1 | Vector Spaces

1.1 The Algebra of Matrices over a Field

Definition. By a **field** F, we mean a non-empty set of elements with two laws of combination, which we call an **addition** $+$ and a **multiplication** $\cdot$ satisfying:

(F1) To every pair of elements $a, b \in F$ there is associated a unique element, called their **sum**, which we denote by $a + b$.
(F2) Addition is associative: $(a + b) + c = a + (b + c)$.
(F3) Addition is commutative: $a + b = b + a$.
(F4) There exists an element, which we denote by 0, such that $a + 0 = a$ for all $a \in F$.
(F5) For each $a \in F$ there exists an element, which we denote by $-a$ such that $a + (-a) = 0$.
(F6) To every pair of elements $a, b \in F$ there is associated a unique element, called their **product**, which we denote by ab, or $a \cdot b$.
(F7) Multiplication is associative: $(ab)c = a(bc)$.
(F8) Multiplication is commutative: $ab = ba$.
(F9) There exists an element different from 0, which we denote by 1, such that $a \cdot 1 = a$ for all $a \in F$.
(F10) For each $a \in F$, $a \neq 0$, there exists an element which we denote by a^{-1}, such that $a \cdot a^{-1} = 1$.
(F11) Multiplication is distributive with respect to addition: $(a + b)c = ac + bc$.

Remark. Note that in a field F, $0 + 0 = 0$.

We write $\mathbb{Q}$ for the set of rational numbers, $\mathbb{R}$ for the set of real numbers and $\mathbb{C}$ for the set of complex numbers. These sets are fields. The rigorous definition and treatments on fields can be found in any abstract algebra courses including 2301337 Abstract Algebra I. The definition of field was presented once in Linear Algebra I. In this course, F always denotes any of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or other fields. Its members are called **scalars**. However, almost nothing essential is lost if we assume that F is the real field $\mathbb{R}$ or the complex field $\mathbb{C}$.

Example 1.1.1. A non-empty subset F of $\mathbb{C}$ such that for any $x, y \in F$, $x - y \in F$ and $xy \in F$ and for any non-zero $z \in F$, $1/z \in F$ is also a field. It is called a **subfield of** $\mathbb{C}$. For example, $\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$.

Example 1.1.2. Let p be a prime and $\mathbb{F}_p = \{0, 1, \ldots, p - 1\}$. For a and b in $\mathbb{F}_p$, we define

$$\begin{aligned} a + b &= \text{the remainder when we divide } a + b \text{ with } p, \text{ and} \\ ab &= \text{the remainder when we divide } ab \text{ with } p. \end{aligned}$$

Then $(\mathbb{F}_p, +, \cdot)$ is a finite field of p elements. Note that if $p = 2$, we have $1 + 1 = 0$.

Definition. Let F be a field. An $m \times n$ (m by n) **matrix** A with m rows and n columns with entries over F is a rectangular array of the form

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix},$$

where $a_{ij} \in F$ for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$. We write $M_{m,n}(F)$ for the set of $m \times n$ matrices with entries in F and we write $M_n(F)$ for $M_{n,n}(F)$ the set of square matrices of order n.

Remark. As a shortcut, we often use the notation $A = [a_{ij}]$ to denote the matrix A with entries a_{ij}. Notice that when we refer to the matrix we put parentheses—as in "$[a_{ij}]$", and when we refer to a specific entry we do not use the surrounding parentheses—as in "a_{ij}."

Definition. Two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are **equal** if $a_{ij} = b_{ij}$ for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$.

Definition. The $m \times n$ **zero matrix** $\mathbf{0}_{m\times n} \in M_{m,n}(F)$ is the matrix with 0_F's everywhere,

$$\mathbf{0}_{m\times n} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

When $m = n$ we write $\mathbf{0}_n$ as an abbreviation for $\mathbf{0}_{n\times n}$.
The $n \times n$ **identity matrix** $I_n \in M_n(F)$ is the matrix with 1's on the main diagonal and 0's everywhere else,

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

Definition. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices and a scalar $r \in F$. The matrix $A + rB$ is the matrix $C \in M_{m,n}(F)$ with entries $C = [c_{ij}]$ where

$$c_{ij} = a_{ij} + rb_{ij}.$$

Theorem 1.1.1. *Let A, B and C be matrices of the same size, and let r and s be scalars in F. Then*

(a) $A + B = B + A$
(b) $(A + B) + C = A + (B + C)$
(c) $A + \mathbf{0} = A$
(d) $r(A + B) = rA + rB$
(e) $r\mathbf{0} = \mathbf{0}$ *and* $0A = \mathbf{0}$
(f) $1A = A$
(g) $(r + s)A = rA + sA$
(h) $r(sA) = (rs)A = (sr)A = s(rA)$

Definition. Let A be an $m \times n$ matrix with columns $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ and $\vec{x}$ is a column vector in F^n. The **product of** A **and** $\vec{x}$ denoted by $A\vec{x}$ is the linear combination of the columns of A using the corresponding entries in $\vec{x}$ as weights. That is,

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} := x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n.$$

If B is an $n \times p$ matrix with columns $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p$, then the **product of** A **and** B, denoted by AB, is the $m \times p$ matrix with columns $A\vec{b}_1, A\vec{b}_2, \ldots, A\vec{b}_p$. In other words,

$$AB = A\begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{bmatrix} := \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \end{bmatrix}.$$

The above definition of AB is a good for theoretical work. When A and B have small sizes, the following method is more efficient when working by hand. Let $A = [a_{ij}] \in M_{m,n}(F)$ and $B = [b_{ij}] \in M_{n,p}(F)$. Then the matrix product AB is defined as the matrix $C = [c_{ij}] \in M_{m,p}(F)$ with entries

$$c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj},$$

that is,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & \cdots & \vdots \\ \cdots & c_{ij} & \cdots \\ \vdots & \cdots & \vdots \\ c_{m1} & \cdots & c_{mp} \end{bmatrix}.$$

If A is a square matrix of order n, then we write A^k for $\underbrace{A \cdots A}_{k \text{ copies}}$.

Theorem 1.1.2. *Let A be $m \times n$ and let B and C have sizes for which the indicated sums and products are defined.*
(a) $A(B + C) = AB + AC$ *and* $(B + C)A = BA + CA$
(b) $r(AB) = (rA)B = A(rB)$ *for any scalar* r
(c) $A\mathbf{0}_{n\times k} = \mathbf{0}_{m\times k}$ *and* $\mathbf{0}_{k\times m}A = \mathbf{0}_{k\times n}$
(d) $I_mA = A = AI_n$
(e) $A(BC) = (AB)C$

Remarks. Properties above are analogous to properties of real numbers. But **NOT ALL** real number properties correspond to matrix properties.
1. It is not the case that AB always equal to BA.
2. Even if $A \neq \mathbf{0}$ and $AB = AC$, then B may not equal to C. (A must have an inverse!)
3. It is possible for $AB = 0$ even if $A \neq 0$ and $B \neq 0$. E.g., $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Definition. The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix obtained from A by interchanging the rows and columns. We denote the transpose of A by A^T. That is, if $A = [a_{ij}]_{m\times n}$, then $A^T = [b_{ji}]_{n\times m}$ where $b_{ji} = a_{ij}$ for all i, j. Moreover,

$$\vec{x}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}$$

and so if $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$, then

$$A^T = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_n^T \end{bmatrix}.$$

Theorem 1.1.3. *Let A and B denote matrices whose sizes are appropriate for the following sums and products.*

(a) $(A^T)^T = A$
(b) $(A + B)^T = A^T + B^T$
(c) $(rA)^T = rA^T$ *for any scalar* r
(d) $(AB)^T = B^T A^T$

1.2 Axioms of a Vector Space

Definition. A **vector space** V over a field F is a nonempty set of elements called **vectors**, which two laws of combination, called **vector addition** (or **addition**) and **scalar multiplication**, satisfying the following conditions.

(A1) $\forall \vec{u}, \vec{v} \in V, \vec{u} + \vec{v} \in V$.
(A2) $\forall \vec{u}, \vec{v} \in V, \vec{u} + \vec{v} = \vec{v} + \vec{u}$.
(A3) $\forall \vec{u}, \vec{v} \in V, \vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$.
(A4) $\exists \vec{0} \in V, \forall \vec{u} \in V, \vec{u} + \vec{0} = \vec{u} = \vec{0} + \vec{u}$.
(A5) $\forall \vec{u} \in V, \exists \vec{u}' \in V, \vec{u} + \vec{u}' = \vec{0} = \vec{u}' + \vec{u}$.
(SM1) $\forall a \in F, \forall \vec{u} \in V, a\vec{u} \in V$.
(SM2) $\forall a \in F, \forall \vec{u}, \vec{v} \in V, a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$.
(SM3) $\forall a, b \in F, \forall \vec{u} \in V, (a + b)\vec{u} = a\vec{u} + b\vec{u}$.
(SM4) $\forall a, b \in F, \forall \vec{u} \in V, (ab)\vec{u} = a(b\vec{u})$.
(SM5) $\forall \vec{u} \in V, 1\vec{u} = \vec{u} \quad (1 \in F)$.

We call $\vec{0}$ the **zero vector** and $\vec{u}'$ the **negative of** $\vec{u}$.

Theorem 1.2.1. *Let V be a vector space over a field F. Then*

1. *(Cancellation)* $\forall \vec{u}, \vec{v}, \vec{w} \in V, \vec{u} + \vec{w} = \vec{v} + \vec{w} \Rightarrow \vec{u} = \vec{v}$ *and* $\forall \vec{u}, \vec{v}, \vec{w} \in V, \vec{w} + \vec{u} = \vec{w} + \vec{v} \Rightarrow \vec{u} = \vec{v}$.
2. *The zero vector and the negative of $\vec{u}$ are unique. We shall denote the negative of $\vec{u}$ by $-\vec{u}$.*
3. $\forall \vec{v} \in V, -(-\vec{v}) = \vec{v}$.
4. $\forall \vec{v} \in V, 0\vec{v} = \vec{0}$.
5. $\forall a \in F, a\vec{0} = \vec{0}$.
6. $\forall a \in F, \forall \vec{v} \in V, (-a)\vec{v} = -(a\vec{v}) = a(-\vec{v})$. *In particular,* $(-1)\vec{v} = -(1\vec{v}) = -\vec{v}$.
7. $\forall a \in F, \forall \vec{v} \in V, a\vec{v} = \vec{0} \Rightarrow (a = 0 \vee \vec{v} = \vec{0})$.

Examples 1.2.1. 1. For any field F and $n \geq 1$, we have F^n is a vector space over F where

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and

$$a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n)$$

for all $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in F^n$ and $a \in F$

2. Let $m, n \in \mathbb{N}$, F is a field and $M_{m,n}(F)$ the set of all $m \times n$ matrices over F. Then $M_{m,n}(F)$ is a vector space over F under the usual addition and scalar multiplication of matrices.
3. [The space of functions from a set to a field] Let S be a nonempty set and F a field. Let $F^S = \{f \mid f : S \to F\}$. Then F^S is a vector space over F by defining $f + g$ and af for functions $f, g \in F^S$ and a scalar $c \in F$ as follows:

$$(f+g)(t) = f(t) + g(t) \quad \text{and} \quad (cf)(t) = cf(t)$$

for all $t \in S$. The zero function from S into F is the zero vector of F and the negative of $f \in V$ is $-f$ defined by $(-f)(t) = -f(t)$ for all $t \in S$.
4. [The sequence space] Let $F^{\mathbb{N}} = \{(x_n) : (x_n) \text{ is a sequence in } F\}$. Then $F^{\mathbb{N}}$ is a vector space over F under the usual addition and scalar multiplication of sequences. That is, for sequences (a_n) and (b_n) in $F^{\mathbb{N}}$ and a scalar $c \in F$,

$$(a_n) + (b_n) = (a_n + b_n) \quad \text{and} \quad c(a_n) = (c\, a_n).$$

Its zero is the zero sequence (z_n) where $z_n = 0$ for all n and the negative of (a_n) is the sequence (b_n) given by $b_n = -a_n$ for all n.
5. Let n be a non-negative integer and $F_n[x]$ be the set of polynomials over F of degree at most n. That is,

$$F_n[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : a_i \in F \text{ for all } i \in \{0, 1, 2, \ldots, n\}\}.$$

We define the addition and scalar multiplication by

$$\begin{aligned} p(x) + q(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n \\ \text{and } c(p(x)) &= (ca_0) + (ca_1)x + (ca_2)x^2 + \cdots + (ca_n)x^n \end{aligned}$$

for all polynomials $p(x) = a_0+a_1x+a_2x^2+\cdots+a_nx^n$ and $q(x) = b_0+b_1x+b_2x^2+\cdots+b_nx^n$ in $F_n[x]$ and $c \in F$. Then $F_n[x]$ is a vector space over F. Observe that for each positive integer n, we have $F_{n-1}[x] \subset F_n[x]$.
6. [The space of polynomials over a field] Let $F[x]$ be the set of all polynomials over F. That is,

$$F[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : n \geq 0 \text{ and } a_i \in F \text{ for all } i \in \{0, 1, 2, \ldots, n\}\}.$$

Then $F[x] = \bigcup_{n \geq 0} F_n[x]$. If we use the addition and scalar multiplication defined for $F_n[x]$, then $F[x]$ is a vector space over F. The zero polynomial $0(x) = 0 + 0x + 0x^2 + \cdots$ is its zero vector and for $f(x) = c_0 + c_1x + \cdots + c_nx^n \in F[x]$, the negative of $f(x)$ is $(-f)(x) = (-c_0) + (-c_1)x + \cdots + (-c_n)x^n$.

Theorem 1.2.2. *Let $(V_1, +_1, \cdot_1), (V_2, +_2, \cdot_2), \ldots, (V_n, +_n, \cdot_n)$ be vector spaces over a field F. For $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n), (\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n) \in V$ and $c \in F$, we define the addition and scalar multiplication on $V = V_1 \times V_2 \times \cdots \times V_n$ by*

$$\begin{aligned} (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) + (\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n) &= (\vec{v}_1 +_1 \vec{w}_1, \vec{v}_2 +_2 \vec{w}_2, \ldots, \vec{v}_n +_n \vec{w}_n) \\ \textit{and} \qquad c(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) &= (c \cdot_1 \vec{v}_1, c \cdot_2 \vec{v}_2, \ldots, c \cdot_n \vec{v}_n). \end{aligned}$$

Then V is a vector space over F with the zero vector $\vec{0} = (\vec{0}_1, \vec{0}_2, \ldots, \vec{0}_n)$ and the negative of $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ is $(-\vec{v}_1, -\vec{v}_2, \ldots, -\vec{v}_n)$. V is called **the direct product of** $V_1, V_2, \ldots, V_n$.

1.3 Subspaces

Definition. Let V be a vector space over a field F. A **subspace** of V is a subset of V which is itself a vector space over F with the operations of vector addition and scalar multiplication of V.

Theorem 1.3.1. *Let W be a nonempty subset of V. Then the following statements are equivalent.*

(i) *W is a subspace of V.*
(ii) $\forall \vec{u}, \vec{v} \in W, \forall c \in F, \vec{u} + \vec{v} \in W$ *and* $c\vec{u} \in W$.
(iii) $\forall \vec{u}, \vec{v} \in W, \forall c, d \in F, c\vec{u} + d\vec{v} \in W$.
(iv) $\forall \vec{u}, \vec{v} \in W, \forall c \in F, c\vec{u} + \vec{v} \in W$.

Examples 1.3.1.

1. For any vector space V over a field F, we have $\{\vec{0}_V\}$ and V are subspaces of V, called **trivial subspaces**.
2. For a non-negative integer n, we have $F_n[x]$ is a subspace of $F[x]$.
3. Let $\alpha \in F$ and $V_\alpha = \{(x_1, x_2) : x_1 = \alpha x_2\}$. Then V_α is a subspace of F^2.
4. Let $\mathrm{Bd}(\mathbb{R}) = \{(a_n) \in \mathbb{R}^{\mathbb{N}} : (a_n)$ is a bounded sequence$\}$,
   $\mathrm{C}(\mathbb{R}) = \{(a_n) \in \mathbb{R}^{\mathbb{N}} : (a_n)$ is a convergent sequence$\}$ and
   $\mathrm{C}_0(\mathbb{R}) = \{(a_n) \in \mathbb{R}^{\mathbb{N}} : a_n \to 0 \text{ as } n \to \infty\}$.
   Then $\mathrm{Bd}(\mathbb{R})$, $\mathrm{C}(\mathbb{R})$ and $\mathrm{C}_0(\mathbb{R})$ are subspaces of $\mathbb{R}^{\mathbb{N}}$.
5. Let $C^0(-\infty, \infty) = \{f \in \mathbb{R}^{\mathbb{R}} : f$ is continuous on $(-\infty, \infty)\}$.
   Then $C^0(-\infty, \infty)$ is a subspace of $\mathbb{R}^{\mathbb{R}}$.
6. Let $W = \{f : \mathbb{R} \to \mathbb{R} \mid f'' = f\}$. Then W is a subspace of $\mathbb{R}^{\mathbb{R}}$.
7. Let $W_1 = \{p(x) \in F[x] : p(1) = 0\}$ and $W_2 = \{p(x) \in F[x] : p(0) = 1\}$.
   Then W_1 is a subspace of $F[x]$ but W_2 is not.
8. Let $A \in M_{m,n}(F)$. Then $\mathrm{Nul}\, A = \{\vec{x} \in F^n : A\vec{x} = \vec{0}_m\}$ is a subspace of F^n, called the **null space of** A.

Theorem 1.3.2. *Let V be a vector space over a field F. The intersection of any collection of subspaces of V is a subspace of V.*

Definition. For non-empty subsets $S_1, S_2, \ldots, S_n$ of V, we define

$$S_1 + S_2 + \cdots + S_n = \sum_{i=1}^{n} S_i = \{x_1 + x_2 + \cdots + x_n : x_1 \in S_1, x_2 \in S_2, \ldots, x_n \in S_n\}.$$

Theorem 1.3.3. *If $W_1, \ldots, W_n$ are subspaces of V, then $W_1 + \cdots + W_n$ is a subspace of V.*

Remark. $W_1 + W_2$ is the smallest subspace of V containing W_1 and W_2, i.e., any subspace containing W_1 and W_2 must contain $W_1 + W_2$.

Definition. Let V be a vector space over a field F.
A vector $\vec{v}$ is said to be a **linear combination of** $\vec{v}_1, \ldots, \vec{v}_n \in V$ if

$$\exists a_1, \ldots, a_n \in F, \vec{v} = a_1\vec{v}_1 + \cdots + a_n\vec{v}_n.$$

Definition. Let $S \subseteq V$. The **subspace of** V **spanned by** S is defined to be the intersection of all subspaces of V containing S. We denote this subspace by $\mathrm{Span}\ S$.
For $\vec{v}_1, \ldots, \vec{v}_p \in V$, we call $\mathrm{Span}\{\vec{v}_1, \ldots, \vec{v}_p\}$ the **subspace of** V **spanned by** $\vec{v}_1, \ldots, \vec{v}_p$.

Since $\emptyset \subset \{\vec{0}_V\}$ which is the smallest of all subspaces of V, we have $\operatorname{Span} = \{\vec{0}_V\}$. Moreover, if W is a subspace of V, then $\operatorname{Span} W = W$. In particular, $\operatorname{Span}(\operatorname{Span} S) = \operatorname{Span} S$.

Remark. Let S be a non-empty subset of V and let W be a subspace of V containing S. Note that for $c_1, \ldots, c_m \in F$ and $\vec{v}_1, \ldots \vec{v}_m \in S$, we have $\vec{v}_1, \ldots \vec{v}_m \in W$ and so

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m \in W.$$

Thus, $Y := \{c_1\vec{v}_1 + \cdots + c_m\vec{v}_m : c_1, \ldots, c_m \in F \text{ and } \vec{v}_1, \ldots, \vec{v}_m \in S \text{ for some } m \in \mathbb{N}\} \subseteq W$ for all subspaces W of V containing S. Hence, $Y \subseteq \operatorname{Span} S$.

