
2. If E multiplies row i by $\alpha \neq 0$, let G be the elementary matrix that multiplies row i by α^{-1}. Then $EG = GE = I_n$.
3. If E adds row i to row j, let H be the elementary matrix that multiplies row i by -1. Then $(HEH)E = E(HEH) = I_n$.

$\square$

2. The Rank of a Matrix

Remark 2.1. Given $A \in M_{m\times n}(F)$, we can define an associated linear transformation $L_A : F^n \to F^m$, given by

$$L_A(x) = Ax = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Example 2.2. If $A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 3 & 1 \end{pmatrix} \in M_{2\times 3}(\mathbb{R})$, then

$$L_A(x) = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_1 + 3x_2 + x_3 \end{pmatrix} \in \mathbb{R}^2.$$

Definition 2.3. The rank of $A \in M_{m\times n}(F)$ is

$$r(A) = \dim_F L_A = r(L_A).$$

Proposition 2.4. *Let $A \in M_n(F)$. Then A is invertible $\iff$ L_A is invertible $\iff$ L_A is an isomorphism $\iff$ $r(A) = n$.*

Proof. By the Rank-Nullity Theorem, we need only prove the first equivalence. Suppose then that A is invertible. For every $y \in F^n$, we have

$$L_A(A^{-1}y) = AA^{-1}y = I_n y = y,$$

so L_A is surjective and therefore bijective by the Rank-Nullity Theorem.
Conversely, if we let $\mathcal{B}$ be the standard basis of F^n, then it easy to see that $[L_A]_{\mathcal{B}} = A$. Thus we have

$$I_n = [I_{F^n}]_{\mathcal{B}} = [L_A L_A^{-1}]_{\mathcal{B}} = [L_A]_{\mathcal{B}}[L_A^{-1}]_{\mathcal{B}} = A[L_A^{-1}]_{\mathcal{B}}.$$

Similarly, $I_n = [L_A^{-1}]_{\mathcal{B}} A$, so A is invertible. $\square$

Lemma 2.5. *Let $A \in M_{m\times n}(F)$, and let $P \in M_m(F)$ and $Q \in M_n(F)$ be invertible. Then*

1) $r(AQ) = r(A)$;
2) $r(PA) = r(A)$;
3) $r(PAQ) = r(A)$.

Proof. $r(AQ) = \dim(\operatorname{Im} L_{AQ}) = \dim(L_{AQ}(F^n)) = \dim(L_A L_Q(F^n)) = \dim(L_A(F^n)) = \dim \operatorname{Im} L_A = r(A)$. We leave the other two statements as exercises. □

Lemma 2.6. *Elementary row and column operations preserve the rank of a matrix.*

Proof. A row (column) operation can be viewed as multiplying on the left (right) by the corresponding elementary matrix, which is invertible. □

Theorem 2.7. *Let $A \in M_{m\times n}(F)$. Then $r(A)$ is the maximum number of columns of A that form a linearly independent set in F^m.*

Proof. Let $\mathcal{B} = \{e_1, \dots, e_n\}$ be the standard basis of F^n. Then

$$\operatorname{Im} L_A = \operatorname{Span}\{L_A(e_1), \dots, L_A(e_n)\}.$$

But if $A = (\alpha_{ij})$, then

$$L_A(e_j) = A \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix},$$

the j-th column of A. □

Theorem 2.8. *Let $0 \neq A \in M_{m\times n}(F)$ with $r(A) = r$. Then using row and column operations, we can transform A into*

$$\left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

Proof. We use induction on m, the number of rows of A. If $m = 1$, then $A = \begin{pmatrix} \alpha_1 & \cdots & \alpha_j & \cdots & \alpha_n \end{pmatrix}$, where $\alpha_j \neq 0$. So we proceed:

$$\begin{aligned} \begin{pmatrix} \alpha_1 & \cdots & \alpha_j & \cdots & \alpha_n \end{pmatrix} &\to \begin{pmatrix} \alpha_j & \cdots & \alpha_1 & \cdots & \alpha_n \end{pmatrix} \\ &\to \begin{pmatrix} 1 & \cdots & \frac{\alpha_1}{\alpha_j} & \cdots & \frac{\alpha_n}{\alpha_j} \end{pmatrix} \\ &\to \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \end{pmatrix}. \end{aligned}$$

Assume for induction that any $(m-1)\times n$ matrix can be transformed as desired, and let $A \in M_{m\times n}(F)$ where $m > 1$ and and at least one entry $\alpha_{ij} \neq 0$. Then

$$\begin{aligned}
\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \\ \vdots & & \vdots & & \vdots \\ \alpha_{i1} & \cdots & \alpha_{ij} & \cdots & \alpha_{in} \\ \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mj} & \cdots & \alpha_{mn} \end{pmatrix} &\to \begin{pmatrix} \alpha_{ij} & \cdots & \alpha_{i1} & \cdots & \alpha_{1n} \\ \vdots & & \vdots & & \vdots \\ \alpha_{1j} & \cdots & \alpha_{11} & \cdots & \alpha_{in} \\ \vdots & & \vdots & & \vdots \\ \alpha_{mj} & \cdots & \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \\
&\to \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & B \end{array}\right) \\
&\to \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline 0 & I_{r-1} & 0 \\ \hline 0 & 0 & 0 \end{array}\right) \\
&= \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).
\end{aligned}$$

□

Example 2.9. This theorem gives us an algorithm to calculate the rank of any matrix. For example,

$$\begin{aligned}
A = \begin{pmatrix} 1 & 2 & -1 & 5 \\ -2 & 0 & 6 & -7 \\ 0 & 4 & 4 & 3 \end{pmatrix}
&\to \begin{pmatrix} 1 & 2 & -1 & 5 \\ 0 & 4 & 4 & 3 \\ 0 & 4 & 4 & 3 \end{pmatrix} \\
&\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 4 & 3 \\ 0 & 4 & 4 & 3 \end{pmatrix} \\
&\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \frac{3}{4} \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&\to \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right).
\end{aligned}$$

Thus $r(a) = 2$.

Notation 2.10. If $A \in M_{m\times n}(F)$, it's sometimes convenient to write

$$A = \begin{pmatrix} C_1 & \cdots & C_n \end{pmatrix} = \begin{pmatrix} R_1 \\ \vdots \\ R_m \end{pmatrix},$$

where $C_j \in F^m$ (or $M_{m\times 1}(F)$) is the j-th column of A and where $R_i \in F^n$ (or $M_{1\times n}(F)$) is the i-th row. We see then that the transpose becomes

$$A^t = \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} = \begin{pmatrix} R_1 & \cdots & R_m \end{pmatrix}.$$

A more important application of this notation comes when we multiply two matrices. For clarity, we will denote the rows and columns of a matrix A simply as vectors, and let the shape of the expression make it obvious which we mean. That is, if $A \in M_{m\times n}$ and $B \in M_{n\times p}$, then

$$AB = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} \begin{pmatrix} B_1 & \cdots & B_p \end{pmatrix} = \begin{pmatrix} A_1 \bullet B_1 & \cdots & A_1 \bullet B_p \\ & \vdots & \\ A_m \bullet B_1 & \cdots & A_m \bullet B_p \end{pmatrix} = (A_i \bullet B_k),$$

where $i = 1, \ldots, m, k = 1, \ldots, p$, and $\bullet$ is simply the familiar dot product in F^n.

Proposition 2.11. *$A \in M_{m\times n}$ and $B \in M_{n\times p}$. Then $(AB)^t = B^t A^t$.*

Proof.

$$\begin{aligned}
(AB)^t &= \left(\begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} \begin{pmatrix} B_1 & \cdots & B_p \end{pmatrix} \right)^t \\
&= \begin{pmatrix} A_1 \bullet B_1 & \cdots & A_1 \bullet B_p \\ & \vdots & \\ A_m \bullet B_1 & \cdots & A_m \bullet B_p \end{pmatrix}^t \\
&= \begin{pmatrix} A_1 \bullet B_1 & \cdots & A_m \bullet B_1 \\ & \vdots & \\ A_1 \bullet B_p & \cdots & A_m \bullet B_p \end{pmatrix} \\
&= \begin{pmatrix} B_1 \bullet A_1 & \cdots & B_1 \bullet A_m \\ & \vdots & \\ B_p \bullet A_m & \cdots & B_p \bullet A_m \end{pmatrix} \\
&= \begin{pmatrix} B_1 \\ \vdots \\ B_p \end{pmatrix} \begin{pmatrix} A_1 & \cdots & A_m \end{pmatrix} \\
&= B^t A^t.
\end{aligned}$$

□

Proposition 2.12. *Let $A \in M_{m\times n}$. have rank r. Then $r(A^t) = r$.*

Proof. We can find invertible matrices $P \in M_m(F)$ and $Q \in M_n(F)$ such that

$$PAQ = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array} \right).$$

Then

$$Q^t A^t P^t = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array} \right).$$

Since Q^t and P^t are also invertible, the result follows. □

Remark 2.13. This last proposition can be stated in a remarkable way: *the number of linearly independent columns in a matrix is the same as the number of linearly independent rows!!*

Application 2.14. We're now in a position to develop an inefficient but interesting procedure for calculating the inverse of a matrix $A \in M_n(F)$. First, we build the augmented $n \times 2n$ matrix $(A|I_n)$, and notice that

$$A^{-1}(A|I_n) = (A^{-1}A|A^{-1}I_n) = (I_n|A).$$

If we express the inverse as the product of elementary matrices $A^{-1} = E_k \cdots E_1$, then

$$A^{-1}A = I_n = E_k \cdots E_1 A.$$

So we could proceed as follows:

- Use row operations to transform A into I_n.
- Perform that sequence of operations on I_n.
- Notice that the result is to transform I_n into A^{-1}.

Example 2.15. Let $A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$.

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -1 & -1 & 1 \end{array}\right)$$
$$\to \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -1 \end{array}\right)$$
$$\to \left(\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{array}\right)$$

Thus $\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$.

3. Systems of Equations

Definition 3.1. Let $A = (\alpha_{ij}) \in M_{m\times n}(F)$, $B = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \in F^m$, and $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. A solution to the system $AX = B$ is a vector $x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in F^n$ such that $Ax = B$.

Proposition 3.2. *The set of all solutions of a homogeneous system $AX = 0$ is a subspace of F^n.*

Proof. The set is just $\operatorname{Ker} L_A$. □

Proposition 3.3. *The set of all solutions of a nonhomogeneous system $AX = B$ is*

$$\mathscr{S} = a + S = \{a + x : x \in S\},$$

where a is any solution of the system (that is, $Aa = B$) and S is the solution space of the homogeneous system $AX = 0$.

Proof. If $x \in S$, then

$$A(a + x) = Aa + Ax = B + 0 = B,$$

so $a + S \subseteq \mathscr{S}$. Conversely, if $y \in \mathscr{S}$, then

$$A(y - a) = Ay - Aa = B - B = 0,$$

so $y - a \in S$. Thus $y \in a + S$ and hence $\mathscr{S} \subseteq a + S$. $\square$

Example 3.4. Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \end{pmatrix} \in M_{2\times 3}(\mathbb{R})$ and $B = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \in \mathbb{R}^2$. To solve $AX = B$, we first solve the homogeneous system $AX = 0$:

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

Thus $S = \operatorname{Span}\{(-1, 1, 1)\}$. We now need a particular solution of the nonhomogeneous system $AX = B$, and any one will do:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 2 & 1 & 1 & 6 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & -3 & 3 & 0 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & 0 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 0 \end{array}\right).$$

Now it's easy to see that $(3, 0, 0)$ works, so

$$\mathscr{S} = (3, 0, 0) + \operatorname{Span}\{(-1, 1, 1)\} = \{(3 - t, t, t) : t \in \mathbb{R}\}.$$

Remark 3.5. Notice that the solution space S of the homogeneous system in this last example is a line through the origin, a one-dimensional subspace of $\mathbb{R}^3$. This is because $r(A) = 2$ so $n(A) = 3 - 2 = 1$. But the solution set $\mathscr{S}$ of the nonhomogeneous system is not a subspace, but rather the set S translated by the vector $(3, 0, 0)$. A different particular solution will give a different translation, but result in the same set!

Theorem 3.6. *The system $AX = B$ has a solution (is consistent) if and only if $r(A) = r(A|B)$.*

Proof. Clearly, the system has a solution if and only if $B \in \operatorname{Im} L_A$. But if $A = \begin{pmatrix} A_1 & \cdots & A_n \end{pmatrix}$, then

$$\operatorname{Im} L_A = \operatorname{Span}\{A_1, \cdots, A_n\} \subseteq \operatorname{Span}\{A_1, \cdots, A_n, B\} = \operatorname{Im} L_{(A|B)}.$$

Thus these linear transformation have the same rank if and only if $B \in \operatorname{Span}\{A_1, \cdots, A_n\}$. $\square$

Example 3.7. Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 2 \\ 1 & -4 & 7 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$ and $B = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^3$. To decide if $AX = B$ is consistent, we simultaneously compute $r(A)$ and $r(A|B)$:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 2 & 1 & 2 & 3 \\ 1 & -4 & 7 & 4 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & -3 & 4 & 1 \\ 0 & -6 & 8 & 3 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & -3 & 4 & 1 \\ 0 & -6 & 8 & 3 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & -3 & 4 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

Since $r(A) = 2$ but $r(A|B) = 3$, the system is inconsistent; that is, there are no solutions.

4. Determinants

Definition 4.1. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F)$. The determinant of A is $\det A = ad - bc$.

Proposition 4.2. *The determinant is multiplicative:* $\det(AB) = \det A \cdot \det B, \forall A, B \in M_2(F)$.

Proof.

$$\begin{aligned}
\det A \cdot \det B &= \det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) \\
&= \det \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} \\
&= (ae + bg)(cf + dh) - (af + bh)(ce + dg) \\
&= aecf + aedh + bgcf + bgdh - afce - afdg - bhce - bhdg \\
&= (ad - bc)(eh - fg) \\
&= \det A \cdot \det B.
\end{aligned}$$

□

Remark 4.3. The determinant is *not* additive. For example,

$$\det \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \det \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0 \text{ but } \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

So the function $\det : M_2(F) \to F$ is *not* a linear transformation.

Proposition 4.4. *The determinant is linear in each row and each column. As one example,*

$$\det(A_1 + \alpha A_1' \ \ A_2) = \det(A_1 \ \ A_2) + \alpha \det(A_1' \ \ A_2).$$

Proof.

$$\begin{aligned}\det\begin{pmatrix} a+\alpha a' & b \\ c+\alpha c' & d \end{pmatrix} &= (a+\alpha a')d - b(c+\alpha c') \\ &= (ad-bc) + \alpha(a'd - bc') \\ &= \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \alpha \det\begin{pmatrix} a' & b \\ c' & d \end{pmatrix}.\end{aligned}$$

The proofs of linearity in the other column and in each of the two rows are similar. □

Proposition 4.5. $\det A \neq 0 \iff A$ *is invertible.*

Proof. ($\Longrightarrow$) It's easy to check that

$$\begin{pmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$$

is an inverse for $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

($\Longleftarrow$)

$$1 = \det I_2 = \det(AA^{-1}) = \det A \cdot \det A^{-1} \implies \det A \neq 0.$$

□

Remark 4.6. We would now like to define the determinant for larger square matrices. Once we do that, we'll want to check that the important properties we've seen in the 2×2 case still hold.

Definition 4.7. Let $A \in M_n(F)$ where $n > 1$. For each $1 \leq i, j \leq n$, $\overline{A_{ij}} \in M_{(n-1)\times(n-1)}(F)$ is the matrix obtained by deleting row i and column j of A.

Definition 4.8. Let $A \in M_n(F)$. We inductively define the determinant of A as follows:

- if $n = 1$, then $\det A = \det(\alpha) = \alpha$;
- if $n > 1$, then

$$\det A = \det(\alpha_{ij}) = \sum_{j=1}^{n} (-1)^{1+j} \alpha_{1j} \det \overline{A_{1j}}.$$

Example 4.9. We should check that this definition does in fact generalize our definition in the 2×2 case:

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

Example 4.10.

$$\begin{aligned}
\det\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} &= 1\cdot\det\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2\cdot\det\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3\cdot\det\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\
&= 1\cdot(-3) - 2\cdot(-6) + 3\cdot(-3) \\
&= 0.
\end{aligned}$$

There's a good sign: the rank of this matrix is 2 (*check!*) so it's not invertible!

Proposition 4.11. $\det I_n = 1$.

Proof. We proceed by induction. If $n = 1, \det(1) = 1$. Suppose that $\det I_{n-1} = 1$. Then

$$\begin{aligned}
\det I_n &= \det\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\
&= \det\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & I_{n-1} \end{array}\right) \\
&= 1\cdot\det I_{n-1} - 0\cdot\overline{(I_n)_{12}} + 0\cdot\overline{(I_n)_{13}} - \dots \\
&= \det I_{n-1} \\
&= 1.
\end{aligned}$$

□

Theorem 4.12. $\det$ *is linear in each row. That is, if $A \in M_n(F)$,*

$$\det\begin{pmatrix} A_1 \\ \vdots \\ A_i + \alpha A_i' \\ \vdots \\ A_n \end{pmatrix} = \det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} + \alpha\det\begin{pmatrix} A_1 \\ \vdots \\ A_i' \\ \vdots \\ A_n \end{pmatrix}.$$

Proof. We use induction of n. If $n = 1$,

$$\det(a + \alpha a') = a + \alpha a' = \det(a) + \alpha\det(a').$$

Suppose the statement is true for any matrix in $M_{n-1}(F)$. If $i = 1$,

$$\begin{aligned}\det\begin{pmatrix} A_1 + \alpha A_1' \\ \vdots \\ A_n \end{pmatrix} &= \sum_{j=1}^{n} (-1)^{1+j}(\alpha_{1j} + \alpha\alpha_{1j}') \det \overline{A_{1j}} \\ &= \sum_{j=1}^{n} (-1)^{1+j}\alpha_{1j} \det \overline{A_{1j}} + \alpha \sum_{j=1}^{n} (-1)^{1+j}\alpha_{1j}' \det \overline{A_{1j}} \\ &= \det A + \alpha \det A'.\end{aligned}$$

If $i > 1$,

$$\begin{aligned}\det\begin{pmatrix} A_1 \\ \vdots \\ A_i + \alpha A_i' \\ \vdots \\ A_n \end{pmatrix} &= \sum_{j=1}^{n} (-1)^{1+j}\alpha_{1j} \det \begin{pmatrix} \vdots \\ \alpha_{i1}+\alpha\alpha_{i1}' \cdots \alpha_{i(j-1)}+\alpha\alpha_{i(j-1)}' \;\; \alpha_{i(j+1)}+\alpha\alpha_{i(j+1)}' \cdots \alpha_{in}+\alpha\alpha_{in}' \\ \vdots \end{pmatrix} \\ &= \sum_{j=1}^{n} (-1)^{1+j}\alpha_{1j} (\det \overline{A_{1j}} + \alpha \det \overline{A_{1j}'}) \text{ (by induction)} \\ &= \det A + \alpha \det A'.\end{aligned}$$

$\square$

Corollary 4.13. *If a row of $A \in M_n(F)$ consists of all zeroes, then* $\det A = 0$.

Proof. Apply linearity to the row of zeroes:

$$\det\begin{pmatrix} A_1 \\ \vdots \\ 0 \\ \vdots \\ A_n \end{pmatrix} = 0 \cdot \det\begin{pmatrix} A_1 \\ \vdots \\ 0 \\ \vdots \\ A_n \end{pmatrix} = 0.$$

$\square$

Lemma 4.14. *Let $A \in M_n(F), n > 1$. Suppose row i of A is e_k, one of the standard basis vectors of F^n:*

$$A = \begin{array}{c} \begin{array}{ccccc} & & k & & \end{array} \\ \begin{pmatrix} & & \vdots & & \\ 0 & \cdots & 1 & \cdots & 0 \\ & & \vdots & & \end{pmatrix} i \end{array}$$

Then

$$\det A = (-1)^{i+k} \det \overline{A_{ik}}.$$

Proof. We use induction on n. If $n = 2$, we can simply check the four possible cases. For example,

$$\det \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} = -b = (-1)^{1+2} \det \overline{A_{12}}.$$

The other three are left as exercises. Suppose then that the statement holds for all matrices in $M_{n-1}(F)$, and suppose row i of A is e_k. If $i = 1$, the statement follows immediately from the definition of det:

$$\det A = 0 \cdot \det \overline{A_{11}} - \cdots + (-1)^{1+k} \cdot 1 \cdot \det \overline{A_{1k}} + \cdots + (-1)^{1+n} \cdot 0 \cdot \det \overline{A_{1n}} = (-1)^{1+k} \det \overline{A_{1k}}.$$

Suppose now that $1 < i \leq n$. Let $\overline{C_{ij}}$ be the matrix obtained from A by deleting rows 1 and i and columns j and k (with $j \neq k$.