Theorem 1.3.4. *$\operatorname{Span} S$ is the smallest subspace of V containing S. That is, any subspace of V containing S must also contain $\operatorname{Span} S$. Moreover, $\operatorname{Span} \emptyset = \{\vec{0}\}$ and*

$$\operatorname{Span} S = \{c_1\vec{v}_1 + \cdots + c_m\vec{v}_m : c_1, \ldots, c_m \in F \text{ and } \vec{v}_1, \ldots, \vec{v}_m \in S \text{ for some } m \in \mathbb{N}\} \text{ if } S \neq \emptyset.$$

In particular,

$$\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_p\} = \{c_1\vec{v}_1 + \cdots + c_p\vec{v}_p : c_1, \ldots, c_p \in F\}.$$

Definition. Let $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \ldots & \vec{a}_n \end{bmatrix}$ be an $m \times n$ matrix over a field F. Then $\vec{a}_i \in F^m$ for all $i = 1, 2, \ldots, n$ and $\operatorname{Span}\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ is a subspace of F^m, called the **column space of** A. We denote this space by $\operatorname{Col} A$.

By Theorem 1.3.4, we have

$$\operatorname{Col} A = \{c_1\vec{a}_1 + c_2\vec{a}_2 + \cdots + c_n\vec{a}_n : c_1, c_2, \ldots, c_n \in F\}.$$

Definition. Let V and W be vector spaces over a field F. A function $T : V \to W$ is said to be a **linear transformation** if the following conditions are satisfied:
(i) $\forall \vec{u}, \vec{v} \in V, T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$ and
(ii) $\forall \vec{u} \in V, \forall c \in F, T(c\vec{u}) = cT(\vec{u})$.

Theorem 1.3.5. *Let V and W be vector spaces over a field F and $T : V \to W$ a linear transformation. Then $T(\vec{0}_V) = \vec{0}_W$ and $\forall \vec{v} \in V, T(-\vec{v}) = -T(\vec{v})$.*

Theorem 1.3.6. *The following statements are equivalent.*
(i) *T is a linear transformation.*
(ii) $\forall \vec{u}, \vec{v} \in V, \forall c \in F, T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) \wedge T(c\vec{u}) = cT(\vec{u})$.
(iii) $\forall \vec{u}, \vec{v} \in V, \forall c, d \in F, T(c\vec{u} + d\vec{v}) = cT(\vec{u}) + dT(\vec{v})$.
(iv) $\forall \vec{u}, \vec{v} \in V, \forall c \in F, T(c\vec{u} + \vec{v}) = cT(\vec{u}) + T(\vec{v})$.

Definition. Let V and W be vector spaces over a field F and $T : V \to W$ a linear transformation. Recall that the **image** or **range** of T is given by

$$\operatorname{im} T = \operatorname{range} T = \{\vec{w} \in W : \exists \vec{v} \in V, T(\vec{v}) = \vec{w}\} = \{T(\vec{v}) : \vec{v} \in V\}.$$

The **kernel** of T is defined by

$$\ker T = \{\vec{v} \in V : T(\vec{v}) = \vec{0}_W\} = T^{-1}(\{\vec{0}_W\}).$$

Theorem 1.3.7. *The kernel of T is a subspace of V and the image of T is a subspace of W.*

Example 1.3.2. Let $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix}$ be an $m \times n$ matrix over a field F. Then the matrix transformation $T : F^n \to F^m$ given by

$$T(\vec{x}) = A\vec{x}$$

is a linear transformation. Its kernel is

$$\operatorname{Nul} A = \{\vec{x} \in F^n : A\vec{x} = \vec{0}_m\}$$

the null space of A, and its image is

$$\operatorname{im} T = \{A\vec{x} : \vec{x} \in F^n\} = \{x_1\vec{a}_1 + \cdots + x_n\vec{a}_n : x_1, \dots, x_n \in F\} = \operatorname{Col} A,$$

which is the column space of A.

Remark. Since the image of $T : \vec{x} \to A\vec{x}$ the column space of A,

$$T \text{ is onto} \Leftrightarrow \operatorname{im} T = F^m \Leftrightarrow \operatorname{Col} A = F^m.$$

If $\operatorname{Col} A = F^m$, we say that the **columns of A span F^m**.

Example 1.3.3. Let $T : \mathbb{R}[x] \to \mathbb{R}$ be defined by

$$T(p(x)) = p(1)$$

for all $p(x) \in \mathbb{R}[x]$. Show that T is an onto linear transformation and find its kernel.

Example 1.3.4. Let V be the space of differentiable functions on $(-\infty, \infty)$ with continuous derivative. Define a function $T : V \to C^0(-\infty, \infty)$ by

$$T(f(x)) = f'(x)$$

for all $f \in V$. Show that T is an onto linear transformation and find its kernel.

Definition. Let V be a vector space over a field F. Vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ in V are **linearly independent** if

$$\forall c_1, c_2, \dots, c_n \in F, c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_n\vec{u}_n = \vec{0} \Rightarrow c_1 = c_2 = \cdots = c_n = 0.$$

If there is a linear combination $c_1\vec{u}_1 + c_2\vec{u}_2 + \dots c_n\vec{u}_n = \vec{0}$ with the scalars $c_1, c_2, \dots, c_n$ not all zero, we say that $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ are **linearly dependent**.

Example 1.3.5. Determine whether the set of vectors

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

is dependent or independent in $\mathbb{R}^3$.

Example 1.3.6. Determine whether the set of vectors

$$\vec{u}_1 = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

is dependent or independent in $M_{2,3}(\mathbb{R})$.

Remarks.
1. The empty set is linearly independent.
2. If $\vec{0}_V$ is in S, then S is linearly dependent.
3. The singleton $\{\vec{0}_V\}$ is linearly dependent and $\{\vec{u}\}$ is linearly independent unless $\vec{u} = \vec{0}_V$.

Theorem 1.3.8. *Let V be a vector space over a field F and $S_1 \subseteq S_2 \subseteq V$. Then*
1. $\operatorname{Span} S_1 \subseteq \operatorname{Span} S_2$.
2. *If S_1 is linearly dependent, then S_2 is linearly dependent.*
3. *If S_2 is linearly independent, then S_1 is linearly independent.*

Example 1.3.7. Consider the space of continuous functions $C^0[-1,1]$.
Determine whether the functions $1, x, x^2$ are dependent or independent.

Remark. Observe that the question of dependence and independence of sets of functions is related to the interval over which the space is defined. Consider the same interval $[-1,1]$ with the functions f, g and h defined as follows:

$$f(x) = 1, -1 \leq x \leq 1,$$
$$g(x) = \begin{cases} 0 & \text{if } -1 \leq x \leq 0, \\ x & \text{if } 0 \leq x \leq 1. \end{cases} \quad \text{and} \quad h(x) = \begin{cases} 0 & \text{if } -1 \leq x \leq 0, \\ x^2 & \text{if } 0 \leq x \leq 1. \end{cases}$$

These functions are linearly independent. However, if we restrict these same functions to the interval $[-1,0]$, then they are dependent because

$$0 \cdot f(x) + 1 \cdot g(x) + 0 \cdot h(x) = 0$$

for $-1 \leq x \leq 0$.

Theorem 1.3.9. *Let $T : V \to W$ be a linear transformation. Then T is 1-1 $\Leftrightarrow \ker T = \{\vec{0}_V\}$.*

1.4 Bases and Dimensions

Definition. Let V be a vector space over F. A subset $\mathcal{B} \subset V$ is a **basis** for V if $\mathcal{B}$ is linearly independent and $\operatorname{Span} \mathcal{B} = V$.

Theorem 1.4.1. *Let V be a vector space over a field F and $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\} \subseteq V$ linearly independent.*
1. *If $\vec{v} \in \operatorname{Span} \mathcal{B}$, then there exist unique $c_1, \ldots, c_n \in F$ such that*
$$\vec{v} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n.$$
2. *If $\mathcal{B}$ is a basis for V, then every vector in V can be expressed uniquely as a linear combination of $\vec{v}_1, \ldots, \vec{v}_n$.*
3. *Let W be a vector space over a field F and $\vec{w}_1, \ldots, \vec{w}_n \in W$ (not necessarily distinct). If $\mathcal{B}$ is a basis for V, then there is a unique linear transformation from V to W such that $T(\vec{v}_i) = \vec{w}_i$ for all $i \in \{1, \ldots, n\}$.*

Examples 1.4.1.
1. Find a linear transformation T that satisfies the following conditions
   (i) $T : \mathbb{C} \to \mathbb{R}_2[x]$ with $T(1-i) = 2x^2$ and $T(1+i) = 1-x$,
   (ii) $T : \mathbb{R}_2[x] \to \mathbb{R}^2$ with $T(1) = (2,1), T(1-x) = (0,1)$ and $T(x+x^2) = (1,1)$.
2. Let $T : \mathbb{R}_1[x] \to \mathbb{R}^3$ be a linear transformation with
$$T(2-x) = (1,-1,1) \quad \text{and} \quad T(1+x) = (0,1,-1).$$
   Find $T(-1+2x)$.

Lemma 1.4.2. *1. If $\vec{u}, \vec{v}_1, \ldots, \vec{v}_n \in S$ and $\vec{u} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$, then $\operatorname{Span} S = \operatorname{Span}(S \setminus \{\vec{u}\})$.*
2. If S is a linearly independent subset of V and $\vec{u} \notin \operatorname{Span} S$, then $S \cup \{\vec{u}\}$ is linearly independent.

Theorem 1.4.3. *Let V be a vector space over F.*
1. If $\mathcal{B}$ is a linearly independent subset of V which is maximal with respect to the property of being linearly independent (i.e., $\forall \mathcal{B} \subseteq S, S \neq \mathcal{B} \Rightarrow S$ is not linearly independent), then $\mathcal{B}$ is a basis of V.
2. If $\mathcal{B}$ is a spanning set for V which is minimal with respect to the property of spanning (i.e., $\forall S \subseteq \mathcal{B}, S \neq \mathcal{B} \Rightarrow \operatorname{Span} S \subsetneqq V$), then $\mathcal{B}$ is a basis of V.

Theorem 1.4.4. [Replacement Theorem] *Let V be a vector space that is spanned by a set G containing exactly n vectors. Let L be a linearly independent subset of V with m vectors. Then*
1. $m \leq n$,
2. there exists a subset H of G with $n - m$ vectors such that $L \cup H$ spans V.

Example 1.4.2. Extend $\{(1,1,1)\}$ to a basis of $\mathbb{R}^3$.

Corollary 1.4.5. *If a vector space V has a finite spanning set $\{\vec{v}_1, \ldots, \vec{v}_n\}$, then*
1. $\{\vec{v}_1, \ldots, \vec{v}_n\}$ has a subset which is a basis,
2. any linearly independent set in V can be extended to a basis,
3. V has a basis,
4. any two bases have the same finite number of elements, necessarily $\leq n$.

Definition. If a vector space V has a finite spanning set, then we say that V is **finite-dimensional**, and the number of elements in a basis is called the **dimension** of V, written $\dim V$. If V has no finite spanning set, we say that V is **infinite-dimensional**.

Examples 1.4.3.
1. The vector space $\{\vec{0}\}$ has dimension zero with basis $\emptyset$.
2. The vector space F^n, $n \geq 1$, is of dimension n with standard basis $\{(1,0,\ldots,0), (0,1,\ldots,0), \ldots, (0,0,\ldots,1)\}$. Similarly, $M_{m,n}(F)$ is of dimension mn where $m, n \in \mathbb{N}$.
3. The vector space $F_n[x]$ is of dimension $n+1$ with standard basis $\{1, x, x^2, \ldots, x^n\}$.
4. The vector spaces $F^{\mathbb{N}}$ and $F[x]$ are infinite-dimensional. A basis for $F[x]$ is $\{1, x, x^2, \ldots\}$.
5. If we consider $\mathbb{C}$ as a vector space over $\mathbb{C}$, it has dimension one with basis $\{1\}$. But if we consider $\mathbb{C}$ as a vector space over $\mathbb{R}$ it has dimension two with basis $\{1, i\}$.

Remark. The above corollary is valid for a "finite" dimensional vector space. For a general (finite/infinite dimensional) vector space V, consider $\mathscr{L} = \{L \subseteq V : L \text{ is linearly independent}\}$. Then $\emptyset \in \mathscr{L}$. Partially ordering $\mathscr{L}$ by $\subseteq$. We now show that every chain in $\mathscr{L}$ has an upper bound. Let $\mathscr{C}$ be a chain in $\mathscr{L}$. Consider $\bigcup \mathscr{C}$. Let $\vec{v}_1, \ldots, \vec{v}_n \in \bigcup \mathscr{C}$ and $c_1, \ldots, c_n \in F$ be such that $c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}_V$. Suppose $\vec{v}_i \in L_i$ for some $L_i \in \mathscr{C}$ for all $i \in \{1, \ldots, n\}$. Since $\mathscr{C}$ is a chain, we may suppose that $L_1 \subseteq \ldots \subseteq L_n$. Thus, $\vec{v}_1, \ldots, \vec{v}_n$ are in L_n which is a linearly independent set. This implies $c_1 = \cdots = c_n = 0$. Hence, $\bigcup \mathscr{C}$ is a linearly independent set, so $\bigcup \mathscr{C}$ is in $\mathscr{L}$. By Zorn's lemma—"If a partially ordered set P has the property that every chain (i.e., totally ordered subset) has an upper bound in P, then the set P contains at least one maximal element.", $\mathscr{L}$ contains a maximal element, say $\mathcal{B}$. This is a maximal linearly independent subset of V. By Theorem 1.4.3 (1), $\mathcal{B}$ is a basis for V. *Hence, every vector space has a basis.* Note that a basis for $F^{\mathbb{N}}$ exists in this way and is not constructible explicitly.

Corollary 1.4.6. *If V is a finite-dimensional vector space with $\dim V = n$, then any spanning set of n elements is a basis of V, and any linearly independent set of n elements is a basis of V. Consequently, if W is an n-dimensional subspace of V, then $W = V$.*

Corollary 1.4.7. *If V is a finite-dimensional vector space and U is a proper subspace of V, then U is finite-dimensional and $\dim U < \dim V$.*

Theorem 1.4.8. *If W_1 and W_2 are finite dimensional subspaces of a vector space V over a field F, then $W_1 + W_2$ is finite dimensional and*

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

Example 1.4.4. Consider two subspaces of $\mathbb{R}^5$

$$W_1 = \left\{ \begin{bmatrix} a \\ a-b \\ b \\ a+b \\ 0 \end{bmatrix} \in \mathbb{R}^5 : a, b \in \mathbb{R} \right\} \text{ and } W_2 = \left\{ \begin{bmatrix} c \\ d \\ 0 \\ e \\ d-e \end{bmatrix} \in \mathbb{R}^5 : c, d, e \in \mathbb{R} \right\}.$$

Find bases for W_1, W_2 and $W_1 \cap W_2$. Determine the dimension of $W_1 + W_2$.

Definition. Let V and W be vector spaces over a field and $T : V \to W$ a linear transformation. If V is finite dimensional, the **rank of** T, denoted by $\operatorname{rank} T$, is $\dim(\operatorname{im} T)$ and the **nullity of** T, denoted by $\operatorname{nullity} T$, is $\dim(\ker T)$.

Theorem 1.4.9. *Let V and W be vector spaces over a field F and $T : V \to W$ a linear transformation. If V is finite dimensional, then*

$$\operatorname{rank} T + \operatorname{nullity} T = \dim V.$$

Theorem 1.4.10. *Let V and W be finite dimensional and $T : V \to W$ a linear transformation and $\dim V = \dim W$. Then T is one-to-one $\Leftrightarrow$ T is onto.*

Corollary 1.4.11. *If V is finite dimensional, S and T are linear transformations from V to V, and $T \circ S$ is the identity map, then $T = S^{-1}$.*

From Theorem 1.4.1, we have known that the representation of a given vector $\vec{v} \in V$ in terms of a given basis is unique.

Definition. Let V be an n-dimensional vector space over a field F with ordered basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\vec{v} \in V$. Then $\forall \vec{v} \in V, \exists!(c_1, \ldots, c_n) \in F^n, \vec{v} = c_1\vec{u}_1 + \cdots + c_n\vec{u}_n$. The vector

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in F^n$$

is called the **coordinate vector of $\vec{v}$ relative to the ordered basis $\mathcal{B}$**.

Theorem 1.4.12. *For $\vec{v}, \vec{w} \in V$ and $c \in F$, we have $[\vec{v} + \vec{w}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{B}} + [\vec{w}]_{\mathcal{B}}$ and $[c\vec{v}]_{\mathcal{B}} = c[\vec{v}]_{\mathcal{B}}$.*

Definition. A one-to-one linear transformation from V onto W is called an **isomorphism**. If there exists an isomorphism from V onto W, then we say that V is **isomorphic** to W and we write $V \cong W$.

Note that $\cong$ is an equivalence relation.

Theorem 1.4.13. *Let V be an n-dimensional vector space over F.*
If $\mathcal{B}$ is a basis for V, then the map $\vec{v} \mapsto [\vec{v}]_{\mathcal{B}}$ is an isomorphism from V onto F^n. Hence, $V \cong F^n$.

Therefore, the theory of finite-dimensional vector spaces can be studied from column vectors and matrices which we shall pursue in the next chapter.

Corollary 1.4.14. *If V and W are finite dimensional, then $\dim V = \dim W \Leftrightarrow V \cong W$.*

Exercises for Chapter 1.

1. Let $V = \mathbb{R}^+$ the set of all positive integers. Define a vector addition and a scalar multiplication on V as
$$v \oplus w = vw \quad \text{and} \quad \alpha \odot v = v^\alpha$$
for all positive real numbers v and w, and $\alpha \in \mathbb{R}$. Show that $(V, \oplus, \odot)$ is a vector space over $\mathbb{R}$.
2. Let V be a vector space over a field F. For $c \in F$ and $\vec{v} \in V$, if $c\vec{v} = \vec{v}$, prove that $c = 1$ or $\vec{v} = \vec{0}_V$.
3. Which of the following are subspaces of $M_2(\mathbb{R})$?
(a) $\{A \in M_2(\mathbb{R}) : \det A = 0\}$ (b) $\{A \in M_2(\mathbb{R}) : A = A^T\}$
(c) $\{A \in M_2(\mathbb{R}) : A = -A^T\}$ (d) $\{A \in M_2(\mathbb{R}) : A^2 = A\}$
4. Which of the following are subspaces of $\mathbb{R}^{\mathbb{N}}$?
(a) All sequences like $(1, 0, 1, 0, \dots)$ that include infinitely many zeros.
(b) $\{(a_n) \in \mathbb{R}^{\mathbb{N}} : \exists n_0 \in \mathbb{N}, \forall j \geq n_0, a_j = 0\}$. (c) All decreasing sequences: $a_{j+1} \leq a_j$ for all $j \in \mathbb{N}$.
(d) All arithmetic sequences: $\{(a_n) \in \mathbb{R}^{\mathbb{N}} : \exists a, d \in \mathbb{R}, \forall n \in \mathbb{N}, a_n = a + (n-1)d\}$.
(e) All geometric sequences: $\{(a_n) \in \mathbb{R}^{\mathbb{N}} : \exists a, r \in \mathbb{R}, \forall n \in \mathbb{N}, r \neq 0 \wedge a_n = ar^{n-1}\}$.
5. Which of the following are subspaces of $V = C^0[0, 1]$?
(a) $\{f \in V : f(0) = 0\}$ (b) $\{f \in V : \forall x \in [0, 1], f(x) \geq 0\}$
(c) All increasing functions: $\forall x, y \in [0, 1], x < y \Rightarrow f(x) \leq f(y)$.
6. Let V and W be vector spaces over a field F and $T : V \to W$ a linear transformation.
(a) If V_1 is a subspace of V, then $T(V_1) = \{T(\vec{x}) : \vec{x} \in V_1\}$ is a subspace of W.
(b) If W_1 is a subspace of W, then $T^{-1}(W_1) = \{\vec{x} \in V : T(\vec{x}) \in W_1\}$ is a subspace of V.
7. If L, M and N are three subspaces of a vector space V such that $M \subseteq L$, then show that
$$L \cap (M + N) = (L \cap M) + (L \cap N) = M + (L \cap N).$$
Also give an example, in which the result fails to hold when $M \nsubseteq L$. (*Hint*. Consider V_α of F^2.)
8. Let S_1 and S_2 be subsets of a vector space V. Prove that $\mathrm{Span}(S_1 \cup S_2) = \mathrm{Span}\ S_1 + \mathrm{Span}\ S_2$.
9. If $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in V$ such that $\vec{v}_1 + \vec{v}_2 + \vec{v}_3 = \vec{0}$, prove that $\mathrm{Span}\{\vec{v}_1, \vec{v}_2\} = \mathrm{Span}\{\vec{v}_2, \vec{v}_3\}$.
10. Let $S = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $c_1, \dots, c_n \in F \setminus \{0\}$. Prove that:
(a) $\mathrm{Span}\ S = \mathrm{Span}\{c_1\vec{v}_1, \dots, c_n\vec{v}_n\}$
(b) S is linearly independent $\Leftrightarrow \{c_1\vec{v}_1, \dots, c_n\vec{v}_n\}$ is linearly independent.
11. If $\{\vec{y}, \vec{v}_1, \dots, \vec{v}_n\}$ is linearly independent, show that $\{\vec{y} + \vec{v}_1, \dots, \vec{y} + \vec{v}_n\}$ is also linearly independent.
12. Determine (with reason or counter example) whether the following statements are TRUE or FALSE.
(a) If W_1 and W_2 are subspaces of V, then $W_1 \cup W_2$ is a subspace of V.
(b) If $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis of $\mathbb{R}^3$, then $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is a basis of $\mathbb{R}^3$.
13. Determine whether the following subsets are linearly independent.
(a) $\{(1, i, -1), (1 + i, 0, 1 - i), (i, -1, -i)\}$ in $\mathbb{C}^3$ (b) $\{x, \sin x, \cos x\}$ in $C^0(\mathbb{R})$
14. Let V be a vector space over a field F. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be vectors in V.
If $\vec{w} \in \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \setminus \mathrm{Span}\{\vec{v}_2, \dots, \vec{v}_n\}$, then $\vec{v}_1 \in \mathrm{Span}\{\vec{w}, \vec{v}_2, \dots, \vec{v}_n\} \setminus \mathrm{Span}\{\vec{v}_2, \dots, \vec{v}_n\}$.
15. Prove that if U and V are finite dimensional vector spaces, then $\dim(U \times V) = \dim U + \dim V$.
16. Find a basis and the dimension of the following subspaces of $M_2(\mathbb{R})$.
(a) $\{A \in M_2(\mathbb{R}) : A = A^T\}$ (b) $\{A \in M_2(\mathbb{R}) : A = -A^T\}$
(c) $\{A \in M_2(\mathbb{R}) : \forall B \in M_2(\mathbb{R}), AB = BA\}$
17. Let $B \in M_2(\mathbb{R})$ and $W = \{A \in M_2(\mathbb{R}) : AB = BA\}$.
Prove that W is a subspace of $M_2(\mathbb{R})$ and $\dim W \geq 2$.
18. Find a basis for the subspace $W = \{p(x) \in \mathbb{R}_3[x] : p(2) = 0\}$ and extend to a basis for $\mathbb{R}_3[x]$.