Notice first that

$$\text{row } i-1 \text{ of } \overline{A_{1j}} = \begin{cases} e_{k-1}, & j < k \\ 0, & j = k \\ e_k, & j > k \end{cases}$$

Thus, by induction,

$$\det \overline{A_{1j}} = \begin{cases} (-1)^{i-1+k-1} \det \overline{C_{ij}}, & j < k \\ 0, & j = k \\ (-1)^{i-1+k} \det \overline{C_{ij}}, & j > k \end{cases}$$

Therefore,

$$\begin{aligned}
\det A &= \sum_{j=1}^{n} (-1)^{1+j} \alpha_{1j} \det \overline{A_{1j}} \\
&= \sum_{j=1}^{k-1} (-1)^{1+j} \alpha_{1j} \det \overline{A_{1j}} + \sum_{j=k+1}^{n} (-1)^{1+j} \alpha_{1j} \det \overline{A_{1j}} \\
&= \sum_{j=1}^{k-1} (-1)^{1+j} \alpha_{1j} (-1)^{i-1+k-1} \det \overline{C_{ij}} + \sum_{j=k+1}^{n} (-1)^{1+j} \alpha_{1j} \det (-1)^{i-1+k} \det \overline{C_{ij}} \\
&= (-1)^{i+k} \left(\sum_{j=1}^{k-1} (-1)^{1+j} \alpha_{1j} \det \overline{C_{ij}} + \sum_{j=k+1}^{n} (-1)^{j} \alpha_{1j} \det \overline{C_{ij}} \right) \\
&= (-1)^{i+k} \det \overline{A_{ik}}.
\end{aligned}$$

$\square$

Theorem 4.15. *If $A \in M_n(F)$, then $\det A$ can be calculated by expanding along any row. That is, for any $1 \leq i \leq n$,*

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

Proof.

$$\det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} = \det\begin{pmatrix} A_1 \\ \vdots \\ \alpha_{i1}e_1 + \ldots + \alpha_{in}e_n \\ \vdots \\ A_n \end{pmatrix} = \sum_{j=1}^{n} \alpha_{ij} \det\begin{pmatrix} A_1 \\ \vdots \\ e_j \\ \vdots \\ A_n \end{pmatrix} = \sum_{j=1}^{n} (-1)^{i+j} \alpha_{ij} \det \overline{A_{ij}}.$$

□

Corollary 4.16. *If $A \in M_n(F)$ has two identical rows, then* $\det A = 0$.

Proof. We use induction on n. If $n = 2$, then

$$\det\begin{pmatrix} a & b \\ a & b \end{pmatrix} = ab - ba = 0.$$

Assume that the statement is true for all matrices in $M_{n-1}(F)$ with $n \geq 3$, and suppose that

$$A = \begin{array}{c} \begin{pmatrix} A_1 \\ \vdots \\ A_r \\ \vdots \\ A_r \\ \vdots \\ A_n \end{pmatrix} \end{array} \begin{array}{l} \\ \\ r \\ \\ s \\ \\ \\ \end{array}$$

Use the Theorem to expand the determinant along any row $i \neq r, s$. Then clearly $\overline{A_{ij}}$ has two identical rows, so by induction, has determinant 0. □

Remark 4.17. We now examine what effect performing an elementary row operation on A has on $\det A$. One operation is covered by Theorem 4.12:

$$\det\begin{pmatrix} A_1 \\ \vdots \\ \alpha A_i \\ \vdots \\ A_n \end{pmatrix} = \alpha \det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix}.$$

Proposition 4.18. *Adding a multiple of one row of $A \in M_n(F)$ to another leaves* $\det A$ *unchanged.*

Proof.

$$\det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j + \alpha A_i \\ \vdots \\ A_n \end{pmatrix} = \det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} + \alpha \det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} = \det A + \alpha \cdot 0 = \det A.$$

□

Corollary 4.19. *Let $A \in M_n(F)$. If $r(A) < n$, then* $\det A = 0$*.*

Proof. If $r(A) < n$, then the rows of A are linearly dependent, so we can write $\alpha_1 A_1 + \ldots + \alpha_n A_n = 0$, where at least one coefficient, say α_i, is not 0. Then

$$\begin{aligned} \det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} &= \alpha_i^{-1} \det\begin{pmatrix} A_1 \\ \vdots \\ \alpha_i A_i \\ \vdots \\ A_n \end{pmatrix} \\ &= \alpha_i^{-1} \det\begin{pmatrix} A_1 \\ \vdots \\ \alpha_1 A_1 + \ldots + \alpha_n A_n \\ \vdots \\ A_n \end{pmatrix} \\ &= \alpha_i^{-1} \det\begin{pmatrix} A_1 \\ \vdots \\ 0 \\ \vdots \\ A_n \end{pmatrix} \\ &= \alpha^{-1} \cdot 0 = 0. \end{aligned}$$

□

Proposition 4.20. *Let $A \in M_n(F)$. Switching two rows of A changes the sign of the determinant.*

Proof. Say we want to switch rows i and j. We see that

$$0 = \det \begin{pmatrix} A_1 \\ \vdots \\ A_i + A_j \\ \vdots \\ A_i + A_j \\ \vdots \\ A_n \end{pmatrix} = \det \begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} + \det \begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} + \det \begin{pmatrix} A_1 \\ \vdots \\ A_j \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} + \det \begin{pmatrix} A_1 \\ \vdots \\ A_j \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix}$$

$$= \det \begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} + \det \begin{pmatrix} A_1 \\ \vdots \\ A_j \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix}.$$

The result follows immediately. □

Remark 4.21. Now that we have established the effects of elementary row operations on $\det A$, we can easily calculate the determinants of elementary matrices:

Type 1: (switching rows) $\det E = -\det I_n = -1$;
Type 2: (multiplying a row by $\alpha \neq 0$) $\det E = \alpha I_n = \alpha$;
Type 3: (adding a row to another) $\det E = \det I_n = 1$.

Theorem 4.22. *The determinant is multiplicative. That is, if $A, B \in M_n(F)$,*

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

Proof. If $A = E$ is elementary, then we have:

Type 1: $\det(EB) = -\det B = \det E \cdot \det B$;
Type 2: $\det(EB) = \alpha \det B = \det E \cdot \det B$;
Type 3: $\det(EB) = \det B = \det E \cdot \det B$.

If $r(A) < n$, then $r(AB) < n$ as well, since if L_A is not surjective, neither is L_{AB}. In that case then,

$$\det(AB) = 0 = 0 \cdot \det B = \det A \cdot \det B.$$

If on the other hand $r(A) = n$, so that A is invertible, then we have $A = E_1 \cdots E_k$, where the E_i are elementary. Therefore,

$$\begin{aligned}
\det(AB) &= \det(E_1 \cdots E_k B) \\
&= \det E_1 \det(E_2 \cdots E_k B) \\
&\vdots \\
&= \det E_1 \cdots \det E_k \det B \\
&= (\det(E_1 E_2) \cdots \det E_k) \det B \\
&\vdots \\
&= \det(E_1 \cdots E_k) \det B \\
&= \det A \cdot \det B.
\end{aligned}$$

$\square$

Corollary 4.23. *$A \in M_n(F)$ is invertible $\iff$ $\det A \neq 0$. In that case,* $\det A^{-1} = (\det A)^{-1}$.

Proof. ($\Longleftarrow$) If A is not invertible, then $r(A) < n$, so $\det A = 0$.
($\Longrightarrow$) If A is invertible,

$$1 = \det I_n = \det(AA^{-1}) = \det A \cdot \det A^{-1} \implies \det A \neq 0.$$

$\square$

Theorem 4.24. *If $A \in M_n(F)$, then* $\det A = \det A^t$.

Proof. If A is not invertible, then $r(A) = r(A^t) < n$, so $\det A = \det A^t = 0$. Suppose than that A is invertible, and write $A = E_1 \cdots E_k$, where the E_i are elementary.. It's easy to see that the theorem holds for elementary matrices, so we have

$$\begin{aligned}
\det A^t &= \det(E_1 \cdots E_k)^t \\
&= \det(E_k^t \cdots E_1^t) \\
&= \det E_k^t \cdots \det E_1^t \\
&= \det E_k \cdots \det E_1 \\
&= \det E_1 \cdots \det E_k \\
&= \det(E_1 \cdots E_k) \\
&= \det A.
\end{aligned}$$

$\square$

Corollary 4.25. *The determinant can be calculated by expanding along any row or column. That is, if $A \in M_n(F)$ and $1 \le i, j \le n$,*

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

5. Cramer's Rule

Theorem 5.1 (***Cramer***)**.** *Let $AX = B$ be a system of n equations in n unknowns, with $\det A \neq 0$. Then*

(1) the system is consistent with a unique solution $X \in F^n$;
(2) if $\overline{A}_i$ is the matrix obtained by replacing column i of A with B, then $x_i = \frac{\det \overline{A}_i}{\det A}$.

Proof.

(1) $AX = B \iff A^{-1}AX = A^{-1}B \iff X = A^{-1}B$.