19. Let $W_1 = \text{Span}\{(1,0,2),(1,-2,2)\}$ and $W_2 = \text{Span}\{(1,1,0),(0,1,-1)\}$ in $\mathbb{R}^3$.
    Find $\dim(W_1 \cap W_2)$ and $\dim(W_1 + W_2)$.
20. If $T : V \to W$ is a linear transformation and $\mathcal{B}$ is a basis for V, prove that $\text{Span } T(\mathcal{B}) = \text{im}\, T$.
21. Let $T : \mathbb{R}_2[x] \to \mathbb{R}_3[x]$ be given by $T(p(x)) = xp(x)$.
    (a) Prove that T is a linear transformation and determine its rank and nullity.
    (b) Does T^{-1} exist? Explain.
22. Suppose that U and V are subspaces of $\mathbb{R}^{13}$, with $\dim U = 7$ and $\dim V = 8$.
    (a) What is the smallest and largest possible dimensions of $U \cap V$? Explain.
    (b) What is the smallest and largest possible dimensions of $U + V$? Explain.
23. If V and W are finite-dimensional vector spaces such that $\dim V > \dim W$, then there is no one-to-one linear transformation $T : V \to W$.
24. Let U and W be subspaces of a vector space V. If $\dim V = 3$, $\dim U = \dim W = 2$ and $U \neq W$, prove that $\dim(U \cap W) = 1$.
25. Let U and W be subspaces of a vector space V such that $U \cap W = \{\vec{0}\}$.
    Assume that $\vec{u}_1, \vec{u}_2$ are linearly independent in U and $\vec{w}_1, \vec{w}_2, \vec{w}_3$ are linearly independent in W.
    (a) Prove that $\{\vec{u}_1, \vec{u}_2, \vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is a linearly independent set in V.
    (b) If $\dim V = 5$, show that $\dim U = 2$ and $\dim V = 3$.

2 | Inner Product Spaces

2.1 Inner Products

We shall need the following properties of complex numbers.

Proposition 2.1.1. *Let $z = a + bi$ where $a, b \in \mathbb{R}$.*

1. $\operatorname{Re} z = a$ (real part) *and* $\operatorname{Im} z = b$ (imaginary part)
2. *The conjugate $\bar{z} = a - bi$, and the absolute value $|z| = \sqrt{a^2 + b^2}$. Moreover, $z\bar{z} = |z|^2$.*
3. $\bar{\bar{z}} = z$ *and* $|z| = 0 \Leftrightarrow a = b = 0$.
4. *If $z, w \in \mathbb{C}$, then $\overline{z + w} = \bar{z} + \bar{w}$ and $\overline{zw} = \bar{z}\bar{w}$.*

Definition. Let $F = \mathbb{R}$ or $\mathbb{C}$ and let V be a vector space over F. Let $\vec{u}$ and $\vec{v}$ be vectors in V. An **inner product** or **scalar product** on V is a function from $V \times V$ to F, denoted by $\langle \cdot, \cdot \rangle$, with following properties:

(IN1) $\forall \vec{u}, \vec{v}, \vec{w} \in V, \langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$.

(IN2) $\forall \vec{u}, \vec{v} \in V, \forall c \in F, \langle c\vec{u}, \vec{v} \rangle = c\langle \vec{u}, \vec{v} \rangle$.

(IN3) $\forall \vec{u}, \vec{v} \in V, \langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$. Here, $\bar{\cdot}$ is the complex conjugation.

(IN4) $\forall \vec{u} \in V, \langle \vec{u}, \vec{u} \rangle \geq 0$ and $[\langle \vec{u}, \vec{u} \rangle = 0 \Rightarrow \vec{u} = \vec{0}]$.

A vector space over F, in which an inner product is defined, is called an **inner product space**.

Remarks.

1. For all $\vec{u}, \vec{v} \in V$, $\langle \vec{0}, \vec{u} \rangle = 0 = \langle \vec{u}, \vec{0} \rangle$ and $\langle \vec{u}, \vec{v} \rangle = 0 \Leftrightarrow \langle \vec{v}, \vec{u} \rangle = 0$.
2. If $F = \mathbb{R}$, then (IN3) reads $\forall \vec{u}, \vec{v} \in V, \langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$.

Example 2.1.1. Consider the complex vector space $\mathbb{C}^n$ of n-tuples of complex numbers. Let $\vec{u} = (u_1, u_2, \ldots, u_n)$ and $\vec{v} = (v_1, v_2, \ldots, v_n)$. We define

$$\langle \vec{u}, \vec{v} \rangle = u_1\bar{v}_1 + u_2\bar{v}_2 + \ldots + u_n\bar{v}_n.$$

Show that this is an inner product.

Remark. If we consider, on the other hand, $\mathbb{R}^n$ the space of n-tuples of real numbers, we have a real-valued scalar product $\langle \vec{u}, \vec{v} \rangle = u_1v_1 + u_2v_2 + \ldots + u_nv_n$ and the verification of the properties is exactly like Example 2.1.1, where all conjugation symbols are removed.

Example 2.1.2. Consider $V = C^0[a, b]$ the vector space of real-valued continuous functions defined on the interval $[a, b]$. Let

$$\langle f, g \rangle = \int_a^b f(x)g(x)\, dx.$$

Show that this defines an inner product.

We can add to the list of properties of the scalar product by proving some theorems, assuming of course that we are dealing with a complex vector space with a scalar product.

Theorem 2.1.2. *1.* $\forall \vec{u}, \vec{v}, \vec{w} \in V, \langle \vec{u}, \vec{v} + \vec{w} \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}, \vec{w} \rangle$.
2. $\forall \vec{u}, \vec{v} \in V, \forall c \in F, \langle \vec{u}, c\vec{v} \rangle = \bar{c}\langle \vec{u}, \vec{v} \rangle$.
3. $\big(\forall \vec{u} \in V, \langle \vec{u}, \vec{v} \rangle = 0\big) \Rightarrow \vec{v} = \vec{0}$.
4. $\big(\forall \vec{u} \in V, \langle \vec{u}, \vec{v} \rangle = \langle \vec{u}, \vec{w} \rangle\big) \Rightarrow \vec{v} = \vec{w}$. *In fact, if* $\langle \vec{v} - \vec{w}, \vec{v} \rangle = \langle \vec{v} - \vec{w}, \vec{w} \rangle$, *then* $\vec{v} = \vec{w}$.

Remark. Let $c_1, c_2 \in F$ and $\vec{u}, \vec{v} \in V$. Then

$$\langle c_1\vec{u} + c_2\vec{v}, c_1\vec{u} + c_2\vec{v} \rangle = c_1\bar{c}_1\langle \vec{u}, \vec{u} \rangle + c_1\bar{c}_2\langle \vec{u}, \vec{v} \rangle + \bar{c}_1 c_2\langle \vec{v}, \vec{u} \rangle + c_2\bar{c}_2\langle \vec{v}, \vec{v} \rangle.$$

Moreover, if $\langle \vec{u}, \vec{v} \rangle = 0$, then $\langle \vec{v}, \vec{u} \rangle = 0$, so

$$\langle c_1\vec{u} + c_2\vec{v}, c_1\vec{u} + c_2\vec{v} \rangle = c_1\bar{c}_1\langle \vec{u}, \vec{u} \rangle + c_2\bar{c}_2\langle \vec{v}, \vec{v} \rangle = |c_1|^2\langle \vec{u}, \vec{u} \rangle + |c_2|^2\langle \vec{v}, \vec{v} \rangle.$$

The quantity $\langle \vec{u}, \vec{u} \rangle$ is non-negative and is zero if and only if $\vec{u} = \vec{0}$. Therefore, we associate with it the square of the length of the vector.

Definition. For $\vec{v} \in V$, we define the **length** or **norm** of $\vec{v}$ to be $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$.

Some of the properties of the norm are given by the next theorem.

Theorem 2.1.3. *If V is an inner product space over F, then the norm $\|\cdot\|$ has the following properties:*
1. $\forall \vec{u} \in V, \|\vec{u}\| \geq 0$ *and* $\|\vec{u}\| = 0 \Leftrightarrow \vec{u} = \vec{0}$
2. $\forall \vec{u} \in V, \forall a \in F, \|a\vec{u}\| = |a|\|\vec{u}\|$
3. $\forall \vec{u}, \vec{v} \in V, |(\vec{u}, \vec{v})| \leq \|\vec{u}\|\|\vec{v}\|$ *(the Cauchy-Schwarz inequality)*
4. $\forall \vec{u}, \vec{v} \in V, \|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$ *(the triangle inequality).*

Example 2.1.3. Let f be a real-valued continuous function defined on the interval $[a, b]$. Prove that

$$\left| \int_a^b f(x)dx \right| \leq (b-a)M, \text{ where } M = \max_{x \in [a,b]} |f(x)|.$$

2.2 Orthonormal Bases

Definition. Let V be an inner product space over F. Two nonzero vectors $\vec{u}$ and $\vec{v}$ are **orthogonal** if $(\vec{u}, \vec{v}) = 0$. A vector $\vec{u}$ is a **unit vector** if $\|\vec{u}\| = 1$.

Definition. A subset S of V is called an **orthogonal set** if $\forall \vec{u}, \vec{v} \in S, \vec{u} \neq \vec{v} \Rightarrow \vec{u}$ and $\vec{v}$ are orthogonal. Moreover, S is called an **orthonormal set** if S is orthogonal and $\forall \vec{v} \in S, \|\vec{v}\| = 1$.

Example 2.2.1. 1. The standard basis of $F^n, n \in \mathbb{N}$ is an orthonormal set.

2. Let $V = C^0[0, 2\pi]$ with inner product $\langle f, g \rangle = \displaystyle\int_0^{2\pi} f(x)g(x)\,dx$. Then

$$S = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \dots \right\}$$

is an orthonormal set.

Let V be an inner product space.

Lemma 2.2.1. *Let $S = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthogonal set.*

1. $\forall \alpha_1, \ldots, \alpha_n \in F, \forall k \in \{1, \ldots, n\}, \left\langle \sum_{i=1}^{n} \alpha_i \vec{v}_i, \vec{v}_k \right\rangle = \alpha_k \|\vec{v}_k\|^2.$
2. $\forall \vec{v} \in \text{Span } S, \vec{v} = \sum_{i=1}^{n} \frac{\langle \vec{v}, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i.$

Theorem 2.2.2. *If S is an orthogonal set, then S is linearly independent.*

Theorem 2.2.3. [Gram-Schmidt Process] *Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$ be linearly independent. Then $\forall m \in \{1, \ldots, n\}, \exists \vec{w}_1, \ldots, \vec{w}_m \in V$ such that $\{\vec{w}_1, \ldots, \vec{w}_m\}$ is an orthogonal set and it is a basis for* $\text{Span}\{\vec{v}_1, \ldots, \vec{v}_m\}$.

Proof. We prove this theorem by induction on $m \geq 1$.

If $m = 1$, $\{\vec{v}_1\}$ is an orthogonal set. Choose $\vec{w}_1 = \vec{v}_1$. Then $\text{Span}\{\vec{w}_1\} = \text{Span}\{\vec{v}_1\}$. Let $k \in \{1, 2, \ldots, n-1\}$ and assume that there exist $\vec{w}_1, \ldots, \vec{w}_k \in V$ such that $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is an orthogonal set $\text{Span}\{\vec{w}_1, \ldots, \vec{w}_k\} = \text{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$. Choose

$$\vec{w}_{k+1} = \vec{v}_{k+1} - \hat{v}_{k+1} = \vec{v}_{k+1} - \sum_{i=1}^{k} \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\|\vec{w}_i\|^2} \vec{w}_i. \tag{2.2.1}$$

We have to show that:

(1) $\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}$ is an orthogonal set. By induction hypothesis, $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is an orthogonal set, so it suffices to show that $\vec{w}_{k+1}$ is orthogonal to $\vec{w}_j$ for all $j \in \{1, \ldots, k\}$. Let $j \in \{1, \ldots, k\}$.

$$\begin{aligned}
\langle \vec{w}_{k+1}, \vec{w}_j \rangle &= \left\langle \vec{v}_{k+1} - \sum_{i=1}^{k} \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\|\vec{w}_i\|^2} \vec{w}_i, \vec{w}_j \right\rangle \\
&= \langle \vec{v}_{k+1}, \vec{w}_j \rangle - \sum_{i=1}^{k} \left\langle \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\|\vec{w}_i\|^2} \vec{w}_i, \vec{w}_j \right\rangle \\
&= \langle \vec{v}_{k+1}, \vec{w}_j \rangle - \langle \vec{v}_{k+1}, \vec{w}_j \rangle = 0.
\end{aligned}$$

(2) $\text{Span}\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\} = \text{Span}\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}\}$. Again, by induction hypothesis,

$$\text{Span}\{\vec{w}_1, \ldots, \vec{w}_k\} = \text{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}.$$

From Eq. (2.2.1), we have

$$\vec{w}_{k+1} \in \text{Span}\{\vec{w}_1, \ldots, \vec{w}_k, \vec{v}_{k+1}\} = \text{Span}\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}\}.$$

Then $\text{Span}\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\} \subseteq \text{Span}\{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}\}$. For the reverse, we note that

$$\vec{v}_{k+1} = \vec{w}_{k+1} + \sum_{i=1}^{k} \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\|\vec{w}_i\|^2} \vec{w}_i \in \text{Span}\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}.$$

Since an orthogonal set is linearly independent, $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is a basis for $\text{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$. □

Corollary 2.2.4. *If V is a finite dimensional inner product space, then V has an orthonormal basis.*

Proof. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_m\}$ be a basis for V. Then $\mathcal{B}$ is linearly independent. By the Gram-Schmidt Process, we can construct an orthogonal subset $\{\vec{w}_1, \ldots, \vec{w}_m\}$ of V which is a basis for $\text{Span}\{\vec{v}_1, \ldots, \vec{v}_m\} = V$. Hence, $\{\vec{w}_1, \ldots, \vec{w}_m\}$ is an orthogonal basis for V in which we can normalize each vector to obtain an orthonormal basis as desired. □

Example 2.2.2. Let $H = \text{Span}\left\{\begin{bmatrix}1\\2i\\0\end{bmatrix}, \begin{bmatrix}2i\\6\\-3\end{bmatrix}\right\} \subset \mathbb{C}^3$. Find an orthonormal basis for H.

Example 2.2.3. Let V be the space of continuous functions on $[0,1]$ and $H = \text{Span}\{1, 3\sqrt{x}, 10x\}$ a 3-dimensional subspace of V. Use the Gram-Schmidt process to find an orthogonal basis for H.

2.3 Orthogonal Complements

Definition. Let V be an inner product space over F. For $S \subseteq V$, the **orthogonal complement of** S is the set $S^\perp$, read "S perp", defined by

$$S^\perp = \{\vec{v} \in V : \langle \vec{v}, \vec{u} \rangle = 0 \text{ for all } \vec{u} \in S\}.$$

Remark. $\emptyset^\perp = V = \{\vec{0}\}^\perp$, $V^\perp = \{\vec{0}\}$ and $S^\perp = (\text{Span } S)^\perp$.

Theorem 2.3.1. *For any subset S of V, $S^\perp$ is a subspace of V.*

Lemma 2.3.2. *Let $S = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a set of distinct nonzero vectors.*
If S is an orthogonal set, then $\vec{v} - \sum_{i=1}^{n} \frac{\langle \vec{v}, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i \in S^\perp$ for all $\vec{v} \in V$.

Theorem 2.3.3. [Bessel's inequality] *Let $S = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a set of distinct nonzero vectors. If S is an orthogonal set, then for all $\vec{v} \in V$,*

$$\sum_{i=1}^{n} \frac{|\langle \vec{v}, \vec{v}_i \rangle|^2}{\|\vec{v}_i\|^2} \leq \|\vec{v}\|^2$$

and equality holds if and only if $\vec{v} \in \text{Span } S$.

Let W_1 and W_2 be subspaces of a vector space V. We know that $W_1 + W_2$ is subspace of V. If $V = W_1 + W_2$, we say that V is a **sum** of W_1 and W_2. The sum is **direct**, denoted by $W_1 \oplus W_2$, if $W_1 \cap W_2 = \{\vec{0}_V\}$. That is,

$$V = W_1 \oplus W_2 \Leftrightarrow [(1)\ V = W_1 + W_2 \quad \text{and} \quad (2)\ W_1 \cap W_2 = \{\vec{0}_V\}].$$

Theorem 2.3.4. $V = W_1 \oplus W_2$
$\Leftrightarrow$ *every vector $\vec{v} \in V$ can be expressed uniquely as $\vec{v} = \vec{w}_1 + \vec{w}_2$ with $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$.*

Theorem 2.3.5. [Orthogonal Decomposition Theorem] *Let W be a finite dimensional subspace of an inner product space V. Then*

1. *$V = W \oplus W^\perp$. In other words, every $\vec{v}$ in V decomposes uniquely as $\vec{v} = \vec{y} + \vec{z}$ with $\vec{y} \in W$ and $\vec{z} \in W^\perp$.*
2. $\dim W + \dim W^\perp = \dim V$.

Exercises for Chapter 2. 1. Let $V_n = \{A \in M_n(\mathbb{R}) : A = A^T\}$ be the vector space of all $n \times n$ *symmetric matrices* over $\mathbb{R}$, and define the product of two matrices A and B by

$$\langle A, B\rangle = \operatorname{tr}(AB).$$

where tr denotes the trace of matrix.
(a) Show that this is an inner product on V_n.
(b) Obtain an orthonormal basis for the subspace $H = \operatorname{Span}\left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix}\right\}$ of V_2.

2. Find an orthonormal basis for $\mathbb{R}_2[x]$ with respect to the inner product

$$\langle p(x), q(x)\rangle = \int_0^1 p(x)q(x)\,dx.$$

3. Let $W = \{y(x) \in \mathbb{R}^{\mathbb{R}} : y'' + 4y = 0\}$. Then W is a real vector space generated by $\{\cos 2x, \sin 2x\}$. Define an inner product $\langle y, z\rangle = \int_0^\pi y(x)z(x)\,dx$ for all $y, z \in W$. Find an orthonormal basis for W.
4. Let V and W be two vector spaces and T a one-to-one linear transformation from V into W. If W is an inner product space with inner product $(\cdot,\cdot)$, prove that the function $\langle\ ,\ \rangle : V \times V \to F$ defined by

$$\langle \vec{u}, \vec{v}\rangle = (T(\vec{u}), T(\vec{v}))$$

for all $\vec{u}, \vec{v} \in V$ is an inner product on V.
5. Let V be an inner product space over F. Prove the following statements.
(a) If $F = \mathbb{R}$, then $\forall \vec{u}, \vec{v} \in V, \langle \vec{u}, \vec{v}\rangle = \frac{1}{4}\|\vec{u}+\vec{v}\|^2 - \frac{1}{4}\|\vec{u}-\vec{v}\|^2$.
(b) If $F = \mathbb{C}$, then $\forall \vec{u}, \vec{v} \in V, \langle \vec{u}, \vec{v}\rangle = \frac{1}{4}\|\vec{u}+\vec{v}\|^2 - \frac{1}{4}\|\vec{u}-\vec{v}\|^2 + \frac{i}{4}\|\vec{u}+i\vec{v}\|^2 - \frac{i}{4}\|\vec{u}-i\vec{v}\|^2$.
(c) $\forall \vec{u}, \vec{v} \in V, \|\vec{u}+\vec{v}\|^2 + \|\vec{u}-\vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2$.
(a) and (b) are called the **polarization identity** and (c) is called the **parallelogram law**.
6. Show that $|\|\vec{u}\| - \|\vec{v}\|| \le \|\vec{u}-\vec{v}\|$ for all $\vec{u}, \vec{v} \in V$.
7. From the Cauchy-Schwarz inequality, $|\langle \vec{u}, \vec{v}\rangle| \le \|\vec{u}\|\|\vec{v}\|$, prove that equality holds if and only if $\vec{u}$ and $\vec{v}$ are linearly dependent.
8. By choosing a suitable vector $\vec{b}$ in the Cauchy-Schwarz inequality, prove that

$$(a_1 + \cdots + a_n)^2 \le n(a_1^2 + \cdots + a_n^2).$$

When does equality hold?
9. Consider $V = C^0[a, b]$. Let $f \in V$. Prove that

$$\int_a^b |f(x)|^2\,dx \le \left(\int_a^b |f(x)|\,dx\right)^{1/2}\left(\int_a^b |f(x)|^3\,dx\right)^{1/2}.$$

10. Prove that the finite sequence $a_0, a_1, \ldots, a_n$ of positive real numbers is a geometric progression if and only if

$$(a_0a_1 + a_1a_2 + \cdots + a_{n-1}a_n)^2 = (a_0^2 + a_1^2 + \cdots + a_{n-1}^2)(a_1^2 + a_2^2 + \cdots + a_n^2).$$

11. Let $P(x)$ be a polynomial with positive real coefficients. Prove that

$$\sqrt{P(a)P(b)} \ge P(\sqrt{ab})$$

for all $a, b \ge 0$.
12. Let V be an n-dimensional inner product space and $m < n$. If $\{\vec{v}_1, \ldots, \vec{v}_m\}$ is an orthonormal set, then there exists $\vec{v}_{m+1}, \ldots, \vec{v}_n \in V$ such that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal basis for V.
13. Prove the following statements.
(a) $\forall S_1, S_2 \subseteq V, S_1 \subseteq S_2 \Rightarrow S_1^\perp \supseteq S_2^\perp$. (b) $\forall S \subseteq V, (\operatorname{Span} S)^\perp = S^\perp$.
(c) For $S \subseteq V$, if $\vec{u} \in S$ and $\vec{v} \in S^\perp$, then $\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2$.
(d) For $\vec{v}_1, \ldots, \vec{v}_n \in V$, $\{\vec{v}_1\}^\perp \cap \cdots \cap \{\vec{v}_n\}^\perp = (\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_n\})^\perp$.
14. Construct an orthonormal basis for the subspace $H = \{(1, -i, i)\}^\perp$ of $\mathbb{C}^3$.
15. Let W be a subspace of an inner product space V over F. If $\vec{v} \in V$ satisfies

$$\langle \vec{v}, \vec{w}\rangle + \langle \vec{w}, \vec{v}\rangle \le \langle \vec{w}, \vec{w}\rangle \quad \text{for all } \vec{w} \in W,$$

show that $\langle \vec{v}, \vec{w}\rangle = 0$ for all $\vec{w} \in W$.

16. Consider the inner product space $C^0[-1,1]$. Suppose that f and g are continuous on $[-1,1]$ and $\|f-g\| \le 5$. Let
$$u_1(x) = \frac{1}{\sqrt{2}} \quad \text{and} \quad u_2(x) = \sqrt{\frac{3}{2}}x \quad \text{for } x \in [-1,1].$$
Write
$$a_j = \int_{-1}^{1} u_j(x)f(x)\,dx \quad \text{and} \quad b_j = \int_{-1}^{1} u_j(x)g(x)\,dx$$
for $j = 1,2$. Show that $|a_1 - b_1|^2 + |a_2 - b_2|^2 \le 25$. (*Hint*. Use Bessel's inequality.)
17. If V is a finite dimensional inner product space and W is a subspace of V, prove that $(W^\perp)^\perp = W$.
18. If $\{\vec{v}_1, \vec{v}_2\}$ is a basis for V, show that $V = \operatorname{Span}\{\vec{v}_1\} \oplus \operatorname{Span}\{\vec{v}_2\}$.
19. Consider the subspace V_α, $\alpha \in \mathbb{R}$, of $\mathbb{R}^2$. Prove that if $\alpha \ne \beta$, then $\mathbb{R}^2 = V_\alpha \oplus V_\beta$.
20. Let $V = \mathbb{R}^{\mathbb{R}}$ be the space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Let
$$V_e = \{f \in V : \forall x \in \mathbb{R}, f(-x) = f(x)\} \quad \text{and} \quad V_o = \{f \in V : \forall x \in \mathbb{R}, f(-x) = -f(x)\},$$
the sets of all **even** and **odd** functions, respectively. Prove the following statements.
(a) V_e and V_o are subspaces of V. (b) $V = V_e \oplus V_o$.
21. Let S be a set of vectors in a finite dimensional inner product space V. Suppose that "$\langle \vec{u}, \vec{v} \rangle = 0$ for all $\vec{u} \in S$ implies $\vec{v} = \vec{0}$". Show that $V = \operatorname{Span} S$.
22. Let $\mathbb{R}^{\mathbb{N}}$ be the sequence space of real numbers. Let $V = \{(a_n) \in \mathbb{R}^{\mathbb{N}} : \text{only finitely many } a_i \ne 0\}$.
(a) Prove that V is a subspace of $\mathbb{R}^{\mathbb{N}}$.
(b) Given $(a_n), (b_n) \in V$, define
$$\langle (a_n), (b_n) \rangle = \sum_{n=1}^{\infty} a_n b_n.$$
(Note that this makes sense since only finitely many a_i and b_i are nonzero.) Show that this defines an inner product on V.
(c) Let $U = \left\{ (a_n) \in V : \sum_{n=1}^{\infty} a_n = 0 \right\}$.
Show that U is a subspace of V such that $U^\perp = \{\vec{0}\}$, $U + U^\perp \ne V$ and $U \ne U^{\perp\perp}$.

3 | Matrices

3.1 Solutions of Linear Systems

Definition. For any system of m linear equations in n unknowns with coefficients over a field F

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & \vdots & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m, \end{array}$$

we can use the matrix notation

$$A\vec{x} = \vec{b},$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

considered as matrices over F. In this case, we usually call A the **coefficient matrix** of the system. It is clear that $A\vec{x} = \vec{b}$ has a solution $\Leftrightarrow \vec{b} \in \operatorname{Col} A$. If all $b_1, \dots, b_m$ are equal to 0, the linear system is said to be **homogeneous**. Note that all solutions of a homogeneous system form the null space of A.

There is another matrix which plays an important role in the study of linear systems. This is the **augmented matrix**, which is formed by inserting $\vec{b}$ as a new last column into the coefficient matrix. In other words, the augmented matrix is

$$\left[A : \vec{b}\right] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right].$$

Remark. A homogeneous linear system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ & \vdots & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0, \end{array}$$

always has a **trivial solution**, namely the solution obtained by letting all $x_j = 0$. Other nonzero solutions (if any) are called **nontrivial solutions**.

Definition. The **rank of a matrix** A is the dimension of the column space of A.

Remark. If A is an $m \times n$ matrix, then $\operatorname{rank} A \leq n$ and $\operatorname{rank} A$ is the *maximum number* of linearly independent columns of A by Corollary 1.4.5.

Theorem 3.1.1. *Let A be an $m \times n$ matrix over a field F.*
1. *The homogeneous system $A\vec{x} = \vec{0}_m$ has only the trivial solution $\vec{x} = \vec{0}_n$ $\Leftrightarrow$ the columns of A are linearly independent $\Leftrightarrow \operatorname{rank} A = n$.*
2. *If $\operatorname{rank} A < n$, then a homogeneous linear system has a nontrivial solution in F.*

For an $m \times n$ matrix A over a field F, recall that the matrix transformation

$$T : \vec{x} \mapsto A\vec{x}$$

is a linear transformation from F^n to F^m. Its kernel is $\operatorname{Nul} A$ and its image is $\operatorname{Col} A$.

Definition. The dimension of $\operatorname{Nul} A$ is called the **nullity of** A, denoted by $\operatorname{nullity} A$.

By Theorem 1.4.9, we have:

Corollary 3.1.2. *Let A be an $m \times n$ matrix over a field F. Then*

$$\operatorname{rank} A + \operatorname{nullity} A = n = \text{ the number of columns of } A.$$

Examples 3.1.1. Consider the following augmented matrices. Write down their general solutions (if any).

1. $\left[\begin{array}{ccc|c} 1 & -3 & 4 & 7 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 5 \end{array}\right]$
2. $\left[\begin{array}{ccc|c} 1 & -3 & 7 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $\left[\begin{array}{ccc|c} 1 & -3 & 7 & 1 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right]$
3. $\left[\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 1 & -2 & 0 \end{array}\right]$ $\left[\begin{array}{ccc|c} 1 & 0 & 3 & 5 \\ 0 & 1 & -2 & 1 \end{array}\right]$
4. $\left[\begin{array}{ccccc|c} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

Theorem 3.1.3. *Let A be an $m \times n$ matrix over a field F and $\vec{b} \in F^m$.*
1. *$A\vec{x} = \vec{b}$ has a solution $\Leftrightarrow \vec{b} \in \operatorname{Col} A \Leftrightarrow \operatorname{rank}[A : \vec{b}] = \operatorname{rank} A$.*
2. *If $\vec{z} \in F^n$ is a solution of $A\vec{x} = \vec{b}$, then*

$$\vec{z} = \vec{y} + \vec{y}_p,$$

where $\vec{y}$ is a solution of the homogeneous system $A\vec{x} = \vec{0}_m$ and $A\vec{y}_p = \vec{b}$.

Hence, the solution set of $A\vec{x} = \vec{b}$ is empty or given by

$$\vec{y}_p + \{\vec{y} \in F^n : A\vec{y} = \vec{0}_m\},$$

where $\vec{y}_p$ is a solution of $A\vec{x} = \vec{b}$, called a **particular solution**.

Corollary 3.1.4. *Let A be an $m \times n$ matrix over a field F and $\vec{b} \in F^m$.*
1. *If $A\vec{x} = \vec{b}$ has a unique solution, then $A\vec{x} = \vec{0}_m$ has a unique solution and $\operatorname{rank} A = n$.*
2. *If $A\vec{x} = \vec{0}_m$ has a nontrivial solution, then $A\vec{x} = \vec{b}$ has no solution or more than one solutions.*

3.2 Inverse of a Matrix and Elementary Matrices

Definition. The main part of the algorithms used for solving simultaneous linear systems with coefficients in F is called **elementary row operations**. It makes repeatedly used of three operations on the linear system or on its augmented matrix, each of which preserves the set of solutions because its inverse is an operation of the same kind:

1. (Interchange, R_{ij}) Interchange the ith row and the jth row.
2. (Scaling, cR_i) Multiply the ith row by a *nonzero* scalar c.
3. (Replacement, $R_i + cR_j$) Replace the ith row by the sum of it and a scalar c multiple of the jth row.

The **elementary column operations** are defined in a similar way.

Remark. The elementary row operations are reversible as follows.

Operation	Reverse
R_{ij}	R_{ij}
cR_i, $c \neq 0$	$(1/c)R_i$
$R_i + cR_j$	$R_i - cR_j$

Definition. Two linear systems are said to be **equivalent** if they have the same set of solutions.

Theorem 3.2.1. *Suppose that a sequence of elementary operations is performed on a linear system. Then the resulting system has the same set of solutions as the original, so the two linear systems are equivalent.*

Proof. It is clear from the way we do the row reductions that if $c_1, c_2, \ldots, c_n$ satisfy the original system, then they also satisfy the reduced system. Since the elementary row operations are reversible if we start with the reduced system, the original system can be recovered. Now, it is clear that any solutions of the reduced system is also a solution of the original system. □

Definition. A rectangular matrix is in **echelon form** (or **row-echelon form**) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zero. If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form** (or **reduced row-echelon form**):
4. The leading entry in each nonzero row is 1, called the **leading** 1.
5. Each leading 1 is the only nonzero entry in its column.

An **echelon matrix** (respectively, **reduced echelon matrix**) is one that is in echelon form (respectively, reduced echelon form).

Theorem 3.2.2. *Every matrix can be brought to a reduced echelon matrix by a finite sequence of elementary row operations.*

Proof. This can be done by an algorithm, called the **Gaussian Algorithm**.

1. If the matrix consists entirely of zeros, stop–it is already in row-echelon form.
2. Otherwise, find the first column from the left containing a nonzero entry (call it a), and move the row containing that entry to the top position.

3. Now multiply the new top row by $1/a$ to create a leading 1.
4. By subtracting multiples of that row from rows below it, make each entry below the leading 1 zero.

This completes the first row, and all further row operations are carried out on the remaining rows.

5. Repeat steps 1–4 on the matrix consisting of the remaining rows.

The process stops when either now rows remain at Step 5 of the remaining rows consist entirely of zeros. Observe that the Gaussian algorithm is recursive. □

Definition. A **pivot position** in a matrix A is a location in A that corresponds to a leading entry in an echelon form of A. A **pivot column** is a column of A that contains a pivot position.

Definition. Let A be an $n \times n$ matrix. We say that A is **invertible** or nonsingular and has the $n \times n$ matrix B as **inverse** if $AB = BA = I_n$.

If B and C are $n \times n$ matrices with $AB = I_n$ and $CA = I_n$, then the associativity of multiplication implies that

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

Hence an inverse for A is *unique* if it exists and we write A^{-1} for this inverse.

Theorem 3.2.3. *Suppose A and B are invertible matrices of the same size. Then the following results hold:*

(a) A^{-1} is invertible and $(A^{-1})^{-1} = A$, i.e., A is the inverse of A^{-1}.
(b) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
(c) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Theorem 3.2.4. [Invertible Matrix Theorem]
The following statements are equivalent for an $n \times n$ matrix A.

(i) A is invertible.
(ii) The homogeneous system $A\vec{x} = \vec{0}$ has only the trivial solution $\vec{x} = \vec{0}_n$.
(iii) A can be carried to the identity matrix I_n by elementary row operations.
(iv) The system $A\vec{x} = \vec{b}$ has at least one solution for any vector $\vec{b} \in F^n$.
(v) There is an $n \times n$ matrix C such that $AC = I_n$.

Corollary 3.2.5. *If A and C are square matrices such that $AC = I$, then also $CA = I$. In particular, A and C are invertible, $C = A^{-1}$ and $A = C^{-1}$.*

Corollary 3.2.6. *An $n \times n$ matrix A is invertible if and only if* $\operatorname{rank} A = n$.

Definition. An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

Example 3.2.1. Consider the following elementary matrices:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \text{and} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Compute the products E_1A, E_2A and E_3A.

Theorem 3.2.7. *If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.*

Remark. Elementary matrices are invertible because row operations are reversible. To find the inverse of an elementary matrix E, determine the elementary row operation needed to transform E back into I and apply this operation to I to obtain the inverse.

Corollary 3.2.8. *An elementary matrix is invertible. Moreover,*

1. *If $I \xrightarrow{R_{ij}} E_1$, then $I \xrightarrow{R_{ij}} E_1^{-1}$.*
2. *If $c \neq 0$ and $I \xrightarrow{cR_i} E_2$, then $I \xrightarrow{(1/c)R_i} E_2^{-1}$.*
3. *If $c \in F$ and $I \xrightarrow{R_i + cR_j} E_3$, then $I \xrightarrow{R_i - cR_j} E_3^{-1}$.*

Example 3.2.2. Find the inverses of the elementary matrices given in Example 3.2.1

Theorem 3.2.9. *Suppose A is an $m \times n$ matrix and $A \to B$ by elementary row operations.*

1. *$B = UA$ for some $m \times m$ invertible matrix U.*
2. *U can be computed by $[A : I_m] \to [B : U]$ using the operations carrying $A \to B$.*
3. *$U = E_k E_{k-1} \dots E_2 E_1$, where $E_1, E_2, \dots, E_{k-1}, E_k$ are the elementary matrix corresponding (in order) to the elementary row operations carrying $A \to B$.*

Example 3.2.3. If $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$, express the reduced row-echelon form R of A as $R = UA$ where U is invertible.

Theorem 3.2.10. *A square matrix is invertible if and only if it is a product of elementary matrices.*

Remark. From the above theorem, we obtain an algorithm to find A^{-1} if A is invertible. Namely, we start with the block matrix $[A : I]$ and row reduce it until we reach the final reduced echelon form $[I : U]$ (because A is row equivalent to I by Theorem 3.2.4). Then we have $U = A^{-1}$.

Example 3.2.4. Express $A = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix}$ as a product of elementary matrices.

3.3 More on Ranks

Definition. Let A be an $m \times n$ matrix.
The **column space**, $\operatorname{Col} A$, of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.
The **row space**, $\operatorname{Row} A$, of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

Note that $\operatorname{Col} A = \operatorname{Row} A^T$.

Lemma 3.3.1. *Let V be a vector space over a field F. Let $\vec{v}_1, \dots, \vec{v}_n$ be in V.*

1. *$\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_n\} = \operatorname{Span}\{\vec{v}_1, \dots, c\vec{v}_i, \dots \vec{v}_n\}$ for all $i \in \{1, \dots, n\}$ and $c \in F$ nonzero.*
2. *$\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_n\} = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n\}$ for all $i \neq j$ and $c \in F$.*

Lemma 3.3.2. *Let A and B denote $m \times n$ matrices.*

1. *If $A \to B$ by elementary row operations, then $\operatorname{Row} A = \operatorname{Row} B$.*
2. *If $A \to B$ by elementary column operations, then $\operatorname{Col} A = \operatorname{Col} B$.*

If A is any matrix, we can carry $A \to R$ by elementary row operations where R is a row-echelon matrix. Hence, $\operatorname{Row} A = \operatorname{Row} R$ by Lemma 3.3.2.

Lemma 3.3.3. *If R is a row-echelon matrix, then*
1. *The nonzero rows of R form a basis for $\operatorname{Row} R$.*
2. *The columns of R containing leading ones form a basis for $\operatorname{Col} R$.*

Theorem 3.3.4. *Let A denote any $m \times n$ matrix of rank r. Then*

$$\dim \operatorname{Col} A = r = \dim \operatorname{Row} A.$$

Moreover, if A is carried to a row-echelon matrix R by row operations, then
1. *The r nonzero rows of R form a basis for $\operatorname{Row} A$.*
2. *If the pivot positions lie in columns $j_1, j_2, \dots, j_r$ of R, then columns $j_1, j_2, \dots, j_r$ of A are a basis of $\operatorname{Col} A$. That is, the pivot columns of A form a basis for $\operatorname{Col} A$.*

Corollary 3.3.5.
1. *If A is any matrix, then $\operatorname{rank} A = \operatorname{rank} A^T$.*
2. *If A is an $m \times n$ matrix, then $\operatorname{rank} A \leq m$ and $\operatorname{rank} A \leq n$.*
3. *$\operatorname{rank} A = \operatorname{rank} UA = \operatorname{rank} AV$ whenever U and V are invertible.*

Corollary 3.3.6. *Let A, B, U and V be matrices of sizes for which the indicated products are defined.*
1. *$\operatorname{Col}(AV) \subseteq \operatorname{Col} A$, with equality if V is (square and) invertible.*
2. *$\operatorname{Row}(UA) \subseteq \operatorname{Row} A$, with equality if U is (square and) invertible.*
3. *$\operatorname{rank} AB \leq \operatorname{rank} A$ and $\operatorname{rank} AB \leq \operatorname{rank} B$.*

Let A be an $m \times n$ matrix of rank r, and let R be the reduced row-echelon form of A. Theorem 3.2.9 shows that $R = UA$ where U is invertible, and that U can be found by $[A : I_m] \to [R : U]$.