(2) Let A_i be column i of A, so that we can write $A = \begin{pmatrix} A_1 & \cdots & A_i & \cdots & A_n \end{pmatrix}$. Let $\overline{I}_i$ be the matrix obtained by replacing column i of I_n with X; that is,

$$\overline{I}_i = \begin{pmatrix} e_1 & \cdots & A_i & \cdots & e_n \end{pmatrix} = \begin{pmatrix} 1 & \cdots & x_1 & \cdots & 0 \\ & & \vdots & & \\ 0 & \cdots & x_i & \cdots & 0 \\ & & \vdots & & \\ 0 & \cdots & x_n & \cdots & 1 \end{pmatrix}.$$

To calculate $\det \overline{I}_i$, expand along row i:

$$\det \overline{I}_i = (-1)^{i+i} x_i \det I_{n-1} = x_i.$$

Since $Ae_j = A_j$, we also see that

$$A\overline{I}_i = \begin{pmatrix} Ae_1 & \cdots & AX & \cdots & Ae_n \end{pmatrix} = \begin{pmatrix} A_1 & \cdots & B & \cdots & A_n \end{pmatrix} = \overline{A}_i.$$

Therefore,

$$\det \overline{A}_i = \det A \cdot \det \overline{I}_i = \det A \cdot x_i \implies x_i = \frac{\det \overline{A}_i}{\det A}.$$

□

Example 5.2. To solve the system

$$\begin{aligned} 2x_1 + x_2 - 3x_3 &= 5 \\ x_1 - 2x_2 + x_3 &= 10 \\ 3x_1 + 4x_2 - 2x_3 &= 0 \end{aligned}$$

we calculate

$$\det A = \det \begin{pmatrix} 2 & 1 & -3 \\ 1 & -2 & 1 \\ 3 & 4 & -2 \end{pmatrix} = -25;$$

$$\det \overline{A}_1 = \det \begin{pmatrix} 5 & 1 & -3 \\ 10 & -2 & 1 \\ 0 & 4 & -2 \end{pmatrix} = -100;$$

$$\det \overline{A}_2 = \det \begin{pmatrix} 2 & 5 & -3 \\ 1 & 10 & 1 \\ 3 & 0 & -2 \end{pmatrix} = 75;$$

$$\det \overline{A}_3 = \det \begin{pmatrix} 2 & 1 & 5 \\ 1 & -2 & 10 \\ 3 & 4 & 0 \end{pmatrix} = 0.$$

Thus the unique solution is $X = \begin{pmatrix} 4 \\ -3 \\ 0 \end{pmatrix}$.

Remark 5.3. Cramer's Rule is a beautiful mathematical result, but completely impractical: the computing time necessary to calculate determinants of large matrices is prohibitive.

6. Exercises

Exercise 6.1.

(a) Show that if $A, B \in M_n(F)$, then $tr(AB) = tr(BA)$.
(b) Show that if $A \in M_n(F)$, then $tr(A) = tr(A^t)$.
(c) Show that if $A, B \in M_n(F)$ are similar, then $tr(A) = tr(B)$.

Exercise 6.2. Let $A \in M_{m\times n}(F)$, and let $P \in M_m(F)$ and $Q \in M_n(F)$ be invertible.

(a) Prove that $r(PA) = r(A)$.
(b) Prove that $r(PAQ) = r(A)$

Exercise 6.3.

(a) Suppose that $A, B \in M_n(F)$ are invertible. Prove that AB is also invertible.
(b) Suppose that $A \in M_n(F)$ is invertible. Prove that A^t is also invertible.
(c) Let $A \in M_{m\times n}(F)$. Show that $r(A) = r$ if and only if there exist invertible matrices $P \in M_m(F)$ and $Q \in M_n(F)$ such that

$$PAQ = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

Exercise 6.4. For each of the following matrices A, use the augmented matrix procedure to find A^{-1} or determine that A is not invertible.

(a) $A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 4 \\ 2 & 3 & -1 \end{pmatrix}$

Exercise 6.5. Show that every invertible matrix $A \in M_n(F)$ is the product of elementary matrices.

Exercise 6.6. Let $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show that there exist $A, B \in M_2(F)$ such that $X = AB - BA \iff a + d = 0$.

Exercise 6.7. Let $n < m$, $A \in M_{m\times n}(F)$, and $B \in M_{n\times m}(F)$, Show that AB is not invertible.

Exercise 6.8. Let $A \in M_{m\times n}(F)$ have rank m and $B \in M_{n\times p}(F)$ have rank n. Determine, with proof, the rank of AB.

Exercise 6.9. Let $A \in M_{m\times n}(F)$ have rank m. Prove that there exists $B \in M_{n\times m}(F)$ such that $AB = I_m$.

Exercise 6.10. The *classical adjoint* of $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F)$ is $\operatorname{Adj} A = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

(a) Show that $\operatorname{Adj} A \cdot A = A \cdot \operatorname{Adj} A = \det A \cdot I_2$.
(b) Show that $\det \operatorname{Adj} A = \det A$.
(c) Show that $(\operatorname{Adj} A)^t = \operatorname{Adj} A^t$.
(d) Show that if A is invertible, $A^{-1} = (\det A)^{-1} \operatorname{Adj} A$.

Exercise 6.11. Let $\delta : M_2(F) \to F$ be a function that satisfies:

(i) δ is linear in each row;
(ii) if the two rows of A are the same, then $\delta(A) = 0$;
(iii) $\delta(I_2) = 1$.

Show that $\delta = \det$.

Exercise 6.12. Compute $\det A$.

(a) $A = \begin{pmatrix} 1 & -2 & 3 \\ -1 & 2 & -5 \\ 3 & -1 & 2 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & -2 & 3 & -12 \\ -5 & 12 & -14 & 19 \\ -9 & 22 & -20 & 31 \\ -4 & 9 & -14 & 15 \end{pmatrix}$

Exercise 6.13. Find the value of α if

$$\det \begin{pmatrix} b_1 + c_1 & b_2 + c_2 & b_3 + c_3 \\ a_1 + c_1 & a_2 + c_2 & a_3 + c_3 \\ a_1 + b_1 & a_2 + b_2 & a_3 + b_3 \end{pmatrix} = \alpha \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

Exercise 6.14. A matrix $A = (\alpha_{ij}) \in M_n(F)$ is *upper triangular* if $\alpha_{ij} = 0$ when $i > j$. Prove that the determinant of an upper triangular matrix is the product of its diagonal entries.

Exercise 6.15. Under what conditions is $\det(-A) = \det A$?

Exercise 6.16. A matrix $A \in M_n(F)$ is *nilpotent* if $A^k = 0$, for some $k \in \mathbb{Z}^+$. Show that if A is nilpotent, then $\det A = 0$.

Exercise 6.17. A matrix $A \in M_n(F)$ is *orthogonal* if $AA^t = I_n$. Show that if A is orthogonal, then $\det A = \pm 1$.

Exercise 6.18. A matrix $A \in M_n(F)$ is *skew symmetric* if $A^t = -A$. Show that if A is skew symmetric and n is odd, then A is not invertible.

Exercise 6.19. Suppose that $M \in M_n(F)$ is of the form

$$M = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right),$$

where A and C are square matrices. Show that $\det M = \det A \cdot \det C$.

CHAPTER V

Eigenvalues

1. Definition and Examples

Definition 1.1. Let $T : V \to V$ be a linear operator. Then T is diagonalizable if there is a basis $\mathcal{B}$ of V such that

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

a diagonal matrix.

Remark 1.2. If $\mathcal{B} = \{v_1, \ldots, v_n\}$, then

$$[Tv_i]_{\mathcal{B}} = [T]_{\mathcal{B}}[v_i]_{\mathcal{B}} = [T]_{\mathcal{B}} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix}.$$

That is, $Tv_i = \lambda_i v_i$.

Definition 1.3. Let $T : V \to V$ be a linear operator. If $Tv = \lambda v$, for some $\lambda \in F$ and some $0 \neq v \in V$, then λ is an eigenvalue of T and v is an eigenvector of T corresponding to λ.

Remark 1.4. Notice that T is diagonalizable $\iff$ V has a basis consisting entirely of eigenvectors.

Example 1.5. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \in M_n(\mathbb{R})$. Then

$$L_A(2,3) = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \end{pmatrix} = 4 \cdot \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and

$$L_A(1,-1) = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus L_A has at least two eigenvalues.

Example 1.6. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_n(\mathbb{R})$. Then

$$L_A(x,y) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$$

But $(x,y) = \lambda(-y,x) \implies x = y = 0$, so L_A has no eigenvalues.

Example 1.7. Let C^∞ be the set of all infinitely differentiable real valued functions of a real variable. Then elementary calculus shows that C^∞ is a vector space over $\mathbb{R}$, and that $T : C^\infty \to C^\infty$ defined by $Tf = f'$ is a linear operator. Then

$$Tf = \lambda f \iff f' = \lambda f \iff f(x) = ce^{\lambda x}, \text{ for some constant } c.$$

Thus *every* $\lambda \in \mathbb{R}$ is an eigenvalue of T.

Definition 1.8. If $A \in M_n(F)$, the eigenvalues and eigenvectors of A are those of the linear operator $L_A : F^n \to F^n$.

2. The Characteristic Polynomial

Theorem 2.1. *$\lambda \in F$ is an eigenvalue of $A \in M_n(F)$* $\iff \det(A - \lambda I_n) = 0$.

Proof.

$$\begin{aligned} \lambda \text{ is an eigenvalue} &\iff Av = \lambda v, \text{ for some } 0 \neq v \in F^n \\ &\iff Av - \lambda v = 0 \\ &\iff Av - \lambda I_n v = 0 \\ &\iff (A - \lambda I_n)v = 0 \\ &\iff v \in \operatorname{Ker} L_{A-\lambda I_n} \\ &\iff A - \lambda I_n \text{ is not invertible} \\ &\iff \det(A - \lambda I_n) = 0. \end{aligned}$$

□

Example 2.2. For $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$, we consider

$$\begin{aligned} \det\left(\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right) &= \det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix} \\ &= \lambda^2 - 3\lambda - 4 = 0 \\ &\iff \lambda = 4, -1. \end{aligned}$$

Example 2.3. For $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, we consider

$$\begin{aligned}\det\left(\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) &= \det\begin{pmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{pmatrix} \\ &= \lambda^2 - 2\cos\theta\lambda + 1.\end{aligned}$$

Using the quadratic formula, we see that this expression can be 0 only when $\cos\theta = \pm 1$, or when $\theta = 0, \pi$. Then

$$\theta = 0 \implies A = I_2 \implies \lambda = 1$$

and

$$\theta = \pi \implies A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \implies \lambda = -1.$$

In both cases, all nonzero vectors are eigenvectors.

Remark 2.4. In this last example, for $F = \mathbb{R}^2$, the matrix A represents a counterclockwise rotation of the plane through an angle θ. Thus no directions are fixed (no eigenvalues!) unless the rotation is the trivial one (so every vector goes to itself) or a half turn (every vector goes to its opposite).