The matrix R has r leading ones (since rank A=r) so, as R is reduced, the $n \times m$ matrix R^T contains each row of I_r in the first r columns. Thus, row operations will carry $R^T \to \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m}$. Hence, Theorem 3.2.9 (again) shows that $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m} = U_1 R^T$ where U_1 is an $n \times n$ invertible matrix. Writing $V = U_1^T$, we obtain

$$UAV = RV = RU_1^T = (U_1 R^T)^T = \left(\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m} \right)^T = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}.$$

Moreover, the matrix $U_1 = V^T$ can be computed by $[R^T : I_n] \to \left[\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m} : V^T \right]$. This proves

Theorem 3.3.7. *Let A be an $m \times n$ matrix of rank r. There exist invertible matrices U and V of size $m \times m$ and $n \times n$, respectively, such that*

$$UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n},$$

called the **Smith normal form of** A.
Moreover, if R is a reduced row-echelon form of A, then:
1. *U can be computed by $[A : I_m] \to [R : U]$.*
2. *V can be computed by $[R^T : I_n] \to \left[\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m} : V^T \right]$.*

Example 3.3.1. Given $A = \begin{bmatrix} 1 & -1 & 1 & 2 \\ 2 & -2 & 1 & -1 \\ -1 & 1 & 0 & 3 \end{bmatrix}$, find invertible matrices U and V such that UAV is in the Smith normal form.

Theorem 3.3.8. [Uniqueness of the reduced row-echelon form]
If a matrix A is carried to reduced row-echelon matrices R and S by row operations, then $R = S$.

Proof. Observe first that $UR = S$ for some invertible matrix U (by Theorem 3.2.9 there exist invertible matrices P and Q such that $R = PA$ and $S = QA$; take $U = QP^{-1}$. We show that $R = S$ by induction on the number m of rows of A. The case $m = 1$ is trivial because we can perform only scaling. If $\vec{r}_j$ and $\vec{s}_j$ denotes the jth column of R and of S, respectively, the fact that $UR = S$ gives

$$U\vec{r}_j = \vec{s}_j \quad \text{for each } j. \tag{3.3.1}$$

Since U is invertible, this shows that R and S have the same zero columns. Hence, by passing to the matrices obtained by deleting the zero columns from R and S, we may assume that R and S have no zero columns.

But then the first column of R and S is the first column of I_m because they are reduced row-echelon so (3.3.1) forces that the first column of U is the first column of I_m. Now, write U, R and S in block form as follows.

$$U = \begin{bmatrix} 1 & X \\ 0 & V \end{bmatrix}, R = \begin{bmatrix} 1 & Y \\ 0 & R' \end{bmatrix} \text{ and } S = \begin{bmatrix} 1 & X \\ 0 & S' \end{bmatrix}.$$

Since $UR = S$, block multiplication gives $VR' = S'$ so, since V is invertible (U is invertible) and both R' and S' are reduced row-echelon, we obtain $R' = S'$ by the induction hypothesis. Thus, R and S have the same number (say r) of leading 1's, and so both have $m - r$ zero rows.

In fact, R and S have leading ones in the same columns, say r of them. Applying (3.3.1) to these columns shows that the first r columns of U are the first r columns of I_m. Hence, we can write U, R and S in block form as follows:

$$U = \begin{bmatrix} I_r & M \\ 0 & W \end{bmatrix}, R = \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix} \text{ and } S = \begin{bmatrix} S_1 & S_2 \\ 0 & 0 \end{bmatrix},$$

where R_1 and S_1 are $r \times r$. Then block multiplication gives $UR = R$. That is, $S = R$. This completes the proof. □

3.4 Permutations and Determinants

Definition. Let $n \in \mathbb{N}$. A **permutation** σ on the set $\{1, 2, \ldots, n\}$ is a one-to-one mapping of the set onto itself or equivalently, a rearrangement of the numbers $1, 2, \ldots, n$. Such a permutation σ is denoted by

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ j_1 & j_2 & \ldots & j_n \end{pmatrix} \quad \text{or} \quad \sigma = j_1 j_2 \ldots j_n, \quad \text{where} \quad j_i = \sigma(i).$$

The set of all such permutations is denoted by S_n, and the number of such permutations is $n!$.

Example 3.4.1. $S_2 = \{12, 21\}$ and $S_3 = \{123, 132, 213, 231, 312, 321\}$.

Remark. If $\sigma \in S_n$, then the inverse mapping $\sigma^{-1} \in S_n$; and if $\sigma, \tau \in S_n$, then the composition mapping $\sigma \circ \tau \in S_n$. Also, the identity mapping $\varepsilon = \sigma \circ \sigma^{-1} = 123 \ldots n \in S_n$.

Definition. For a permutation σ in S_n, let

$$I_\sigma = \{(i,k) : i,k \in \{1,2,\ldots,n\}, i < k \text{ and } \sigma(i) > \sigma(k)\}.$$

We say that σ is an **even permutation** $\Leftrightarrow$ $|I_\sigma|$ is even, and an **odd permutation** $\Leftrightarrow$ $|I_\sigma|$ is odd. We then define the **sign** or **parity** of σ, written $\operatorname{sgn}\sigma$, by

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

Thus, $\operatorname{sgn}\sigma \in \{-1,1\}$ for all $\sigma \in S_n$.

Example 3.4.2. Let $\sigma = 2134$ in S_4 and $\tau = 21543$ in S_5.

1. Find σ^{-1} and τ^{-1}.
2. Compute $\operatorname{sgn}\sigma$ and $\operatorname{sgn}\tau$.

Theorem 3.4.1. *Let $n \geq 2$ and let g be the polynomial given by*

$$g = g(x_1, x_2, \ldots, x_n) = \prod_{i<j}(x_i - x_j).$$

For $\sigma(g) \in S_n$, define the polynomial

$$\sigma(g) = \prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}).$$

Then

$$\sigma(g) = \begin{cases} g & \text{if } \sigma \text{ is even,} \\ -g & \text{if } \sigma \text{ is odd.} \end{cases}$$

That is, $\sigma(g) = (\operatorname{sgn}\sigma)g$.

Theorem 3.4.2. *Let $\sigma, \tau \in S_n$. Then*

$$\operatorname{sgn}(\tau \circ \sigma) = (\operatorname{sgn}\tau)(\operatorname{sgn}\sigma).$$

Thus, the product of two even or two odd permutations is even, and the product of an odd and an even permutation is odd.

Let $[a_{ij}]$ be a square matrix of size $n \times n$.

Consider a product of n elements of A such that one and only one element comes from each row and one and only one element comes from each columns. Such a product can be written in the form

$$a_{1j_i} a_{2j_2} \ldots a_{nj_n}$$

that is, where the factors comes from successive rows, and so the first subscripts are in the natural order $1, 2, \ldots, n$. Now since the factors come from different columns, the sequence of the second subscripts forms a permutation $\sigma = j_1 j_2 \ldots j_n$ in S_n. Conversely, each permutation in S_n determines a product of the above form. Thus the matrix A contains $n!$ such products.

Definition. The **determinant of** $A = [a_{ij}]$, denoted by $\det A$ or $|A|$, is the sum of all the above $n!$ products where each such product is multiplied by $\operatorname{sgn}\sigma$. That is,

$$|A| = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1j_1} a_{2j_2} \ldots a_{nj_n} = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}.$$

Lemma 3.4.3. *Let $A = [a_{ij}]$ be an $n \times n$ matrix and $\sigma \in S_n$.*
1. $\operatorname{sgn} \sigma^{-1} = \operatorname{sgn} \sigma$.
2. $\{(i, \sigma^{-1}(i)) : i \in \{1, 2, \ldots, n\}\} = \{(\sigma(i), i) : i \in \{1, 2, \ldots, n\}\}$.
3. $a_{\sigma(1),1} a_{\sigma(2),2} \ldots a_{\sigma(n),n} = a_{1,\sigma^{-1}(1)} a_{2,\sigma^{-1}(2)} \ldots a_{n,\sigma^{-1}(n)}$.

Theorem 3.4.4. *The determinant of a matrix A and its transpose are equal. That is, $|A| = |A^T|$.*

Remark. By this theorem, any theorem about the determinant of a matrix A that concerns the rows of A will have an analogous theorem concerning the columns of A.

Lemma 3.4.5. *For $k < l$, $\tau = \begin{pmatrix} 1 & 2 & \ldots & k & \ldots & l & \ldots & n \\ 1 & 2 & \ldots & l & \ldots & k & \ldots & n \end{pmatrix}$ is an odd permutation in S_n.*

Proof. Note that $I_\tau = \{(k, j) : j \in \{k+1, k+2, \ldots, l\}\} \cup \{(i, l) : i \in \{k+1, k+2, \ldots, l-1\}\}$. Then $|I_\tau| = (l-k) + (l-k-1) = 2(l-k) - 1$ is odd. Thus, τ is an odd permutation and so $\operatorname{sgn} \tau = -1$. □

Theorem 3.4.6. *If $A \to B$ by interchanging two rows (columns) of A, then $|B| = -|A|$.*

Theorem 3.4.7. *Let A be a square matrix of size $n \times n$.*
(a) *If A has a row (column) of zeros, then $|A| = 0$.*
(b) *If $\sigma \neq 1\,2 \ldots n$, then $\exists i \in \{1, 2, \ldots, n\}, i > \sigma(i)$.*
(c) *If A is* **triangular**, *i.e., A has zeros above or below the diagonal, then $|A|$ is the product of diagonal elements. In particular, $|I| = 1$.*

Theorem 3.4.8. *If A has two identical rows (columns), then $|A| = 0$.*

Proof. Assume that kth and lth rows are identical with $k < l$.
That is, $a_{kj} = a_{lj}$ for all $j \in \{1, \ldots, n\}$.
In particular, for any $\sigma \in S_n$, $a_{k\sigma(l)} = a_{l\sigma(l)}$ and $a_{k\sigma(k)} = a_{l\sigma(k)}$.
Let $\tau = \begin{pmatrix} 1 & 2 & \ldots & k & \ldots & l & \ldots & n \\ 1 & 2 & \ldots & l & \ldots & k & \ldots & n \end{pmatrix}$.
Then $\operatorname{sgn} \tau = -1$ and $\sigma(\tau(j)) = \sigma(j)$ for all $j \in \{1, \ldots, n\} \setminus \{k, l\}$. Also,

$$\operatorname{sgn}(\sigma\tau) = (\operatorname{sgn} \sigma)(\operatorname{sgn} \tau) = -\operatorname{sgn} \sigma.$$

As σ runs through all even permutations, $\sigma\tau$ runs through all odd permutations, and vice versa. Thus

$$\begin{aligned}
|A| &= \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{k\sigma(k)} \ldots a_{l\sigma(l)} \ldots a_{n\sigma(n)} \\
&= \sum_{\substack{\sigma \in S_n \\ \text{even}}} \big((\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{k\sigma(k)} \ldots a_{l\sigma(l)} \ldots a_{n\sigma(n)} \\
&\qquad + (\operatorname{sgn}(\sigma\tau)) a_{1\sigma\tau(1)} a_{2\sigma\tau(2)} \ldots a_{k\sigma\tau(k)} \ldots a_{l\sigma\tau(l)} \ldots a_{n\sigma\tau(n)} \big) \\
&= \sum_{\substack{\sigma \in S_n \\ \text{even}}} \big((\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{k\sigma(k)} \ldots a_{l\sigma(l)} \ldots a_{n\sigma(n)} \\
&\qquad - (\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{k\sigma(l)} \ldots a_{l\sigma(k)} \ldots a_{n\sigma(n)} \big) \\
&= \sum_{\substack{\sigma \in S_n \\ \text{even}}} \big((\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{k\sigma(k)} \ldots a_{l\sigma(l)} \ldots a_{n\sigma(n)} \\
&\qquad - (\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{l\sigma(l)} \ldots a_{k\sigma(k)} \ldots a_{n\sigma(n)} \big) \\
&= 0.
\end{aligned}$$

Hence, we have the theorem. □

Theorem 3.4.9. *If $A \to B$ by multiplying a row (column) of A by a scalar $c \in F$, then $|B| = c|A|$.*

Remark. If A is an $n \times n$ matrix, then $|cA| = c^n|A|$.

Theorem 3.4.10. *If $A \to B$ by adding a multiple of a row (column) of A to another row (column) of A, then $|B| = |A|$.*

Corollary 3.4.11. *If E is an elementary matrix, then $\det E \neq 0$.*

Lemma 3.4.12. *Let E be an elementary matrix. Then $|EA| = |E||A|$ for any matrix A. In particular, if $E_1, E_2 \ldots, E_s$ are elementary matrices, then*

$$|E_1E_2 \ldots E_s| = |E_1||E_2| \ldots |E_s|.$$

Theorem 3.4.13. *Let A be a square matrix. Then, A is invertible $\Leftrightarrow \det A \neq 0$.*

Theorem 3.4.14. *The determinant of a product of two matrices A and B is the product of their determinants; that is $|AB| = |A||B|$.*

Definition. Consider an n-square matrix $A = [a_{ij}]$. Let $M_{ij}(A)$ denote the $(n-1)$-square submatrix of A obtained by deleting its ith row and jth column. The determinant $|M_{ij}(A)|$ is called the **minor** of the element a_{ij} of A, and we define the **cofactor** of a_{ij}, denoted by $C_{ij}(A)$, to be the "signed" minor:

$$C_{ij}(A) = (-1)^{i+j}|M_{ij}(A)|.$$

Recall that

$$\begin{aligned} |A| &= \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)} \\ &= a_{ij}\mathscr{C}_{ij}(A) + \text{ (terms which do not contain } a_{ij} \text{ as a factor)}. \end{aligned}$$

Lemma 3.4.15. *$C_{ij}(A) = \mathscr{C}_{ij}(A)$ for all $i, j \in \{1, \ldots, n\}$.*

Grant this lemma, we observe that

$$|A| = \sum_{j=1}^{n} a_{ij}\mathscr{C}_{ij}(A) = \sum_{j=1}^{n} a_{ij}C_{ij}(A).$$

Therefore, we have shown.

Theorem 3.4.16. [Laplace] *The determinant of a square matrix $A = [a_{ij}]$ is equal to the sum of the products obtained by multiplying the elements of any row (column) by their respective cofactors:*

$$|A| = a_{i1}C_{i1}(A) + a_{i2}C_{i2}(A) + \cdots + a_{in}C_{in}(A) = \sum_{j=1}^{n} a_{ij}C_{ij}(A)$$

$$|A| = a_{1j}C_{1j}(A) + a_{2j}C_{2j}(A) + \cdots + a_{nj}C_{nj}(A) = \sum_{i=1}^{n} a_{ij}C_{ij}(A)$$

for all $i, j \in 1, 2, \ldots, n$.

Remark. The above formulas for $|A|$ is called the **Laplace expansions** of the determinant of A by the ith row and the jth column. Together with the elementary row operations, they offer a method of simplifying the computation of $|A|$.

Next we proceed to prove the lemma.

Proof of Lemma 3.4.15. Note that for *any* matrix B,

$$\begin{aligned}|B| &= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \dots b_{n-1,\sigma(n-1)} b_{n\sigma(n)} \\ &= b_{nn} \sum_{\substack{\sigma\in S_n \\ \sigma(n)=n}} (\operatorname{sgn}\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \dots b_{n-1,\sigma(n-1)} + (\text{other terms which do not contain } b_{nn}).\end{aligned}$$

Thus,

$$\begin{aligned}\mathscr{C}_{nn}(B) &= \sum_{\substack{\sigma\in S_n \\ \sigma(n)=n}} (\operatorname{sgn}\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \dots b_{n-1,\sigma(n-1)} \\ &= \sum_{\tau\in S_{n-1}} (\operatorname{sgn}\tau) b_{1\tau(1)} b_{2\tau(2)} \dots b_{n-1,\tau(n-1)} \\ &= \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1,n-1} \\ b_{21} & b_{22} & \dots & b_{2,n-1} \\ & & \ddots & \\ b_{n-1,1} & b_{n-1,2} & \dots & b_{n-1,n-1} \end{vmatrix} \\ &= \text{ determinant of the matrix obtained from deleting the } n\text{th row and } n\text{th column of } B.\end{aligned}$$

Write

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ & \vdots & & & & \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ & \vdots & & & & \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix}.$$

To compute $\mathscr{C}_{ij}(A)$, we row reduce A to A' by interchanging rows $n-i$ times and columns $n-j$ times as shown:

$$A' = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2,j-1} & a_{2,j+1} & \dots & a_{2n} & a_{2j} \\ & \vdots & & & & & \vdots & \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} & a_{i-1,j} \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} & a_{i+1,j} \\ & \vdots & & & & & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{nn} & a_{nj} \\ a_{i1} & a_{i2} & \dots & a_{i,j-1} & a_{i,j+1} & \dots & a_{in} & a_{ij} \end{bmatrix}.$$

Hence,

$$|A'| = (-1)^{(n-i)+(n-j)}|A| = (-1)^{-i-j}|A|.$$

That is,

$$\begin{gathered}|A| = (-1)^{i+j}|A'| \\ a_{ij}\mathscr{C}_{ij}(A) + (\text{other terms}) = (-1)^{i+j} a_{ij}\mathscr{C}_{nn}(A') + (\text{other terms}).\end{gathered}$$

Therefore,

$$\begin{aligned}\mathscr{C}_{ij}(A) &= (-1)^{i+j}\mathscr{C}_{nn}(A') \\ &= (-1)^{i+j}(\text{the determinant of the matrix obtained from} \\ &\qquad\qquad \text{deleting the } n^{\text{th}} \text{ row and the } n^{\text{th}} \text{ column of } A') \\ &= (-1)^{i+j}|M_{ij}(A)| \\ &= C_{ij}(A).\end{aligned}$$

This completes the lemma. □

Definition. Let $A = [a_{ij}]$ be an $n \times n$ matrix and let $C_{ij}(A)$ denote the cofactor of a_{ij}. The **classical adjoint** of A, denoted by $\operatorname{adj} A$, is the transpose of the matrix of the cofactors of A, namely,

$$\operatorname{adj} A = [C_{ij}(A)]^T.$$

We say "classical adjoint" here instead of simply "adjoint" because the term "adjoint" will be used for an entirely different concept.

Theorem 3.4.17. *Let A be a square matrix. Then*

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = |A|I$$

where I is the identity matrix. Thus, if $|A| \neq 0$, then

$$A^{-1} = \frac{1}{|A|}(\operatorname{adj} A).$$

For any $n \times n$ matrix A and any $\vec{b} \in F^n$, let $A_i(\vec{b})$ be the matrix obtained from A by replacing the ith column by the vector $\vec{b}$, that is,

$$A_j(\vec{b}) = \begin{bmatrix} \vec{a}_1 & \cdots & \underbrace{\vec{b}}_{j\text{th}} & \cdots & \vec{a}_n \end{bmatrix}$$

for all $j = 1, 2, \ldots, n$.

Theorem 3.4.18. [Cramer's rule] *Let A be an invertible $n \times n$ matrix. For any $\vec{b} \in F^n$, the unique solution $\vec{x}$ of $A\vec{x} = \vec{b}$ has entries given by*

$$x_j = \frac{|A_j(\vec{b})|}{|A|}, \quad j = 1, 2, \ldots, n.$$

Exercises for Chapter 3.

1. The following matrices are echelon forms of coefficient matrices of linear systems. Which has a unique solution? Why?

(a) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

2. Find the general solution to the linear system

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & + & x_3 & - & 2x_4 & = & 5 \\ 2x_1 & + & 4x_2 & + & x_3 & + & x_4 & = & 9 \\ 3x_1 & + & 6x_2 & + & 2x_3 & - & x_4 & = & 14 \end{array}$$

3. Consider the linear system with parameter a

$$\begin{array}{rcrcrcl} (2a-1)x & + & ay & - & (a+1)z & = & 1 \\ ax & + & y & - & 2z & = & 1 \\ 2x & + & (3-a)y & + & (2a-6)z & = & 1 \end{array}$$

Determine, with proof, for which a this system has
(a) no solution (b) a unique solution (c) more than one solutions.