Definition 2.5. If $A \in M_n(F)$, the characteristic polynomial of A is

$$p_A(t) = \det(A - \lambda I_n).$$

Thus the eigenvalues of A are the roots of $p_A(t)$.

Remark 2.6. If $T : V \to V$ is a linear operator, and $\mathcal{B}, \mathcal{B}'$ are bases of V, then

$$\begin{aligned}\det[T]_{\mathcal{B}} &= \det([I_V]^{\mathcal{B}}_{\mathcal{B}'}[T]_{\mathcal{B}'}[I_V]^{\mathcal{B}'}_{\mathcal{B}}) \\ &= \det[I_V]^{\mathcal{B}}_{\mathcal{B}'} \cdot \det[T]_{\mathcal{B}'} \cdot \det[I_V]^{\mathcal{B}'}_{\mathcal{B}} \\ &= \det[T]_{\mathcal{B}'}.\end{aligned}$$

Thus is makes sense to define the determinant of T as the determinant of any of its matrices, and hence to also define the characteristic polynomial of T.

Example 2.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(x, y, z) = 7x-4y+10z, 4x-3y+8z, -2x+y-2z$. Using the standard basis $\mathcal{B} = \{e_1, e_2, e_3\}$, we see that

$$[T]_{\mathcal{B}} = \begin{pmatrix} 7 & -4 & 10 \\ 4 & -3 & 8 \\ -2 & 1 & -2 \end{pmatrix},$$

so

$$p_T(t) = \det\begin{pmatrix} 7-t & -4 & 10 \\ 4 & -3-t & 8 \\ -2 & 1 & -2-t \end{pmatrix} = -t^3 + 2t^2 + t - 2 = (1+t)(1-t)(2-t).$$

Thus T has 3 eigenvalues: $-1, 1, 2$.

Definition 2.8. Let $T : V \to V$ be a linear operator with eigenvalue λ. The eigenspace corresponding to λ is

$$E_\lambda = \operatorname{Ker}(T - \lambda I_V) = \{v \in V : Tv = 0\}.$$

Remark 2.9. The eigenspace E_λ is not exactly the set of eigenvectors of T since it includes 0. But as the kernel of a linear transformation, it is a subspace of V.

Example 2.10. To find the eigenspace E_1 from the previous example, we must solve $T(x, y, z) = 1 \cdot (x, y, z)$. So we examine

$$[T - 1 \cdot I_3]_{\mathcal{B}} = \begin{pmatrix} 7-1 & -4 & 10 \\ 4 & -3-1 & 8 \\ -2 & 1 & -2-1 \end{pmatrix} = \begin{pmatrix} 6 & -4 & 10 \\ 4 & -4 & 8 \\ -2 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} -2 & 1 & -3 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

and solve

$$\begin{aligned} -2x + y - 3z &= 0 \\ -y + z &= 0 \end{aligned}$$

to find that $E_1 = \operatorname{Span}\{(-1, 1, 1)\}$.
We can similarly calculate $E_{-1} = \operatorname{Span}\{(1, 2, 0)\}$ and $E_2 = \operatorname{Span}\{(-2, 0, 1)\}$. These three spanning vectors are independent (*check!*), so together form a basis $\mathcal{C} = \{(-1, 1, 1), (1, 2, 0), (-2, 0, 1)\}$.
Thus T is diagonalizable:

$$[T]_{\mathcal{C}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

3. Diagonalizability Criteria

Theorem 3.1. *Eigenvectors corresponding to distinct eigenvalues are independent. That is, if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T with respective eigenvectors $v_1, \ldots, v_k$, then $\{v_1, \ldots, v_k\}$ is a linearly independent set.*

Proof. We use induction on k. If $k = 1$, then the eigenvector v_1 is nonzero, so $\{v_1\}$ is certainly independent. Suppose then that the statement holds for $k - 1$, let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues, and let $v_1, \ldots, v_k$ be corresponding eigenvectors.

To show independence, suppose that

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = 0, \text{ for some } \alpha_i \in F.$$

Then

$$\begin{aligned}(T-\lambda_k I_V)(\alpha_1 v_1+\ldots+\alpha_k v_k) = 0 \\ &= \alpha_1(\lambda_1-\lambda_k)v_1+\ldots+\alpha_k(\lambda_k-\lambda_k)v_k \\ &= \alpha_1(\lambda_1-\lambda_k)v_1+\ldots+\alpha_{k-1}(\lambda_{k-1}-\lambda_k)v_{k-1}.\end{aligned}$$

By induction then,

$$\alpha_1(\lambda_1-\lambda_k)=\ldots=\alpha_{k-1}(\lambda_{k-1}-\lambda_k)=0.$$

The eigenvalues are distinct, so this means that $\alpha_1=\ldots=\alpha_{k-1}=0$. But then the original dependence relation reduces to $\alpha_k v_k=0$, so $\alpha_k=0$ as well, completing the proof. □

Corollary 3.2. *Let $T:V\to V$ be a linear operator, where* $\dim V=n$. *If T has n distinct eigenvalues, then T is diagonalizable.*

Proof. Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues. Then if v_i is an eigenvector corresponding to λ_i, the set $\mathcal{B}=\{v_1,\ldots,v_n\}$ is independent, so forms a basis of eigenvectors. Specifically,

$$[T]_{\mathcal{B}}=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n\end{pmatrix}$$

□

Remark 3.3. The converse of this last statement is false. For example, I_n is certainly diagonalizable, but

$$p_{I_n}(t)=\det\begin{pmatrix}1-t & 0 & \cdots & 0\\ 0 & 1-t & \cdots & 0\\ & & \vdots & \\ 0 & 0 & \cdots & 1-t\end{pmatrix}=(1-t)^n,$$

so 1 is the only eigenvalue.

Definition 3.4. A polynomial $p(t)\in\mathcal{P}_n(F)$ splits over F if

$$p(t)=\gamma(t-\alpha_1)\cdots(t-\alpha_n),$$

where $\gamma,\alpha_1,\ldots,\alpha_n\in F$. That is, $p(t)$ factors completely into linear polynomials with coefficients in F.

Example 3.5. $p(t)=t^2+1$ does not split over $\mathbb{R}$, but does over $\mathbb{C}$, since $p(t)=(t-i)(t+i)$.

Theorem 3.6. *Let $T:V\to V$ be a linear operator, where V is a vector space over the field F. Then if T is diagonalizable, $p_T(t)$ splits over F.*

Proof. Choose a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of eigenvectors, where v_i corresponds to the eigenvalue λ_i. Then

$$p_T(t) = \det \begin{pmatrix} \lambda_1 - t & 0 & \cdots & 0 \\ 0 & \lambda_2 - t & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n - t \end{pmatrix} = (\lambda_1 - t)(\lambda_2 - t)\cdots(\lambda_n - t).$$

$\square$

Definition 3.7. Let λ be an eigenvalue of $TLV \to V$, The multiplicity m_λ of λ is the largest positive integer such that $(t-\lambda)^{m_\lambda}$ is a factor of $p_T(t)$. That is,

$$p_T(t) = (t-\lambda)^{m_\lambda} q(t), \text{ where } q(\lambda) \neq 0.$$

Example 3.8. Let $A = \begin{pmatrix} 2 & 6 & 1 & 0 \\ 0 & 2 & 3 & -2 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & -4 \end{pmatrix}$. Then $p_A(t) = (t-2)^2(t-3)(t+4)$, so $m_2 = 2$ and $m_3 = m_{-4} = 1$.

Theorem 3.9. *Let λ be an eigenvalue of $T : V \to V$ of multiplicity m_λ. Then*

$$1 \leq \dim E_\lambda \leq m_\lambda.$$

Proof. Since λ is an eigenvector, $E_\lambda \neq \{0\}$, so the first inequality is clear. Now choose a basis $\{v_1, \ldots, v_k\}$ of E_λ and extend it to a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V. Then

$$[T]_{\mathcal{B}} = \left(\begin{array}{cccc|c} \lambda & 0 & \cdots & 0 & \cdots \\ 0 & \lambda & \cdots & 0 & \cdots \\ & \vdots & & \vdots & \\ 0 & 0 & \cdots & \lambda & \cdots \\ \hline & \vdots & & \vdots & \end{array}\right) = \left(\begin{array}{c|c} \lambda I_k & B \\ \hline 0 & C \end{array}\right).$$

Thus

$$p_T(t) = \det \left(\begin{array}{c|c} (\lambda - t) I_k & B \\ \hline 0 & C - tI_{n-k} \end{array}\right) = \det((\lambda - t)I_k) \cdot \det(C - tI_{n-k}) = (\lambda - t)^k q(t).$$

Therefore $k \leq m_\lambda$ by the maximality of m_λ. $\square$

Theorem 3.10. *Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, \ldots, k$, let $S_i \subseteq E_{\lambda_i}$ be an independent set. Then*

$$S = S_1 \cup \cdots \cup S_k = \bigcup_{i=1}^{k} S_i$$

is also independent.

Proof. To begin, let $v_i \in E_{\lambda_i}$ and suppose $v_1 + \ldots + v_k = 0$. Then by Theorem 3.1, we must have $v_1 = \ldots = v_k = 0$.

So now let $S_i = \{v_{i1}, \ldots, v_{ik_i}\}$, and take a linear combination of the vectors in S:

$$\sum_{i=1}^{k} \underbrace{\sum_{j=1}^{k_i} \alpha_{ij} v_{ij}}_{\in S_i} = 0, \text{ where } \alpha_{ij} \in F.$$

By the initial remark, we have

$$\sum_{j=1}^{k_i} \alpha_{ij} v_{ij} = 0, \text{ for each } i = 1, \ldots, k.$$

But then by the independence of S_i, $\alpha_{ij} = 0$ for all i, j. Thus S is independent. □

Theorem 3.11. *Let $T : V \to V$ be a linear operator where $\dim V = n$, and let $\lambda_1, \ldots, \lambda_k$ be all the distinct eigenvalues. Suppose that $p_T(t)$ splits. Then*

(1) T is diagonalizable $\iff m_{\lambda_i} = \dim E_{\lambda_i}$, for all $i = 1, \ldots, k$.
(2) If T is diagonalizable and $\mathcal{B}_i$ is a basis of E_{λ_i} for $i = 1, \ldots, k$, then

$$\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k = \bigcup_{i=1}^{k} \mathcal{B}_i$$

is a basis of V.

Proof.

(1) ($\implies$) Let $\mathcal{B}$ be a basis of eigenvectors, and let $\mathcal{B}_i = \mathcal{B} \cap E_{\lambda_i}$ contain n_i vectors. Let $d_i = \dim E_{\lambda_i}$. Then $n_i \leq d_i \leq m_{\lambda_i}$, so

$$n = \sum_{i=1}^{k} n_i \leq \sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} m_{\lambda_i} = \deg p_T(t) = n.$$

Therefore,

$$0 = n - n = \sum_{i=1}^{k} m_{\lambda_i} - \sum_{i=1}^{k} d_i = \sum_{i=1}^{k} (m_{\lambda_i} - d_i).$$

But $m_{\lambda_i} - d_i \geq 0$, so $m_{\lambda_i} = d_i$.

($\impliedby$) Suppose that $m_{\lambda_i} = d_i$. Let $\mathcal{B}_i$ be a basis of E_{λ_i}, and $\mathcal{B} = \cup_{i=1}^{k} \mathcal{B}_i$. Then $\mathcal{B}$ is independent by the previous theorem, and contains $\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} m_{\lambda_i} = n$ vectors. Thus $\mathcal{B}$ is a basis of eigenvectors, so T is diagonalizable.

(2) This follows immediately from the proof of (1).

□

Remark 3.12. To summarize, T is diagonalizable if and only if $p_T(t)$ factors into linear polynomials and, in addition, for every eigenvalue λ,

$$m_\lambda = \dim E_\lambda = n(T - \lambda I_V) = n - r(T - \lambda I_V).$$

Example 3.13. Let $A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{R})$. We calculate

$$p_A(t) = \det \begin{pmatrix} -t & 0 & 1 \\ 1 & -t & -1 \\ 0 & 1 & 1-t \end{pmatrix} = -t^3 + t^2 - t + 1 = (1-t)(t^2+1).$$

Since $p_A(t)$ does not split, A is not diagonalizable.

Example 3.14. Let $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \in M_2(\mathbb{R})$. We calculate

$$p_A(t) = \det \begin{pmatrix} 1-t & 2 \\ 0 & 1-t \end{pmatrix} = (1-t)^2.$$

So $p_A(t)$ does split, and there is one eigenvalue $\lambda = 1$ of multiplicity $m_1 = 2$. But

$$\dim E_1 = 2 - r(A - 1 \cdot I_2) = 2 - r \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = 2 - 1 = 1 \neq 2 = m_1,$$

so A is not diagonalizable.

Example 3.15. Let $A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{pmatrix} \in M_3(\mathbb{R})$. We calculate

$$p_A(t) = \det \begin{pmatrix} 3-t & 1 & 1 \\ 2 & 4-t & 2 \\ -1 & -1 & 1-t \end{pmatrix} = -t^3 + 8t^2 - 20t + 16 = (2-t)^2(4-t).$$

So $p_A(t)$ does split, and there are two eigenvalues $\lambda_1 = 2$ of multiplicity $m_2 = 2$ and $\lambda_2 = 4$ of multiplicity $m_4 = 1$. Also,

$$\dim E_2 = 3 - r(A - 2 \cdot I_3) = 2 - r \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{pmatrix} = 3 - 1 = 2 = m_2.$$

Now $\dim E_4$ must be at least 1 but cannot be greater by Theorem 3.11. Therefore, A is diagonalizable.

Let's find a basis of eigenvectors for V.

$$A - 2 \cdot I_3 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies E_2 = \operatorname{Span}\{(-1,0,1), (-1,1,0)\}$$

and

$$A - 4 \cdot I_3 = \begin{pmatrix} -1 & 1 & 1 \\ 2 & 0 & 2 \\ -1 & -1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \implies E_4 = \text{Span}\{(-1,-2,1)\}.$$

Thus our basis is $\mathcal{B} = \{(-1,0,1), (-1,1,0), (-1,-2,1)\}$. In that case,

$$A = Q \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} Q^{-1},$$

where

$$Q = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & -2 \\ 1 & 0 & 1 \end{pmatrix},$$

the change of basis matrix from $\mathcal{B}$ to the standard basis $\{e_1, e_2, e_3\}$.

Theorem 3.16. *Let $T : V \to V$ be a linear operator, and suppose that $p_T(t)$ splits, with eigenvalues $\lambda_1, \dots, \lambda_n$ (some possibly repeated). Then $\det T = \lambda_1 \lambda_2 \cdots \lambda_n$.*

Proof. We have that

$$p_T(t) = (\lambda_1 - t) \cdots (\lambda_n - t) = (-1)^n t^n + \ldots + \lambda_1 \lambda_2 \cdots \lambda_n,$$

so

$$\det T = \det(T - 0 \cdot I_V) = p_T(0) = \lambda_1 \lambda_2 \cdots \lambda_n.$$

□

Remark 3.17. This result of this last theorem is obvious if T is diagonalizable, since if $\mathcal{B}$ is a basis of eigenvectors,

$$[T]_{\mathcal{B}} = Q \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} Q^{-1} \implies \det T = \lambda_1 \lambda_2 \cdots \lambda_n.$$

In fact, the theorem is always true, even if $p_T(t)$ doesn't split, but we have to view the eigenvalues in a larger field. For example, consider $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$. Then $p_T(t) = t^2 + 1$ doesn't split in $\mathbb{R}$, but it does in $\mathbb{C} : p_T(t) = (t - i)(t + i)$. From that perspective, the product of the eigenvalues is $i \cdot -i = 1 = \det A$.

It's an important theorem of Abstract Algebra that such a larger field always exists. That is, for any polynomial $f(t)$ with coefficients in a field F, there is a field K containing F such that $f(x)$ splits if coefficients in K are allowed. The theorem then holds if we view $A \in M_n(K)$

4. Exercises

Exercise 4.1. For each linear operator $T : V \to V$, find the eigenvalues of T and a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal.

(a) $V = \mathbb{R}^3; T(x, y, z) = (-4x + 3y - 6z, 6x - 7y + 12z, 6x - 6y + 11z)$
(b) $V = \mathcal{P}_2(\mathbb{R}); T(f(x)) = xf'(x) + f''(x) - f(2)$

Exercise 4.2. Prove that a linear operator $T : V \to V$ is invertible if and only if 0 is not an eigenvalue of T.

Exercise 4.3. For any $A \in M_n(F)$, show that A and A^t have the same characteristic polynomial.

Exercise 4.4. Let $T : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be defined by $T(A) = A^t$.

(a) Show that ± 1 are the only eigenvalues of T.
(b) Describe E_1 and E_{-1}.
(c) Find a basis $\mathcal{B}$ of $M_n(\mathbb{R})$ such that $[T]_{\mathcal{B}}$ is diagonal.

Exercise 4.5. Let $A = (\alpha_{ij}) \in M_n(F)$ have characteristic polynomial

$$p_A(t) = (-1)^n t^n a_{n-1} t^{n-1} + \ldots + a_1 t + a_0.$$

(a) Show that $p_A(0) = a_0 = \det A$.
(b) Deduce that A is invertible if and only if $\alpha_0 \neq 0$.
(c) Show that

$$p_A(t) = (\alpha_{11} - t)(\alpha_{22} - t) \cdots (\alpha_{nn} - t) + q(t),$$

where $q(t)$ is a polynomial of degree at most $n - 2$. *(Hint: use induction on n).*
(d) Show that $tr(A) = (-1)^{n-1} a_{n-1}$.

Exercise 4.6. Determine if each of the following matrices $A \in M_n(\mathbb{R})$ is diagonalizable, and if so, find an invertible matrix P and a diagonal matrix D such that $D = PAP^{-1}$.

(a) $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$
(b) $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{pmatrix}$

Exercise 4.7. Suppose that $A \in M_n(F)$ has two distinct eigenvalues λ_1 and λ_2, and that $\dim E_{\lambda_1} = n - 1$. Prove that A is diagonalizable.

Exercise 4.8. Let T be an invertible linear operator on on V, where $\dim_F V = n$.

(a) Show that if λ is an eigenvalue of T, then λ^{-1} is an eigenvalue of T^{-1}.
(b) Show that the eigenspace E_λ of T is the same as the eigenspace $E_{\lambda^{-1}}$ of T^{-1}.
(c) Show that if T is diagonalizable, so is T^{-1}.

Exercise 4.9. Let $A \in M_n(F)$. Recall that A and A^t have the same characteristic polynomial, and hence the same eigenvalues. For a common eigenvalue λ of A and A^t, let E_λ and E'_λ be the corresponding eigenspaces.

(a) Give an example to show that E_λ and E'_λ need not be the same.
(b) Show, however, that $\dim E_\lambda = \dim E'_\lambda$.
(c) Show that if A is diagonalizable, so is A^t.

CHAPTER VI

Inner Product Spaces

1. The Complex Numbers

Definition 1.1. The set of complex numbers $\mathbb{C}$ is constructed from the Cartesian plane

$$\mathbb{R} \times \mathbb{R} = \{(a, b) : a, b \in \mathbb{R}\}$$

by defining two operations:

- $(a, b) + (c, d) = (a + c, b + d)$
- $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$

Remark 1.2. We can easily identify $\mathbb{R}$ with the subset $\{(a, 0) : a \in \mathbb{R}\} \subseteq \mathbb{C}$ because

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

and

$$(a, 0) \cdot (b, 0) = (ab - 0 \cdot 0, a \cdot 0 + 0 \cdot c) = (ab, 0).$$

Remark 1.3. Notice that

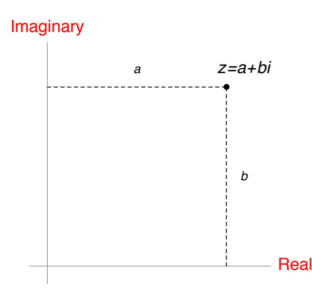
$$(0, 1)^2 = (0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0),$$

which we've identified with the real number 1. Then if we define $i = (0, 1)$ and write

$$z = (a, b) = (a, 0) + (0, b) = a + bi,$$

we recognize $\mathbb{C}$ as the field from Chapter 1 Section 2. We call a the real part of z, denoted $\Re(z)$, and b the imaginary part $\Im(z)$.

Remark 1.4. We can see the Cartesian plane now as the *complex* plane, where each point is a complex number.