4. Consider the linear system

$$\begin{array}{rcrcrcl} x & + & 2y & + & z & = & 3 \\ & & ay & + & 5z & = & 10 \\ 2x & + & 7y & + & az & = & b \end{array}$$

(a) Find those values of a for which the system has a unique solution.
(b) Find those pairs of values (a, b) for which the system has more than one solutions.

5. If $A\vec{x} = \vec{b}$ has more than one solutions, why is it impossible for $A\vec{x} = \vec{c}$ (new right-hand side) to have only one solution? Could $A\vec{x} = \vec{c}$ have no solution?
6. Let $A\vec{x} = \vec{0}$ be a homogeneous system of n linear equations in n unknowns and let Q be an invertible $n \times n$ matrix. Show that $A\vec{x} = \vec{0}$ has a nontrivial solution if and only if $(QA)\vec{x} = \vec{0}$ has a nontrivial solution.
7. If $A\vec{x} = \vec{b}$ has two distinct solutions $\vec{p}$ and $\vec{q}$, find two distinct solutions to $A\vec{x} = \vec{0}$.
8. Under what conditions on b_1 and b_2 (if any) is $A\vec{x} = \vec{b}$ consistent (has a solution)?

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 0 & 7 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

9. Find the number c so that (if possible) the rank of A is (a) 1 (b) 2 (c) 3

$$A = \begin{bmatrix} 6 & 4 & 2 \\ -3 & -2 & -1 \\ 9 & 6 & c \end{bmatrix}$$

10. Suppose $A = \begin{bmatrix} 1 & 2 & 1 & b \\ 2 & a & 1 & 8 \\ * & * & * & * \end{bmatrix}$ has the reduced echelon form $R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.
(a) Find a and b. (b) Solve $A\vec{x} = \vec{0}$.
11. Let A be an $m \times n$ matrix for which

$$A\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ has no solutions and } A\vec{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ has a unique solution.}$$

(a) Give all possible information about m and n and the rank of A.
(b) Find all solutions of $A\vec{x} = \vec{0}$ and explain your answer.
12. Let A be an 3×4 matrix for which

$$A\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ has no solutions} \quad \text{and} \quad A\vec{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ has more than one solutions.}$$

(a) Give all possible values for $\operatorname{rank} A$.
(b) Do we always have more than one solutions for $A\vec{x} = \vec{0}$? Explain your answer.
(c) Is it possible to have a vector $\vec{b}$ such that $A\vec{x} = \vec{b}$ has a unique solution? Why?
13. Find the value for c in the following n by n inverse:

$$\text{if } A = \begin{bmatrix} n & -1 & \dots & -1 \\ -1 & n & \dots & -1 \\ \dots & \dots & \dots & -1 \\ -1 & -1 & \dots & n \end{bmatrix} \text{ then } A^{-1} = \frac{1}{n+1} \begin{bmatrix} c & 1 & \dots & 1 \\ 1 & c & \dots & 1 \\ \dots & \dots & \dots & 1 \\ 1 & 1 & \dots & c \end{bmatrix}.$$

14. If E is an elementary matrix, prove that E^T is an elementary matrix.

15. Let E_1, E_2 and E_3 denote, respectively, the elementary row operations

"Interchange rows R_1 and R_2" "Multiply R_3 by 5" "Replace R_2 by $-3R_1 + R_2$".

(a) Find the corresponding 3-square elementary matrices E_1, E_2 and E_3.
(b) Find the inverses of matrices E_1, E_2 and E_3.

16. Let A be a 3×3 invertible matrix. Construct B by replacing R_3 of matrix A by $R_3 - 4R_1$. How do we find B^{-1} from A^{-1}? Explain.

17. Let A be a 3×3 matrix and $B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Consider the augmented matrix $C = [A : B]$. After row reducing C, we get the following matrix

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -3 & -4 \\ 0 & 1 & 0 & -1 & 2 & 2 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right].$$

Compute A^{-1}.

18. (a) For which values of the parameter c is $A = \begin{bmatrix} -2 & 1 & c \\ 0 & -1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$ invertible?

(b) For which values of e is the matrix $A = \begin{bmatrix} 5 & e & e \\ e & e & e \\ 1 & 2 & e \end{bmatrix}$ *not* invertible?

19. Let $A = \begin{bmatrix} a & b & b \\ a & a & b \\ a & a & a \end{bmatrix}$. If $a \neq 0$ and $a \neq b$, prove that A is invertible and find A^{-1} in terms of a and b.

20. Show that if $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix}$ is an elementary matrix, then at least one entry in the third row must be zero.

21. In each case find an elementary matrix E such that $B = EA$.
(a) $A = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}, B = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}, B = \begin{bmatrix} -1 & -3 \\ 3 & -1 \end{bmatrix}$

22. In each case find an invertible matrix U such that $UA = B$, and express U as a product of elementary matrices.
(a) $A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & -1 & -2 \\ 3 & 0 & 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 3 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix}$

23. In each case find invertible matrices U and V such that UAV is in the Smith normal form.
(a) $A = \begin{bmatrix} 1 & 1 & -1 \\ -2 & -2 & 4 \end{bmatrix}$ (b) $A = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$ (c) $A = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$ (d) $A = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & -1 & 0 & 3 \\ 0 & 1 & -4 & 1 \end{bmatrix}$

24. Let F be a field and $A = [a_{ij}] \in M_n(F)$. Define the **trace** of A to be the sum of the diagonal elements, that is,

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

(a) Show that the trace is a linear transformation from $M_n(F)$ onto F.
(b) If A and B are in $M_n(F)$, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
(c) If B is invertible, then $\operatorname{tr}(B^{-1}AB) = \operatorname{tr} A$.
(d) Prove that there are no square real matrices A and B such that $AB - BA = I_n$.

25. Let $A = \begin{bmatrix} 1 & 2 & 0 & 2 & 1 \\ -1 & -2 & 1 & 1 & 0 \\ 1 & 2 & -3 & -7 & -2 \end{bmatrix}$.
(a) Find bases for $\operatorname{Col} A$ and $\operatorname{Nul} A$. (b) Find $\operatorname{rank} A$ and $\operatorname{nullity} A$.

26. Determine the rank and nullity of $A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$.

27. If A is an $n \times n$ matrix such that $A^2 = A$ and $\operatorname{rank} A = n$, prove that $A = I_n$.

28. Let A be a 5×7 matrix with rank 4.
(a) What is the dimension of the solution space of $A\vec{x} = \vec{0}$?
(b) Does $A\vec{x} = \vec{b}$ have a solution for all $\vec{b} \in \mathbb{R}^5$? Explain.
29. Let A be a square matrix such that $A^k = \mathbf{0}$ for some positive integer k. Prove that $I + A$ is invertible.
30. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. If $m > n$, show that AB is not invertible.
31. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. Show that $AB = \mathbf{0}_{m\times m} \Leftrightarrow \operatorname{Col} B \subseteq \operatorname{Nul} A$.
32. Determine the sign of all permutations in S_4 and expand the determinant $\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$ by using permutations and their signs explicitly.
33. Determine the sign of the following permutations in S_5.
(a) 12354 (b) 12534 (c) 15243 (d) 54321
34. Show that if two rows (columns) of A are proportional, i.e., $R_k = cR_l$ for some $k < l$, then $|A| = 0$.
35. Let $A = [a_{ij}]$ be a square matrix of order n and $\sigma \in S_n$. If $A_\sigma = [a_{\sigma(i),j}]$, show that $|A_\sigma| = (\operatorname{sgn}\sigma)|A|$.
36. Prove that if n is odd, $1 + 1 \neq 0$ and A is a square matrix of order n with $A = -A^T$, then A is not invertible.
37. After the indicated row operations on a 3×3 matrix A with $\det A = -540$, matrices $A_1, A_2, \ldots, A_5$ are successively obtained:

$$A \xrightarrow{R_1+3R_2} A_1 \xrightarrow{R_{23}} A_2 \xrightarrow{3R_2-R_1} A_3 \xrightarrow{R_1-3R_2} A_4 \xrightarrow{2R_1} A_5 \ .$$

Determine the values of $|A_1|, |A_2|, |A_3|, |A_4|$ and $|A_5|$, respectively.
38. If A is an invertible square matrix of order $n > 1$, show that $\det(\operatorname{adj} A) = (\det A)^{n-1}$.
What is $\det(\operatorname{adj} A)$ if A is not invertible? Prove your answer.
39. Let A, B, C be 3×3 matrices with $\det A = 3, \det B^3 = -8, \det C = 2$. Compute
(a) $\det(ABC)$ (b) $\det(5AC^T)$ (c) $\det(A^3B^{-3}C^{-1})$ (d) $\det[B^{-1}(\operatorname{adj} C)]$.
40. Show that $\operatorname{adj} A^T = (\operatorname{adj} A)^T$.
41. Show that if A is invertible and $n > 2$, then $\operatorname{adj}(\operatorname{adj} A) = (\det A)^{n-2}A$.
42. If A and B are invertible, show that

$$\operatorname{adj}(AB) = (\operatorname{adj} B)(\operatorname{adj} A) \quad \text{and} \quad \operatorname{adj}(BAB^{-1}) = B(\operatorname{adj} A)B^{-1}.$$

43. Prove that if A is an invertible upper triangular matrix (all entries lying below the diagonal are zero), then $\operatorname{adj} A$ and A^{-1} are upper triangular.
44. Suppose the set of real-valued functions $f_1(x), f_2(x), \ldots, f_k(x)$ are all defined and are differentiable $k-1$ times on the interval $[a, b]$. The **Wronskian** of the set of functions is defined on this interval to be the determinant

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_k(x) \\ f_1'(x) & f_2'(x) & \cdots & f_k'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_k''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(k-1)}(x) & f_2^{(k-1)}(x) & \cdots & f_k^{(k-1)}(x) \end{vmatrix}.$$

Prove that a set of real-valued functions $\{f_1(x), f_2(x), \ldots, f_k(x)\}$ differentiable $k-1$ times on the interval $[a, b]$, are linearly independent if $W(x_0) \neq 0$ at some point x_0 in the interval.
45. Consider the interval $[-1, 1]$ and the two functions defined by

$$f(x) = \begin{cases} 0 & \text{if } -1 \le x \le 0, \\ x^2 & \text{if } 0 \le x \le 1 \end{cases} \quad \text{and} \quad g(x) = \begin{cases} x^2 & \text{if } -1 \le x \le 0, \\ 0 & \text{if } 0 \le x \le 1. \end{cases}$$

These functions are both differentiable. Show that f and g are linearly independent but $W(x) = 0$ for all $x \in [-1, 1]$. This provides an example to prove that the converse of the previous problem does not hold.
46. (a) Show that the functions $1, x, x^2, \ldots, x^k$ are linearly independent in the function space $C^0[0, 1]$.
(b) Show that the functions $\sin x, \sin 2x, \sin 3x, \ldots, \sin kx$ are linearly independent in the function space $C^0[0, 2\pi]$. (*Hint*. Use the Wronskian.)

47. Use induction to show that

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{vmatrix} = (-1)^{n+1}.$$

48. (a) Let x_1, x_2 and x_3 be numbers. Show that

$$V_2 = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1 \qquad \text{and} \qquad V_3 = \begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix} = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2).$$

(b) If $x_1, x_2, \ldots, x_n$ are numbers, then show by induction that

$$V_n = \begin{vmatrix} 1 & x_1 & \ldots & x_1^{n-1} \\ 1 & x_2 & \ldots & x_2^{n-1} \\ & & \ldots & \\ 1 & x_n & \ldots & x_n^{n-1} \end{vmatrix} = \prod_{i<j} (x_j - x_i).$$

This determinant is called the **Vandermonde determinant**. (*Hint*. To do the induction easily, multiply each column by x_1 and subtract it from the next column on the right starting from the right-hand side. We shall find that $V_n = (x_n - x_1) \ldots (x_2 - x_1) V_{n-1}$.)

4 | Linear Transformations

4.1 Linear Functionals

Definition. Let V and W be two vector spaces over F. We write $\mathcal{L}(V, W)$ for the set of all linear transformations from V to W, that is,

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

Then $\mathcal{L}(V, W)$ is a vector space over F with the operations defined by for $S, T \in \mathcal{L}(V, W)$,

$$(S + T)(\vec{v}) = S(\vec{v}) + T(\vec{v}) \quad \text{and} \quad (cT)(\vec{v}) = cT(\vec{v})$$

for all $\vec{v} \in V$ and $c \in F$. Note that the *zero function* is its zero vector and $(-T)(\vec{v}) = -T(\vec{v})$ for all $\vec{v} \in V$.

Remark. By Theorem 1.4.1, for a given basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for an n-dimensional vector space V, there exists a unique linear transformation $T : V \to W$ such that $T(\vec{v}_i) = \vec{w}_i \in W$ for all $i \in \{1, 2, \ldots, n\}$. Then for $S, T \in \mathcal{L}(V, W)$, $(S(\vec{v}_\alpha) = T(\vec{v}_\alpha)$ for all $i \in \{1, 2, \ldots, n\}) \Rightarrow S = T$. *Hence, to show that two linear transformations are identical, it suffices to see the equality on some basis of V.*

Theorem 4.1.1. *Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for V and let $\mathcal{C} = \{\vec{w}_1, \ldots, \vec{w}_m\}$ be a basis for W. For each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, we define*

$$T_{ij}(\vec{v}_k) = \begin{cases} \vec{w}_j & \text{if } i = k, \\ \vec{0}_W & \text{if } i \neq k, \end{cases}$$

for all $k \in \{1, \ldots, n\}$. By Theorem 1.4.1, $T_{ij} \in \mathcal{L}(V, W)$ for all i, j. Then

$$\{T_{ij} : i \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}\}$$

is a basis for $\mathcal{L}(V, W)$. Hence, if $\dim V = n$ and $\dim W = m$, then $\dim \mathcal{L}(V, W) = mn$.

Definition. Let V be a vector space over a field F.
A linear transformation from V to F is also called a **linear functional**. Let

$$V^* = \mathcal{L}(V, F) \text{ and } V^{**} = (V^*)^* (= \mathcal{L}(V^*, F) = \mathcal{L}(\mathcal{L}(V, F), F)).$$

By Theorem 4.1.1, we have that if V is finite dimensional, then

$$\dim V = \dim V^* = \dim V^{**}$$

and thus, by Corollary 1.4.14, $V \cong V^* \cong V^{**}$.

Definition. The space V^* is called the **dual space of** V and V^{**} is called the **double dual of** V.

Examples 4.1.1. The following functions are linear functionals.

1. $T : C^0[0,1] \to \mathbb{R}$ given by $T(f) = \displaystyle\int_0^1 f(x)\,dx$.
2. $T : F[x] \to F$ given by $T(p(x)) = p(1)$.

Remarks. 1. For $f \in V^*$,
(a) $f \neq 0 \Rightarrow \operatorname{im} f = F$
(b) if V is finite dimensional and $f \neq 0$, then $\operatorname{nullity} f = (\dim V) - 1$.
2. For $\vec{v} \in V$, if $f(\vec{v}) = 0$ for all $f \in V^*$, then $\vec{v} = \vec{0}$.

Theorem 4.1.2. *Let $\dim V = n$ and let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis of V. For each $i \in \{1, \ldots, n\}$, let $f_i \in V^*$ be such that*

$$f_i(\vec{v}_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then the following statements hold.

1. *$\{f_1, \ldots, f_n\}$ is a basis of V^* which is called the* **dual basis of** *$\mathcal{B}$.*
2. *$\forall f \in V^*$, $f = \displaystyle\sum_{i=1}^{n} f(\vec{v}_i) f_i = f(\vec{v}_1) f_1 + \ldots + f(\vec{v}_n) f_n$.*
3. *$\forall \vec{v} \in V$, $\vec{v} = \displaystyle\sum_{i=1}^{n} f_i(\vec{v}) \vec{v}_i = f_1(\vec{v}) \vec{v}_1 + \ldots + f_n(\vec{v}) \vec{v}_n$.*

For $\vec{v} \in V$, define $L_{\vec{v}} : V^* \to F$ by $L_{\vec{v}}(f) = f(\vec{v})$ for all $f \in V^*$.
Then $L_{\vec{v}} \in V^{**}$ for all $\vec{v} \in V$. Hence, $\{L_{\vec{v}} : \vec{v} \in V\} \subseteq V^{**}$.

Theorem 4.1.3. 1. *The map $\theta : \vec{v} \mapsto L_{\vec{v}}$ is a 1-1 linear transformation from V into V^{**}.*
2. *If V is finite-dimensional, then*
(a) the map $\theta : \vec{v} \mapsto L_{\vec{v}}$ is an isomorphism of V onto V^{**}
(b) $\forall L \in V^{**}, \exists! \vec{v} \in V, L = L_{\vec{v}}$.

Corollary 4.1.4. *If V is finite dimensional, then each basis of V^* is the dual of some basis of V.*

Example 4.1.2. Consider $V = \mathbb{R}_2[x]$, the vector space of all polynomials of degree less than 2 over $\mathbb{R}$. Let t_1, t_2, t_3 be three distinct real numbers and let $f_i(p(x)) = p(t_i)$ for all $p(x) \in \mathbb{R}_2[x]$ and $i = 1, 2, 3$.

Show that $\{f_1, f_2, f_3\}$ is a basis of V^* and find a basis of V such that $\{f_1, f_2, f_3\}$ is its dual basis.

Let V be an inner product space over a field $F = \mathbb{R}$ or $\mathbb{C}$.

1. $\forall \vec{w} \in V$, the map $\vec{v} \mapsto (\vec{v}, \vec{w})$ is a linear functional on V.
2. The maps $\vec{v} \mapsto (\vec{v}, \vec{w}_1)$ and $\vec{v} \mapsto (\vec{v}, \vec{w}_2)$ are identical $\Leftrightarrow \vec{w}_1 = \vec{w}_2$.

Theorem 4.1.5. *Let V be a finite dimensional inner product space and $f \in V^*$. Then $\exists! \vec{w} \in V, f(\vec{v}) = (\vec{v}, \vec{w})$ for all $\vec{v} \in V$. Hence, $V^* = \{f_{\vec{w}} : \vec{w} \in V\}$ where $f_{\vec{w}}(\vec{v}) = (\vec{v}, \vec{w})$ for all $\vec{v} \in V$.*

4.2 Quotient Spaces and Isomorphism Theorem

Let V be a vector space over a field F and let W be *any* subspace of V. For $\vec{v} \in V$, define

$$\vec{v} + W = \{\vec{v} + \vec{w} : \vec{w} \in W\}$$

which is called a **coset of** W. Then
(1) $\forall \vec{v}_1, \vec{v}_2 \in V, \vec{v}_1 + W = \vec{v}_2 + W \Leftrightarrow \vec{v}_1 - \vec{v}_2 \in W$,
(2) $\forall \vec{v}_1, \vec{v}_2 \in V, (\vec{v}_1 + W) \cap (\vec{v}_2 + W) = \emptyset$ or $\vec{v}_1 + W = \vec{v}_2 + W$ and
(3) $\forall \vec{v}_1, \vec{v}_2 \in V, (\vec{v}_1 + W) + (\vec{v}_2 + W) = (\vec{v}_1 + \vec{v}_2) + W$.
For $c \in F$ and $\vec{v} \in V$, define $c(\vec{v} + W) = c\vec{v} + W$.

Definition. Let $V/W = \{\vec{v} + W : \vec{v} \in V\}$. It is a vector space over F with respect to the operations

$$(\vec{v}_1 + W) + (\vec{v}_2 + W) = (\vec{v}_1 + \vec{v}_2) + W \text{and} \quad c(\vec{v}_1 + W) = c\vec{v}_1 + W,$$

and $\vec{0} + W$ is the zero vector of V/W and $-(\vec{v} + W) = (-\vec{v}) + W$ for all $\vec{v} \in V$.
The vector space V/W is called the **quotient space of** V **by** W.

Theorem 4.2.1. 1. *There is a linear transformation π from V onto V/W given by*

$$\pi : \vec{v} \mapsto \vec{v} + W \text{ for all } \vec{v} \in V.$$

Its kernel is equal to W. This map π is called the **canonical projection from** V **onto** V/W.
2. *If V is a finite dimensional vector space and W is a subspace of V, then V/W is finite dimensional and* $\dim(V/W) = \dim V - \dim W$.