Definition 1.5. Let $z = a + bi \in \mathbb{C}$. The conjugate of z is $\overline{z} = a - bi$, and the absolute value of z is $|z| = \sqrt{a^2 + b^2}$.

Proposition 1.6. *Let $z = a + bi, w = c + di \in \mathbb{C}$.*

(1) $\overline{\overline{z}} = z$.
(2) $\overline{z+w} = \overline{z} + \overline{w}$.
(3) $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$.
(4) $|z \cdot w| = |z| \cdot |w|$.
(5) $\Re(z) \leq |z|$ *and* $\Im(z) \leq |z|$.
(6) $z \cdot \overline{z} = |z|^2$.
(7) $z + \overline{z} = 2\Re(z)$ *and* $z - \overline{z} = 2\Im(z)$.
(8) $z \neq 0 \implies z^{-1} = \frac{\overline{z}}{|z|^2}$.
(9) The Triangle Inequality: $|z+w| \leq |z| + |w|$.

Proof. We leave (1)-(8) as exercises.

(9)

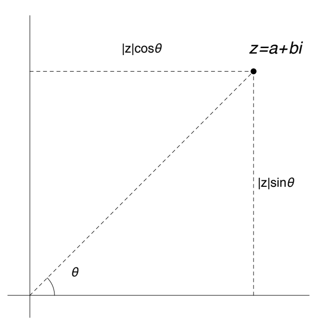
$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) \\
&= (z+w)(\overline{z}+\overline{w}) \\
&= z\overline{z} + z\overline{w} + w\overline{z} + w\overline{w} \\
&= +z\overline{w} + \overline{z\overline{w}} + |w|^2 \\
&= |z|^2 + 2\Re(z\overline{w}) + |w|^2 \\
&\leq |z|^2 + 2|z\overline{w}| + |w|^2 \\
&= (|z| + |w|)^2.
\end{aligned}$$

Taking square roots establishes the inequality.

□

Remark 1.7. Using polar coordinates, we can easily express $z = a+bi \in \mathbb{C}$ in an alternative form, known (not surprisingly) as the polar form of the complex number:

$$z = |z|(\cos\theta + i\sin\theta).$$



Proposition 1.8. *Let* $z = r(\cos\theta + i\sin\theta)$ *and* $w = s(\cos\phi + i\sin\phi)$ *be complex numbers. Then*

$$zw = rs(\cos(\theta+\phi) + i\sin(\theta+\phi)).$$

Proof. Simply multiply and then apply the sum formulas for the trig functions:

$$\begin{aligned}
zw &= r(\cos\theta + i\sin\theta)\cdot s(\cos\phi + i\sin\phi)\\
&= rs((\cos\theta\cos\phi - \sin\theta\sin\phi) + i(\cos\theta\sin\phi + \sin\theta\cos\phi))\\
&= rs(\cos(\theta+\phi) + i\sin(\theta+\phi)).
\end{aligned}$$

□

Remark 1.9. This proposition tells us where the product of two complex numbers lies in the plane: just add the angles and multiply the distances from 0. It also leads to a famous formula for powers of complex numbers.

Theorem 1.10 (***DeMoivre***)**.** *Let* $z = r(\cos\theta + i\sin\theta) \in \mathbb{C}$ *and* $n \in \mathbb{Z}^+$. *Then*

$$z^n = r^n(\cos n\theta + i\sin n\theta).$$

Proof. The statement is trivially true if $n = 1$ and follows from the previous Proposition if $n = 2$. So for induction, suppose it's true for $n = k$. Then

$$\begin{aligned}
z^{k+1} = z^k\cdot z &= r^k(\cos k\theta + i\sin k\theta)\cdot r(\cos\theta + i\sin\theta)\\
&= r^k r(\cos(k\theta+\theta) + i\sin(k\theta+\theta))\\
&= r^{k+1}(\cos(k+1)\theta + i\sin(k+1)\theta).
\end{aligned}$$

□

Definition 1.11 (***Euler***)**.** $e^{i\theta} = \cos\theta + i\sin\theta$.

Motivation 1.12. This strange definition is in fact natural, because what we've just done is show that the usual exponential rules hold:

$$re^{i\theta}\cdot se^{i\phi} = rse^{i(\theta+\phi)}$$

$$(re^{i\theta})^n = r^n re^{in\theta}$$

But Euler was led to the definition because he was the absolute master of infinite series:

$$\begin{aligned}
e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \dots\\
&= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \dots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots\right)\\
&= \cos\theta + i\sin\theta.
\end{aligned}$$

From this brilliant definition, we get perhaps the most famous equation in mathematics, relating the four most important constants...

Theorem 1.13. $e^{i\pi} = -1$.

Proof. Just take $\theta = \pi$ in Euler's definition. □

2. Inner Products and Norms

Remark 2.1. For the rest of this chapter, any reference to a field F will mean that either $F = \mathbb{R}$ or $F = \mathbb{C}$. We'll freely write $\overline{\alpha}$ for the conjugate of $\alpha \in F$, since the conjugate of any real number is itself.

Definition 2.2. Let V be a vector space over F. An inner product on V is a function

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

satisfying

(1) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle, \forall x, y, z \in V$.
(2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \forall x, y \in V, \alpha \in F$.
(3) $\langle x, y\rangle = \overline{\langle y, x\rangle}, \forall x, y \in V$.
(4) $\langle x, x\rangle \in \mathbb{R}$ and if $x \neq 0$, then $\langle x, x\rangle > 0$.

A vector space equipped with such a function is called an inner product space.

Example 2.3. The familiar dot product

$$(\alpha_1, \ldots, \alpha_n) \bullet (\beta_1, \ldots, \beta_n) = \alpha_1\beta_1 + \ldots + \alpha_n\beta_n$$

is an inner product on $\mathbb{R}^n$. It's the model that we're generalizing to other real or complex vector spaces.

Example 2.4. An analogue of the dot product, called the *Frobenius product*, defined by

$$(z_1, \ldots, z_n) \bullet (w_1, \ldots, w_n) = z_1\overline{w_1} + \ldots + z_n\overline{w_n},$$

makes $\mathbb{C}^n$ an inner product space.

Example 2.5. Let $[a, b] \subseteq \mathbb{R}$ be a closed interval, and let $C([a, b])$ be the real vector space of continuous functions $f : [a, b] \to \mathbb{R}$. An inner product on $C([a, b])$ can be defined by

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt.$$

Example 2.6. Let $[a, b] \subseteq \mathbb{R}$ be a closed interval, and let $\mathcal{C}([a, b])$ be the complex vector space of continuous functions $f : [a, b] \to \mathbb{C}$. An inner product on $\mathcal{C}([a, b])$ can be defined by

$$\langle f, g\rangle = \int_a^b f(t)\overline{g(t)}dt.$$

Proposition 2.7. *Let V be an inner product space, $x, y, z \in V$, and $\alpha \in F$.*

(1) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
(2) $\langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle$.
(3) $\langle 0, x\rangle = \langle x, 0\rangle = 0$.
(4) $\langle x, x\rangle = 0 \iff x = 0$.
(5) $\langle x, y\rangle = \langle x, z\rangle, \forall x \iff y = z$.

Proof.

(1)
$$\begin{aligned}\langle x, y+z\rangle &= \overline{\langle y+z, x\rangle}\\ &= \overline{\langle y, x\rangle + \langle z, x\rangle}\\ &= \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle}\\ &= \overline{\overline{\langle x, y\rangle}} + \overline{\overline{\langle x, z\rangle}}\\ &= \langle x, y\rangle + \langle x, z\rangle.\end{aligned}$$

(2)
$$\langle x, \alpha y\rangle = \overline{\langle \alpha y, x\rangle} = \overline{\alpha\langle y, x\rangle} = \overline{\alpha}\,\overline{\langle y, x\rangle} = \overline{\alpha}\,\overline{\overline{\langle x, y\rangle}} = \overline{\alpha}\langle x, y\rangle.$$

(3) $\langle 0, x\rangle = \langle 0\cdot 0, x\rangle = 0\cdot\langle 0, x\rangle = 0$. A similar argument proves the other equality, since $\overline{0} = 0$.

(4) This follows immediately from the third part of this proposition and the fourth condition in the definition of inner product.

(5) ($\Longleftarrow$): If $y = z$, the statement is obvious.
($\Longrightarrow$): If $\langle x, y\rangle = \langle x, z\rangle, \forall x$, then
$$\langle y-z, y-z\rangle = \langle y, y\rangle - \langle y, z\rangle - \langle z, y\rangle + \langle z, z\rangle = 0,$$
since $\langle y, y\rangle = \langle y, z\rangle$ and $\langle z, y\rangle = \langle z, z\rangle$ by the assumption. Thus $y - z = 0$.

□

Definition 2.8. Let V be an inner product space. The norm of $x \in V$ is
$$||x|| = \sqrt{\langle x, x\rangle} \in \mathbb{R}.$$

Proposition 2.9. *Let V be an inner product space, $x, y \in V$, and $\alpha \in F$.*

(1) $||\alpha x|| = |\alpha|\cdot||x||$.
(2) $||x|| = 0 \iff x = 0$.
(3) The Cauchy-Schwarz Inequality: $|\langle x, y\rangle| \leq ||x||\cdot||y||$.
(4) The Generalized Triangle Inequality: $||x+y|| \leq ||x|| + ||y||$.

Proof.

(1) $\langle \alpha x, \alpha x\rangle = \alpha\overline{\alpha}\langle x, x\rangle = |\alpha|^2\langle x, x\rangle \implies ||\alpha x|| = \sqrt{|\alpha|^2\langle x, x\rangle} = |\alpha|\cdot||x||$.

(2) $\sqrt{\langle x, x\rangle} = 0 \iff \langle x, x\rangle = 0$.

(3) If $y = 0$, both sides of the inequality are 0, so we may assume $y \neq 0$. Now if $\alpha \in F$,
$$0 \leq ||x-\alpha y||^2 = \langle x-\alpha y, x-\alpha y\rangle = \langle x, x\rangle - \overline{\alpha}\langle x, y\rangle - \alpha\langle y, x\rangle + \alpha\overline{\alpha}\langle y, y\rangle.$$
If we take $\alpha = \frac{\langle x,y\rangle}{\langle y,y\rangle}$, this inequality becomes
$$0 \leq \langle x, x\rangle - \frac{\overline{\langle x, y\rangle}\langle x, y\rangle}{\langle y, y\rangle} - \frac{\langle x, y\rangle\langle y, x\rangle}{\langle y, y\rangle} + \frac{\langle x, y\rangle\overline{\langle x, y\rangle}}{\langle y, y\rangle}.$$

But the last two terms cancel, so we have

$$0 \leq ||x||^2 - \frac{|\langle x, y\rangle|^2}{||y||^2}.$$

The result follows easily.

(4)

$$\begin{aligned}
||x+y||^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= ||x||^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} + ||y||^2 \\
&= ||x||^2 + 2\Re(\langle x, y\rangle) + ||y||^2 \\
&\leq ||x||^2 + 2|\langle x, y\rangle| + ||y||^2 \text{ (by Proposition 1.6)} \\
&\leq ||x||^2 + 2||x|| \cdot ||y|| + ||y||^2 \text{ (by Cauchy-Schwarz)} \\
&= (||x|| + ||y||)^2.
\end{aligned}$$

Taking square roots completes the proof.

□

Remark 2.10. Applying the Cauchy-Schwarz and Triangle inequalities to the dot product in $\mathbb{R}^n$ give results that become very useful in Mathematical Analysis. Letting $(\alpha_1, \ldots, \alpha_n), (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$, we see that

$$\left|\sum_{i=1}^{n} \alpha_i \beta_i\right| \leq \left(\sum_{i=1}^{n} \alpha_i^2\right)^{1/2} \cdot \left(\sum_{i=1}^{n} \beta_i^2\right)^{1/2}$$

and

$$\left(\sum_{i=1}^{n} (\alpha_i + \beta_i)^2\right)^{1/2} \leq \left(\sum_{i=1}^{n} \alpha_i^2\right)^{1/2} + \left(\sum_{i=1}^{n} \beta_i^2\right)^{1/2}.$$

3. Orthogonality

Definition 3.1. A set S of nonzero vectors in an inner product space V is orthogonal if $\langle x, y\rangle = 0$, for all $x, y \in S$ with $x \neq y$. If in addition $||x|| = 1$ for all $x \in S$, then S is said to orthonormal.

Example 3.2. In $\mathbb{R}^3$, the set $S = \{(1,1,0), (1,-1,1), (-1,1,2)\}$ is orthogonal. We can then create the orthonormal set $S' = \{\frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{3}}(1-1,1), \frac{1}{\sqrt{6}}(-1,1,2)\}$.

Example 3.3. In V, the inner product space of continuous functions $f : [0, 2\pi] \to \mathbb{C}$, let

$$f_n(t) = e^{int} = \cos nt + i \sin nt, \text{ for } n \in \mathbb{Z}.$$

Then if $n \neq m$,

$$\begin{aligned}
\langle f_n, f_m \rangle &= \int_0^{2\pi} e^{int} \cdot \overline{e^{imt}} dt \\
&= \int_0^{2\pi} e^{int} \cdot e^{-imt} dt \\
&= \int_0^{2\pi} e^{i(n-m)t} dt \\
&= \frac{e^{i(n-m)t}}{n-m} \Big[_0^{2\pi} \\
&= \frac{1}{n-m} \cdot (1-1) \\
&= 0.
\end{aligned}$$

Thus $\{f_n : n \in \mathbb{Z}\}$ is an orthogonal set, and $\{\frac{1}{2\pi} \cdot f_n : n \in \mathbb{Z}\}$ an orthonormal set.

Theorem 3.4. *An orthogonal set is linearly independent.*

Proof. Let S be orthogonal, let $x_1, \ldots, x_n \in S$, and suppose $x = \alpha_1 x_1 + \ldots + \alpha_n x_n = 0$. Then if $1 \leq i \leq n$,

$$\begin{aligned}
0 = \langle 0, x_i \rangle = \langle x, x_i \rangle &= \langle \alpha_1 x_1 + \ldots + \alpha_n x_n, x_i \rangle \\
&= \alpha_1 \langle x_1, x_i \rangle + \ldots + \alpha_n \langle x_n, x_i \rangle \\
&= \alpha_i \langle x_i, x_i \rangle.
\end{aligned}$$

Since $x_i \neq 0$, we see that $\alpha_i = 0$. □

Remark 3.5. A very nice computational tool to have available in an inner product space V would be a basis $\mathcal{B} = \{b_1, \ldots, b_n\}$ that was also an orthonormal set. Why? Because we could easily compute the $\mathcal{B}$-coefficients of a vector $x \in V$:

$$x = \alpha_1 b_1 + \ldots + \alpha_n b_n \implies \langle x, b_i \rangle = \alpha_i \langle b_i, b_i \rangle = \alpha_i.$$

Theorem 3.6 (***Gram-Schmidt***)**.** *Every nontrivial finite dimensional inner product space has an orthonormal basis.*

Proof. Let $\{x_1, \ldots, x_n\}$ be a basis of V. Construct a set $\mathcal{B} = \{b_1, \ldots, b_n\}$ as follows:

$$b_1 = x_1$$

$$b_k = x_k - \sum_{j=1}^{k-1} \frac{\langle x_k, b_j \rangle}{||b_j||^2} b_j, \text{ for } 2 \leq k \leq n.$$

We claim that $\mathcal{B}$ is orthogonal. We proceed inductively by noting that $\{b_1\}$ is trivially orthogonal, and assuming that $\{b_1, \ldots b_{k-1}\}$ is orthogonal. Then if $i < k$,

$$\langle b_k, b_i \rangle = \langle x_k, b_i \rangle - \sum_{j=1}^{k-1} \frac{\langle x_k, b_j \rangle}{||b_j||^2} \langle b_j, b_i \rangle = \langle x_k, b_i \rangle - \frac{\langle x_k, b_i \rangle}{||b_i||^2} \langle b_i, b_i \rangle = \langle x_k, b_i \rangle - \langle x_k, b_i \rangle = 0.$$

Thus $\mathcal{B}$ is an orthogonal, and hence independent, set of n vectors, and is therefore a basis. Normalizing (that is, dividing each b_i by $||b_i||$) produces an orthonormal basis. □

Definition 3.7. Let W be a subspace of the inner product space V. The orthogonal complement of W is

$$W^{\perp} = \{x \in V : \langle x, w\rangle = 0\}, \text{ for all } w \in W.$$

Example 3.8. In $\mathbb{R}^3$, $\{0\}^{\perp} = \mathbb{R}^3$ and $(\mathbb{R}^3)^{\perp} = \{0\}$. If ℓ is a 1-dimensional subspace, a line through the origin, then $\ell^{\perp}$ is the plane $\mathcal{P}$ through the origin perpendicular to ℓ, a 2-dimensional subspace. Dually, $\mathcal{P}^{\perp} = \ell$.

Proposition 3.9. *$W^{\perp}$ is a subspace of V.*

Proof. $0 \in W^{\perp}$, since $\langle 0, w\rangle = 0$ for all $w \in W$. If $x, y \in W^{\perp}$,

$$\langle x + y, w\rangle = \langle x, w\rangle + \langle y, w\rangle = 0 + 0 = 0,$$

so $x + y \in W^{\perp}$. Finally, if $\alpha \in F$ and $x \in W^{\perp}$, then

$$\langle \alpha x, w\rangle = \alpha\langle x, w\rangle = \alpha \cdot 0 = 0,$$

so $\alpha x \in W^{\perp}$. $\square$

Proposition 3.10. $W \cap W^{\perp} = \{0\}$.

Proof. If $w \in W \cap W^{\perp}$, then $\langle w, w\rangle = 0$, so $w = 0$. $\square$

Theorem 3.11. *Let $x \in V$. Then for any subspace W of V, x can be uniquely expressed as $x = w + w^{\perp}$, where $w \in W$ and $w^{\perp} \in W^{\perp}$.*

Proof. First we show that such an expression is possible. If $W = \{0\}$, then $W^{\perp} = V$, and of course $x = 0 + x$. Otherwise, take an orthonormal basis $\{w_1, \ldots, w_k\}$ of W, and define

$$w = \sum_{i=1}^{k} \langle x, w_i\rangle w_i \in W \text{ and } w^{\perp} = x - w.$$

Then we need to show that $w^{\perp} \in W^{\perp}$, and it suffices to show $\langle w^{\perp}, w_j\rangle = 0$. But

$$\langle w^{\perp}, w_j\rangle = \langle x - w, w_j\rangle = \langle x, w_j\rangle - \sum_{i=1}^{k} \langle x, w_i\rangle\langle w_i, w_j\rangle = \langle x, w_j\rangle - \langle x, w_j\rangle = 0.$$

For uniqueness, suppose that $w + w^{\perp} = u + u^{\perp}$, where $w, u \in W$ and $w^{\perp}, u^{\perp} \in W^{\perp}$. Then

$$w - u = u^{\perp} - w^{\perp} \in W \cap W^{\perp} = \{0\} \implies w = u \text{ and } w^{\perp} = u^{\perp}.$$

$\square$

Definition 3.12. Let W and U be subspaces of the vector space V such that $V = W + U$ and $W \cap U = \{0\}$. Then V is the direct sum of W and U, written $V = W \oplus U$. Equivalently, $V = W \oplus U$ if and only every element of V is uniquely the sum of elements from W and U.

Corollary 3.13. *If W is a subspace of the inner product space V, then $V = W \oplus W^{\perp}$.* $\square$

4. Exercises

Exercise 4.1. In $C([0,1])$, let $f(t) = t$ and $g(t) = e^t$. Compute $\langle f, g\rangle, ||f||, ||g||$, and $||f+g||$. Verify the Cauchy-Schwarz inequality and the Triangle Inequality.

Exercise 4.2. Prove that the Parallelogram Law holds in any inner product space V:

$$||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2, \forall x, y \in V.$$

Exercise 4.3. Let T be a linear operator on the inner product space V. Show that if $||Tx|| = ||x||, \forall x \in V$, then T is injective.

Exercise 4.4. Prove *Parseval's Identity*: if $\mathcal{B} = \{b_1, \ldots, b_n\}$ is an orthonormal basis of the inner product space V, then

$$\langle x, y\rangle = \sum_{i=1}^{n} \langle x, b_i\rangle \overline{\langle y, b_i\rangle}, \forall x, y \in V.$$