Theorem 4.2.2. [Isomorphism Theorem] *Let V and W be two vector spaces over a field F and $T : V \to W$ a linear transformation. Then*

$$V/(\ker T) \cong \operatorname{im} T.$$

Example 4.2.1. Let A be an $m \times n$ matrix and $T_A : \mathbb{R}^n \to \mathbb{R}^m$ given by $T_A(\vec{x}) = A\vec{x}$. Then we have

$$\mathbb{R}^n/(\operatorname{Nul} A) \cong \operatorname{Col} A.$$

Moreover, if $\vec{b} \in \operatorname{Col} A$, then $A\vec{x} = \vec{b}$ has a solution, say $\vec{y}_p$. Theorem 4.2.2 also gives the correspondence

$$\vec{y}_p + \operatorname{Nul} A \longleftrightarrow \vec{b}.$$

This is Theorem 3.1.3 (2).

4.3 Matrix Representations

Definition. Let V be an n-dimensional vector space over a field F with an ordered basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ and $\vec{v} \in V$. Then $\forall \vec{v} \in V, \exists!(c_1, \ldots, c_n) \in F^n$,

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n \quad \text{and} \quad [\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in F^n$$

is called the **coordinate vector of** $\vec{v}$ **relative to the ordered basis** $\mathcal{B}$.

Example 4.3.1. Let $\mathcal{B} = \{(1,1,0,0),(1,0,1,0),(1,1,1,0),(0,0,0,2i)\}$ be an ordered basis for $\mathbb{C}^4$. Find $[(2,-16,3,-i)]_{\mathcal{B}}$.

We recall Theorems 1.4.12 and 1.4.13 as follows.

Theorem 4.3.1. *Let V be an n-dimensional vector space over F and $\mathcal{B}$ a basis for V.*

1. *For $\vec{v},\vec{w} \in V$ and $c \in F$, we have $[\vec{v}+\vec{w}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{B}} + [\vec{w}]_{\mathcal{B}}$ and $[c\vec{v}]_{\mathcal{B}} = c[\vec{v}]_{\mathcal{B}}$.*
2. *The map $\vec{v} \mapsto [\vec{v}]_{\mathcal{B}}$ is an isomorphism from V onto F^n.*
   This also implies $\forall \vec{u},\vec{v} \in V, [\vec{u}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{B}} \Leftrightarrow \vec{u} = \vec{v}$.

Theorem 4.3.2. *Let $T : V \to W$ be a linear transformation where $\dim V = m$ and $\dim W = n$, and let $\mathcal{B} = \{\vec{v}_1,\ldots,\vec{v}_n\}$ and $\mathcal{C} = \{\vec{w}_1,\ldots,\vec{w}_m\}$ be ordered bases of V and W, respectively. Then for each $j \in \{1,\ldots,n\}$, we have*

$$T(v_j) = d_{1j}\vec{w}_1 + d_{2j}\vec{w}_2 + \cdots + d_{mj}\vec{w}_m.$$

Hence, there exists a unique $m \times n$ matrix over a field F given by

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix}[T(\vec{v}_1)]_{\mathcal{C}} & [T(\vec{v}_2)]_{\mathcal{C}} & \cdots & [T(\vec{v}_n)]_{\mathcal{C}}\end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{m1} & d_{m2} & \ldots & d_{mn} \end{bmatrix}.$$

Furthermore, $\varphi : T \mapsto [T]_{\mathcal{B}}^{\mathcal{C}}$ is an isomorphism of $\mathcal{L}(V,W)$ onto $M_{m,n}(F)$.

Definition. The matrix $[T]_{\mathcal{B}}^{\mathcal{C}}$ is called the **matrix for T relative to the ordered bases $\mathcal{B}$ and $\mathcal{C}$** If $V = W$ and $\mathcal{B} = \mathcal{C}$, then we write $[T]_{\mathcal{B}}$ for $[T]_{\mathcal{B}}^{\mathcal{B}}$. In addition, if $T : F^n \to F^n$ is a linear transformation and $\mathcal{B}$ is the standard basis for F^n, we call $[T]_{\mathcal{B}}$ the **standard matrix for T**.

Note that for $\vec{v} \in V$,

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n,$$

so that

$$T(\vec{v}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n).$$

Thus,

$$[T(\vec{v})]_{\mathcal{C}} = c_1[T(\vec{v}_1)]_{\mathcal{C}} + c_2[T(\vec{v}_2)]_{\mathcal{C}} + \cdots + c_n[T(\vec{v}_n)]_{\mathcal{C}} = \begin{bmatrix}[T(\vec{v}_1)]_{\mathcal{C}} & [T(\vec{v}_2)]_{\mathcal{C}} & \cdots & [T(\vec{v}_n)]_{\mathcal{C}}\end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

That is,

$$[T(\vec{v})]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}\,[\vec{v}]_{\mathcal{B}} \text{ for all } \vec{v} \in V.$$

We conclude this in the following diagram.

$$\begin{tikzcd} \vec{v} \arrow[r, "T"] \arrow[d] & T(\vec{v}) \arrow[d] \\ {[\vec{v}]_{\mathcal{B}}} \arrow[r, "{[T]_{\mathcal{B}}^{\mathcal{C}}\times}"] & {[T(\vec{v})]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}[\vec{v}]_{\mathcal{B}}} \end{tikzcd}$$

Example 4.3.2. 1. Let $\mathcal{B} = \{1+x, x\}$ be an ordered basis for $\mathbb{R}_1[x]$ and $\mathcal{C} = \{1+x, x, x^2-1, x^3\}$ an ordered basis for $\mathbb{R}_3[x]$.
Let $T : \mathbb{R}_1[x] \to \mathbb{R}_3[x]$ be a linear transformation defined by

$$T(a+bx) = x^2(a+bx) \text{ for all } a, b \in \mathbb{R}.$$

Find $[T]_{\mathcal{B}}^{\mathcal{C}}$.

2. Suppose $T : M_{22}(\mathbb{R}) \to \mathbb{R}^3$ is a linear transformation with $[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$ where $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ and $\mathcal{C} = \{(1,0,0), (0,1,0), (0,0,1)\}$.
Compute $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$.

Theorem 4.3.3. *Let V, W and Z be finite-dimensional vector spaces over a field F and let $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ be ordered bases of V, W and Z, respectively.*
If $S : V \to W$ and $T : W \to Z$ are linear transformations, then

$$[T \circ S]_{\mathcal{B}}^{\mathcal{D}} = [T]_{\mathcal{C}}^{\mathcal{D}} [S]_{\mathcal{B}}^{\mathcal{C}}.$$

Moreover, if $V = W = Z$ and $\mathcal{B} = \mathcal{C} = \mathcal{D}$, then $[T \circ S]_{\mathcal{B}} = [T]_{\mathcal{B}} [S]_{\mathcal{B}}$.

Corollary 4.3.4. *Let V be a finite-dimensional vector space, $\mathcal{B}$ an ordered basis and $T : V \to V$ a linear transformation. Then*
1. *T is an isomorphism $\Rightarrow [T]_{\mathcal{B}}$ is invertible and $[T]_{\mathcal{B}}^{-1} = \left[T^{-1}\right]_{\mathcal{B}}$.*
2. *$[T]_{\mathcal{B}}$ is invertible $\Rightarrow T$ is an isomorphism and $\left[T^{-1}\right]_{\mathcal{B}} = [T]_{\mathcal{B}}^{-1}$.*

Theorem 4.3.5. *Let $T : V \to W$ be a linear transformation where $\dim V = n$ and $\dim W = m$. If $\mathcal{B}$ and $\mathcal{C}$ are any ordered bases for V and W, respectively, then $\operatorname{rank} T = \operatorname{rank} [T]_{\mathcal{B}}^{\mathcal{C}}$.*

Example 4.3.3. Define $T : \mathbb{R}_2[x] \to \mathbb{R}^3$ by $T(a+bx+cx^2) = (a-2b, 3c-2a, 3c-4b)$ for all $a, b, c \in \mathbb{R}$. Compute $\operatorname{rank} T$.

4.4 Change of Bases

Definition. Let V be a n-dimensional vector space over a field F. with an ordered basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$. If $\mathcal{B}' = \{\vec{v}_1', \ldots, \vec{v}_n'\}$ is another ordered basis for V, we define the **transition** or **change of coordinate matrix from $\mathcal{B}'$ to $\mathcal{B}$** by $P_{\mathcal{B} \to \mathcal{B}'} = [I]_{\mathcal{B}}^{\mathcal{B}'}$

Theorem 4.4.1. *Let $\mathcal{B}, \mathcal{B}'$ and $\mathcal{B}''$ be bases for V. Then*
1. *$\forall \vec{v} \in V, [\vec{v}]_{\mathcal{B}'} = P_{\mathcal{B} \to \mathcal{B}'} [\vec{v}]_{\mathcal{B}}$,*
2. *$P_{\mathcal{B} \to \mathcal{B}} = I_n$,*
3. *$P_{\mathcal{B} \to \mathcal{B}'}$ is invertible and $(P_{\mathcal{B} \to \mathcal{B}'})^{-1} = P_{\mathcal{B}' \to \mathcal{B}}$,*
4. *$P_{\mathcal{B} \to \mathcal{B}''} = P_{\mathcal{B}' \to \mathcal{B}''} P_{\mathcal{B} \to \mathcal{B}'}$.*

Example 4.4.1. Let $\mathcal{B} = \left\{ \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ and $\mathcal{B}' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$ be ordered bases for $\mathbb{R}^2$.
(a) Find $\underset{\mathcal{B} \to \mathcal{B}'}{P}$.
(b) If $\vec{v} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, find $[\vec{v}]_{\mathcal{B}}$ and $[\vec{v}]_{\mathcal{B}'}$.

Definition. A linear transformation from V to V is called a **linear operator on** V.

Theorem 4.4.2. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two bases for a finite dimensional vector space V. If $T : V \to V$ is a linear operator, then*

$$[T]_{\mathcal{B}'} = [I]_{\mathcal{B}'}^{\mathcal{B}}[T]_{\mathcal{B}}[I]_{\mathcal{B}}^{\mathcal{B}'} = (P_{\mathcal{B}\to\mathcal{B}'})^{-1}[T]_{\mathcal{B}}(P_{\mathcal{B}\to\mathcal{B}'}).$$

Example 4.4.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation with standard matrix $\begin{bmatrix} 2 & 1 & 0 \\ 6 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$. Find $[T]_{\mathcal{B}'}$ where $\mathcal{B}' = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -6 \end{bmatrix} \right\}$.

From the above theorem, we have

$$\det[T]_{\mathcal{B}} = \det[T]_{\mathcal{B}'} \quad \text{and} \quad \operatorname{rank}[T]_{\mathcal{B}} = \operatorname{rank}[T]_{\mathcal{B}'}$$

for any two bases $\mathcal{B}$ and $\mathcal{B}'$ for V.

Definition. If $T : V \to V$ is a linear operator, we define the **determinant of** T by

$$\det T = \det[T]_{\mathcal{B}} \quad \text{for some basis } \mathcal{B} \text{ for } V.$$

Definition. For $n \times n$ matrices A and B, we say that A is **similar** to B, $A \sim B$, if there exists an invertible matrix $P \in M_n(F)$ such that $B = P^{-1}AP$.

Remarks.
1. $\sim$ is an equivalence relation on $M_n(F)$.
2. If $A \sim B$, then $A^T \sim B^T$, $A^k \sim B^k$ for all $k \in \mathbb{N}$, and $A^{-1} \sim B^{-1}$ (if inverses exist).
3. $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$ are similar for any two bases $\mathcal{B}$ and $\mathcal{B}'$ of V.

Definition. The **trace** of an $n \times n$ matrix A is the sum of the diagonal elements.

Theorem 4.4.3. *Let A and B be similar matrices. Then*
1. $\det A = \det B$,
2. $\operatorname{rank} A = \operatorname{rank} B$,
3. $\operatorname{tr} A = \operatorname{tr} B$.

Exercises for Chapter 4.
1. If $T : V \to W$ is an isomorphism and $\mathcal{B}$ is a basis for V, prove that $T(\mathcal{B})$ is a basis for W.
2. Let $T : V \to V$ be a linear transformation. Suppose that there exists a $\vec{v} \in V$ such that $T(T(\vec{v})) \neq \vec{0}$ and $T(T(T(\vec{v}))) = \vec{0}$. Prove that $\{\vec{v}, T(\vec{v}), T(T(\vec{v}))\}$ is linearly independent.
3. Let $S, T \in \mathcal{L}(V, W)$ and $c \in F$. Prove that:
   (a) $\ker S \cap \ker T \subseteq \ker(S + cT)$ (b) $\operatorname{im}(S + T) \subseteq \operatorname{im} S + \operatorname{im} T$.
4. Let E be a linear transformation on a vector space V such that $E \circ E = E$.
   Prove that the following statements hold.
   (a) $\forall \vec{v} \in V, \vec{v} \in \operatorname{im} E \Leftrightarrow E(\vec{v}) = \vec{v}$ (b) $\forall \vec{v} \in V, \vec{v} - E(\vec{v}) \in \ker E$ (c) $V = \ker E \oplus \operatorname{im} E$.
5. Let $f, g \in V^*$. If $\ker f \subseteq \ker g$, prove that $g = cf$ for some $c \in F$.
6. Let V be an n-dimensional vector space over F.
   If $f, g \in V^*$ are linearly independent, find $\dim(\ker f \cap \ker g)$.
7. If V and W are finite dimensional vector spaces which are isomorphic, prove that $V^* \cong W^*$.
8. Let $\mathcal{B} = \{(1, 0, -1), (1, 1, 1), (2, 2, 0)\}$ be a basis for $\mathbb{R}^3$. Find the dual basis of $\mathcal{B}$.

9. Consider $V = \mathbb{R}_1[x]$. Let $f_1 : V \to \mathbb{R}$ and $f_2 : V \to \mathbb{R}$ be defined by

$$f_1(p(x)) = \int_0^1 p(x)\,dx \qquad \text{and} \qquad f_2(p(x)) = \int_0^2 p(x)\,dx.$$

Clearly, $f_1, f_2 \in V^*$. Prove that $\{f_1, f_2\}$ is a basis for V^* and find a basis of V such that $\{f_1, f_2\}$ is its dual basis.

10. (a) Let W be a subspace of a finite dimensional vector space V.
If $\mathcal{B} = \{x_1, \ldots, x_m\}$ is a basis for W and $\{x_1, \ldots, x_m, x_{m+1}, \ldots, x_n\}$ is a basis of V, show that $\{x_{m+1} + W, \ldots, x_n + W\}$ is a basis for V/W.
(b) Let $H = \operatorname{Span}\{(1, 1, -1)\}$. Determine a basis for $\mathbb{R}^3/H$.

11. Let W_1 and W_2 be two subspaces of a vector space V.
Define $T : W_1 + W_2 \to W_2/(W_1 \cap W_2)$ by $T(\vec{w}_1 + \vec{w}_2) = \vec{w}_2 + (W_1 \cap W_2)$ for all $\vec{w}_1 \in W_1$ and $\vec{w}_2 \in W_2$.
(a) Prove that T is well defined and is an onto linear transformation.
(b) Prove that $\ker T = W_1$.
(c) Conclude by Theorem 4.2.2 that $(W_1 + W_2)/W_1 \cong W_2/(W_1 \cap W_2)$.
This is a generalization of Theorem 1.4.8.

12. If W_1 and W_2 are subspaces of V with $W_1 \subseteq W_2$.
Define $T : V/W_1 \to V/W_2$ by $T(\vec{v} + W_1) = \vec{v} + W_2$ for all $\vec{v} \in V$.
(a) Prove that T is well defined and is an onto linear transformation.
(b) Prove that $\ker T = W_2/W_1$.
(c) Conclude by Theorem 4.2.2 that $(V/W_1)/(W_2/W_1) \cong V/W_2$.

13. Let U, V and W be finite dimensional vector spaces over a field F. Let $S : U \to V$ and $T : V \to W$ be linear transformations such that $T \circ S$ is the zero map. Show that

$$\dim(W/\operatorname{im} T) - \dim(\ker T/\operatorname{im} S) + \operatorname{nullity} S = \dim W - \dim V + \dim U.$$

14. Let V and W be finite dimensional vector spaces over a field F. Let U be a subspace of V and $T : V \to W$ a linear transformation.
(a) Prove that $\dim(V/U) \geq \dim(T(V)/T(U))$.
(b) If T is 1-1, prove also that the inequality in (a) becomes an equality.

15. For $S \subseteq V$, let $A(S) = \{f \in V^* : f(\vec{v}) = 0 \text{ for all } \vec{v} \in S\}$. It is called the **annihilator of** S.
Prove that
(a) $A(S)$ is a subspace of V^* (b) If $S_1 \subseteq S_2$, then $A(S_1) \supseteq A(S_2)$
(c) If V is finite dimensional and W is a subspace of V, then $V^*/A(W) \cong W^*$.

16. Prove that $\forall S, T \in \mathcal{L}(V, V), S \circ T \in \mathcal{L}(V, V)$.

17. Let $T : V \to W$ be a linear transformation where $\dim V = \dim W = n$.
Prove that the following statements are equivalent.
(i) T is an isomorphism.
(ii) $[T]_{\mathcal{B}}^{\mathcal{C}C}$ is invertible for all ordered bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively.
(iii) $[T]_{\mathcal{B}}^{\mathcal{C}C}$ is invertible for some pair of ordered bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively.

18. Suppose the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$T(1, 1) = (2, 3) \quad \text{and} \quad T(-1, 1) = (4, 5).$$

Find the standard matrix for T.

19. Let $\mathcal{B} = \{1, x, x^2\}$ be an ordered basis for $\mathbb{R}_2[x]$ and $\mathcal{C} = \{(1, 0), (1, -1)\}$ an ordered basis for $\mathbb{R}^2$. Find $[T]_{\mathcal{B}}^{\mathcal{C}}$ if $T : \mathbb{R}_2[x] \to \mathbb{R}^2$ is a linear transformation defined by $T(a + bx + cx^2) = (a + c, 2b)$ for all $a, b, c \in \mathbb{R}$.

20. Let $\mathcal{B} = \{\sin t, \cos t\}$ and $\mathcal{B}' = \{\sin t + 2\cos t, \sin t - \cos t\}$ be ordered bases for $H = \operatorname{Span} \mathcal{B} = \operatorname{Span} \mathcal{B}'$ which is a subspace of $C^1(-\infty, \infty)$. Let $D : H \to H$ defined by $D(f) = f''$ for all

21. If $T : V \to W$ is an isomorphism, prove that $([T]_{\mathcal{B}}^{\mathcal{C}})^{-1} = [T^{-1}]_{\mathcal{C}}^{\mathcal{B}}$ for all ordered bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively.

22. Let $T : \mathbb{R}_2[x] \to \mathbb{R}^3$ be a linear transformation defined by $T(a + bx + cx^2) = (a - c, b, a - 2c)$ for all $a, b, c \in \mathbb{R}$. If $\mathcal{B} = \{1, x, x^2\}$ and $\mathcal{C} = (1, 0, 0), (0, 1, 0), (0, 0, 1)$, find $[T]_{\mathcal{B}}^{\mathcal{C}}$ and the formula for T^{-1}.

23. Let $T : \mathbb{R}_n[x] \to \mathbb{R}_n[x]$ be a linear transformation defined by

$$T(p(x)) = p(x) + xp'(x),$$

where $p'(x)$ is the derivative of $p(x)$. Show that T is an isomorphism by finding $[T]_{\mathcal{B}}$ where $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$.

24. Let α be a real number. Define a linear transformation $T_\alpha : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ by

$$T_\alpha(A) = A + \alpha A^T \text{ for all } A \in M_2(\mathbb{R}).$$

If $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$, find $[T]_\mathcal{B}$ and conclude that T_α is invertible if $\alpha^2 \neq 1$.

25. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by $T(x, y) = (-y, x)$ for all $x, y \in \mathbb{R}$. Prove that
(a) $\forall c \in \mathbb{R}, (A - cI_2)$ is invertible,
(b) if $\mathcal{B}$ is an ordered basis for $\mathbb{R}^2$ and $[T]_\mathcal{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $bc \neq 0$.

---

5 | Structure Theorems

5.1 Eigenvalues and Eigenvectors

We first recall some numerical examples.

Example 5.1.1. Diagonalize $A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$.

That is, find an invertible matrix P and a diagonal matrix D (if any) such that $A = PDP^{-1}$.

Example 5.1.2. Let $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$ and $T(\vec{x}) = A\vec{x}$ a matrix transformation on $\mathbb{R}^3$.

Find a basis $\mathcal{B}$ (if any) such that $[T]_{\mathcal{B}}$ is a diagonal matrix. Given $\det(A - \lambda I) = -(\lambda - 1)(\lambda - 2)^2$.

Definition. Let V be a vector space over a field F and $T \in \mathcal{L}(V, V)$.
A scalar $\lambda \in F$ is called an **eigenvalue** or **characteristic value of** T if there exists a *nonzero* vector $\vec{v} \in V$ such that $T(\vec{v}) = \lambda\vec{v}$. If λ is is an eigenvalue of T, then $\vec{v} \in V$ such that $T(\vec{v}) = \lambda\vec{v}$ is called an **eigenvector** or **characteristic vector of** T **associated with the characteristic value** λ. We have that

$$E_\lambda(T) = \{\vec{v} \in V : T(\vec{v}) = \lambda\vec{v}\} = \{\vec{v} \in V : (T - \lambda I)(\vec{v}) = \vec{0}_V\} = \ker(T - \lambda I)$$

is a subspace of V, called the **eigenspace** or **characteristic space of** T **associated with** λ.

Remark. λ is an eigenvalue of $T \Leftrightarrow \ker(T - \lambda I) \neq \vec{0}_V \Leftrightarrow T - \lambda I$ is not 1-1.

For matrix theory, we restrict ourselves to the case of V is n-dimensional. Then $\mathcal{L}(V, V) \cong M_n(F)$ with $T \mapsto [T]_{\mathcal{B}}$ for a fixed basis $\mathcal{B}$ of V. Hence, we can only work on $M_n(F)$.

Definition. Let $A \in M_n(F)$. The matrix transformation $T_A : F^n \to F^n$ is given by

$$T_A(\vec{x}) = A\vec{x}$$

for all $\vec{x} \in F^n$. An eigenvalue of T_A is called an **eigenvalue of** A and the eigenspace of T_A is called an **eigenspace of** A. In other words,

$$E_\lambda(A) = \{\vec{x} \in F^n : A\vec{x} = \lambda\vec{x}\} = \{\vec{x} \in F^n : (A - \lambda I_n)\vec{x} = \vec{0}_n\} = \mathrm{Nul}(A - \lambda I_n).$$

Then

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\Leftrightarrow \ker(T_A - \lambda I) \neq \vec{0}_n \\ &\Leftrightarrow \mathrm{Nul}(A - \lambda I_n) \neq \vec{0}_n \\ &\Leftrightarrow A - \lambda I_n \text{ is not invertible} \\ &\Leftrightarrow \det(A - \lambda I_n) = 0. \end{aligned}$$

Definition. The polynomial $c_A(x) = \det(xI_n - A)$ is called the **characteristic polynomial of** A.

Thus we have proved

Theorem 5.1.1. *For $A \in M_n(F)$, λ is an eigenvalue of $A \Leftrightarrow \det(A - \lambda I_n) = 0$, i.e., λ is a root of the characteristic polynomial of A.*

Since an eigenvalue of an $n \times n$ matrix A is a root of $c_A(x) = \det(xI_n - A)$ which has degree n and a polynomial of degree n over a field F has at most n roots in F, A has $\leq n$ eigenvalues.

Theorem 5.1.2. *An $n \times n$ matrix has at most n eigenvalues.*

Remark. If A is similar to B, then $\det A = \det B$ and

$$\det(B - \lambda I_n) = \det(P^{-1}AP - \lambda P^{-1}I_nP) = \det(P^{-1}(A - \lambda I_n)P) = \det(A - \lambda I_n).$$

Therefore, we have the following result.

Theorem 5.1.3. *If A and B are similar $n \times n$ matrices, then A and B have the same characteristic polynomial and eigenvalues (with same multiplicities).*

Example 5.1.3. The matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

have the same determinant, trace, characteristic polynomial and eigenvalue, but they are *not* similar because $PIP^{-1} = I$ for any invertible matrix P.

Definition. A **diagonal matrix** D is a square matrix such that all the entries off the main diagonal are zero, that is if D is of the form

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where $\lambda_1, \lambda_2, \dots, \lambda_n \in F$ (not necessarily distinct).

Definition. An $n \times n$ matrix A over F is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, there are an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$. In this case, we say that P **diagonalizes** A.

Definition. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V, V)$ a linear operator. We say that T is **diagonalizable** if there exists a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}$ is a diagonal matrix.

Theorem 5.1.4. *Let A be an $n \times n$ matrix.*

1. *A is diagonalizable $\Leftrightarrow$*
   A has eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ such that $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ is invertible.
2. *When this is the case, $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where*
   for each i, λ_i is the eigenvalue of A corresponding to $\vec{v}_i$.

Proof. Let $P = \begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_n\end{bmatrix}$ and $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.
Then $AP = PD$ becomes

$$A\begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{bmatrix} = \begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$$
$$\begin{bmatrix}A\vec{v}_1 & A\vec{v}_2 & \cdots & A\vec{v}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \cdots & \lambda_n\vec{v}_n\end{bmatrix}.$$

Comparing columns shows that $A\vec{v}_i = \lambda_i\vec{v}_i$ for each i, so

$$P^{-1}AP = D \Leftrightarrow \ P \text{ is invertible and } A\vec{v}_i = \lambda_i\vec{v}_i \text{ for all } i \in \{1, \dots, n\}.$$

The results follow. □

Theorem 5.1.5. *Let $\vec{v}_1, \dots, \vec{v}_m$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_m$ of an $n \times n$ matrix A. Then $\{\vec{v}_1, \dots, \vec{v}_m\}$ is linearly independent.*

Proof. We use induction on k.
If $k = 1$, then $\{\vec{v}_1\}$ is linearly independent because $\vec{v}_1 \neq \vec{0}$.
Let $k \geq 1$ and the theorem is true for any k eigenvectors.
Let $\vec{v}_1, \dots, \vec{v}_{k+1}$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_{k+1}$ of A.
Let $c_1, \dots, c_{k+1} \in F$ be such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{k+1}\vec{v}_{k+1} = \vec{0}. \tag{5.1.1}$$

Since $A\vec{v}_i = \lambda\vec{v}_i$ for all i, multiplying by A both sides gives

$$c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 + \cdots + c_{k+1}\lambda_{k+1}\vec{v}_{k+1} = \vec{0}. \tag{5.1.2}$$

Subtracting (5.1.2) by $\lambda_1 \times$(5.1.1), we have

$$c_2(\lambda_2 - \lambda_1)\vec{v}_2 + \cdots + c_{k+1}(\lambda_{k+1} - \lambda_1)\vec{v}_{k+1} = \vec{0}.$$

Since $\vec{v}_2, \dots, \vec{v}_{k+1}$ are k eigenvectors, they are linearly independent by induction hypothesis, so

$$c_2(\lambda_2 - \lambda_1) = \cdots = c_{k+1}(\lambda_{k+1} - \lambda_1) = 0.$$

However, $\lambda_1, \dots, \lambda_{k+1}$ are distinct, hence we get

$$c_2 = \cdots = c_{k+1} = 0.$$

This implies $c_1\vec{v}_1 = \vec{0}$, so $c_1 = 0$ because $\vec{v}_1 \neq \vec{0}$.
Therefore, $\{\vec{v}_1, \dots, \vec{v}_{k+1}\}$ is linearly independent. □

Corollary 5.1.6. *If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.*

Proof. Let $\vec{v}_1, \dots, \vec{v}_n$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_n$ of A.
Then they are linearly independent, and so $P = \begin{bmatrix}\vec{v}_1 & \dots & \vec{v}_n\end{bmatrix}$ is invertible
and $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. □

Lemma 5.1.7. *Let $\{\vec{v}_1, \dots, \vec{v}_k\}$ be a linearly independent set of eigenvectors of an $n \times n$ matrix A, extend it to a basis of F^n, and let*

$$P = \begin{bmatrix} \vec{v}_1 & \dots & \vec{v}_k & \vec{v}_{k+1} & \dots & \vec{v}_n \end{bmatrix}$$

which is invertible. If $\lambda_1, \dots, \lambda_k$ are the (not necessarily distinct) eigenvalues of A corresponding to $\vec{v}_1, \dots, \vec{v}_k$, respectively, then $P^{-1}AP$ has block form

$$P^{-1}AP = \begin{bmatrix} \operatorname{diag}(\lambda_1, \dots, \lambda_k) & B \\ 0 & C \end{bmatrix}$$

where B has size $k \times (n-k)$ and C has size $(n-k) \times (n-k)$.

Definition. An eigenvalue λ of a square matrix A is said to have **multiplicity** m if it occurs m times as a root of the characteristic polynomial $c_A(x)$.

In other words,

$$c_A(x) = (x - \lambda)^m g(x)$$

for some polynomial $g(x)$ such that $g(\lambda) \neq 0$.

Lemma 5.1.8. *Let λ be an eigenvalue of multiplicity m of a square matrix A. Then $\operatorname{nullity}(A - \lambda I) = \dim E_\lambda(A) \leq m$.*

Proof. Assume that $\dim E_\lambda(A) = d$ with basis $\{\vec{v}_1, \dots, \vec{v}_d\}$. By Lemma 5.1.7, there exists an invertible $n \times n$ matrix P such that

$$P^{-1}AP = \begin{bmatrix} \lambda I_d & B \\ 0 & C \end{bmatrix} = M$$

where I_d is the $d \times d$ identity matrix. Since M and A are similar,

$$\begin{aligned} c_A(x) = c_M(x) = \det(xI_n - M) &= \begin{vmatrix} (x-\lambda)I_d & B \\ 0 & xI_{n-d} - C \end{vmatrix} \\ &= (\det(x-\lambda)I_d)(\det(xI_{n-d} - C)) \\ &= (x-\lambda)^d c_C(x). \end{aligned}$$

Hence, $d \leq m$ because m is the highest power of $(x - \lambda)$ in $c_A(x)$. □

Theorem 5.1.9. *Let $\lambda_1, \dots, \lambda_k$ be all distinct eigenvalues of an $n \times n$ matrix A. For each $i \in \{1, \dots, k\}$, let m_i denote the multiplicity of λ_i and write $d_i = \operatorname{nullity}(A - \lambda_i I_n)$. Then $1 \leq d_i \leq m_i$ for all i, $n = m_1 + \cdots + m_k$ and*

$$c_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}.$$

Moreover, the following statements are equivalent.

(i) *A is diagonalizable.*
(ii) *$d_i = \operatorname{nullity}(A - \lambda_i I_n) = \dim E_{\lambda_i}(A) = m_i$ for all i.*
(iii) *$n = d_1 + \cdots + d_k$.*

5.2 Annihilating Polynomials

Let A be an $n \times n$ matrix over a field F. Since $\dim M_n(F) = n^2$, the set $\{I_n, A, A^2, \ldots, A^{n^2}\}$ is linearly dependent. Then there exist $c_0, c_1, \ldots, c_{n^2}$ in F not all zero such that

$$c_0 I_n + c_1 A + c_2 A^2 + \cdots + c_{n^2} A^{n^2} = \underline{0}.$$

Let $f(x)$ be the polynomial over F defined by $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n^2} x^{n^2}$. Then $f(x) \neq 0$ and $f(A) = 0$.

Let $g(x) = \alpha^{-1} f(x)$ where α is the leading coefficient of $f(x)$. Then $g(x)$ is monic (leading coefficient $= 1$) and $g(A) = \underline{0}$. Thus there exists a polynomial $p(x)$ over F such that
(a) $p(A) = \underline{0}$
(b) $p(x)$ is monic and
(c) $\forall$ nonzero polynomial $q(x)$, $q(A) = \underline{0} \Rightarrow \deg p(x) \leq \deg q(x)$.
We have that such $p(x)$ is *unique* (Proof!) and it is called the **minimal polynomial**. Note that if $k(x) \in F[x]$ and $k(A) = \underline{0}$, then $p(x) \mid k(x)$.

Remark. If A and B in $M_n(F)$ are similar, then they have the same minimal polynomial.

Recall that the characteristic polynomial of A is given by

$$c_A(x) = \det(xI_n - A).$$

Theorem 5.2.1. *The characteristic polynomial and minimal polynomial for A have the same roots.*

Remark. Although the minimal polynomial and the characteristic polynomial have the same roots, they may not be the same.

Example 5.2.1. The characteristic polynomial for $A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$ is $(x-1)(x-2)^2$ while

$$(A - I)(A - 2I) = 0,$$

so the minimal polynomial of A is $(x-1)(x-2)$. Notice that A is diagonalizable. In general, we have:

Theorem 5.2.2. *If an $n \times n$ matrix A is diagonalizable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then $(x - \lambda_1) \ldots (x - \lambda_k)$ is the minimal polynomial for A.*

Theorem 5.2.3. [Cayley-Hamilton] *If $f(x)$ is the characteristic polynomial of a matrix A, then $f(A) = \underline{0}$.*

Proof. Write $f(x) = \det(xI_n - A) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Let $B = xI_n - A$. Since $\operatorname{adj} B$ is a matrix such that each entry is obtained by using $(n-1) \times (n-1)$ submatrix of A and computing its determinant,

$$C_{ij}(B) = b_{ij}^{(n-1)} x^{n-1} + b_{ij}^{(n-2)} x^{n-2} + \cdots + b_{ij}^{(1)} x + b_{ij}^{(0)}$$

for all $i, j \in \{1, \ldots, n\}$. Thus

$$\begin{aligned} \operatorname{adj} B &= [C_{ij}(B)]_{n\times n}^T \\ &= [b_{ij}^{(n-1)} x^{n-1} + b_{ij}^{(n-2)} x^{n-2} + \cdots + b_{ij}^{(1)} x + b_{ij}^{(0)}]_{n\times n}^T \\ &= B_{n-1} x^{n-1} + B_{n-2} x^{n-2} + \cdots + B_1 x + B_0 \end{aligned}$$

where $B_i \in M_n(F)$. Recall that

$$(\det B)I_n = B(\operatorname{adj} B) = B(B_{n-1}x^{n-1} + B_{n-2}x^{n-2} + \cdots + B_1x + B_0).$$

Then

$$\begin{aligned}
(x^n + a_{n-1}x^{n-1} + \cdots + & a_1x + a_0)I_n \\
&= B(B_{n-1}x^{n-1} + B_{n-2}x^{n-2} + \cdots + B_1x + B_0) \\
&= (xI - A)(B_{n-1}x^{n-1} + B_{n-2}x^{n-2} + \cdots + B_1x + B_0) \\
&= B_{n-1}x^n + B_{n-2}x^{n-1} + \cdots + B_1x^2 + B_0x \\
&\quad - AB_{n-1}x^{n-1} + AB_{n-2}x^{n-2} + \cdots + AB_1x + AB_0.
\end{aligned}$$

This gives

$$\begin{aligned}
I &= B_{n-1} \\
a_{n-1}I &= B_{n-2} - AB_{n-1} \\
a_{n-2}I &= B_{n-3} - AB_{n-2} \\
&\vdots \\
a_1I &= B_0 - AB_1 \\
a_0I &= -AB_0.
\end{aligned}$$

Therefore

$$\begin{aligned}
A^n + & a_{n-1}A^{n-1} + \ldots a_1A + a_0I \\
&= A^nB_{n-1} + A^{n-1}(B_{n-2} - AB_{n-1}) + A^{n-2}(B_{n-3} - AB_{n-2}) + \ldots \\
&\quad + A(B_0 - AB_1) - AB_0 \\
&= \underline{0}
\end{aligned}$$

as desired. □

Example 5.2.2. Determine the minimal polynomial of $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$.

Some consequences of the Cayley-Hamilton are as follows.

Corollary 5.2.4. *The minimal polynomial of A divides its characteristic polynomial.*

Recall that

$$0 \text{ is an eigenvalue of } A \Leftrightarrow 0 = \det(A - 0I) = \det A \Leftrightarrow A \text{ is not invertible.}$$

Corollary 5.2.5. *If $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$ is the characteristic polynomial of an invertible matrix A, then $a_0 \neq 0$*

$$A^{-1} = -\frac{1}{a_0}(a_1I + a_2A + \cdots + A^{n-1}).$$

5.3 Symmetric and Hermitian Matrices

Definition. Let $F = \mathbb{R}$ or $\mathbb{C}$ and $A = [a_{ij}]$ a matrix over F.
The matrix A is said to be **symmetric** if $A = A^T$. We define $A^H = [\bar{a}_{ij}]^T$, the conjugate transpose of A, called A **Hermitian**. We say that A is **Hermitian or self-adjoint** if $A = A^H$.

Notice that symmetric and Hermitian matrices are square matrices and they coincide if $F = \mathbb{R}$.

Example 5.3.1. Let $A = \begin{bmatrix} 3 & 1 \\ 1 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 2+3i \\ 2-3i & 2 \end{bmatrix}$.
Then A is symmetric and both of them are Hermitian.

Theorem 5.3.1. *If A is a Hermitian matrix, then*
(1) $\vec{x}^H A\vec{x}$ is real for all $\vec{x} \in \mathbb{C}^n$ and (2) the eigenvalues of A are real.
That is, if A is Hermitian, then all roots of $c_A(x)$ are real.

Example 5.3.2. For vectors $\vec{x}$ and $\vec{y}$ in $\mathbb{C}^n$, we define $(\vec{x}, \vec{y}) = \vec{x}^H\vec{y}$.
Then $(\cdot, \cdot)$ is an inner product on $\mathbb{C}^n$ so that

$$\|\vec{x}\|^2 = \vec{x}^H\vec{x} = |x_1|^2 + \cdots + |x_n|^2 \quad \text{for all } \vec{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n.$$

Theorem 5.3.2. *Two eigenvectors corresponding to different eigenvalues of a Hermitian matrix are orthogonal to one another.*

Definition. For $F = \mathbb{R}$ or $\mathbb{C}$ and $U \in M_n(F)$, U is called **unitary** if $U^HU = I_n = UU^H$. If $F = \mathbb{R}$, a unitary matrix satisfies $U^TU = I_n = UU^T$ and may be called an **orthonormal matrix**.

Theorem 5.3.3. *Let $U \in M_n(\mathbb{C})$ be a unitary matrix.*
For the inner product defined in Example 5.3.2, we have
$(U\vec{x}, U\vec{y}) = (\vec{x}, \vec{y})$ for all $\vec{x}, \vec{y} \in \mathbb{C}^n$, so $\|U\vec{x}\| = \|\vec{x}\|$ for all $\vec{x} \in \mathbb{C}$.

Corollary 5.3.4. *If $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix} \in M_n(\mathbb{C})$ is a unitary matrix, then for all $j, k \in \{1, 2, \ldots, n\}$ we have*

$$(\vec{u}_j, \vec{u}_k) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Remark. The converse of Corollary 5.3.4 is also true and its proof is left as an exercise.

Example 5.3.3. $U_1 = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ and $U_2 = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}$
are unitary matrices.

Theorem 5.3.5. *Every eigenvalue of an unitary matrix U has absolute value one, i.e., $|\lambda| = 1$. Moreover, eigenvectors corresponding to different eigenvalues are orthogonal to each other.*

We are going to explore some very remarkable facts about Hermitian and real symmetric matrices. These matrices are diagonalizable, and moreover diagonalization can be accomplished by a unitary matrix P. This means that $P^{-1}AP = P^HAP$ is diagonal. In this situation, we say that the matrix A is **unitarily** or **orthogonally diagonalizable**. Orthogonally and unitary are particularly attractive since the calculation is essentially free and error-free as well: $P^H = P^{-1}$.

Theorem 5.3.6. *If a real matrix A is a orthogonally diagonalizable with an orthonormal matrix P, that is $P^T AP$ is a diagonal matrix, then A is symmetric.*

Remark. The converse of Theorem 5.3.6 is also true. In addition, we prove a stronger result.

Theorem 5.3.7. [Principal Axes Theorem] *Every Hermitian matrix is unitarily diagonalizable. In addition, every real symmetric matrix is orthogonally diagonalizable.*

Proof. We shall show this statement by induction on n. It is clear for $n = 1$.
Assume that $n > 1$ and every $(n-1)\times(n-1)$ Hermitian matrix is unitary diagonalizable. Consider an $n \times n$ Hermitian matrix A.
Let λ_1 be a real eigenvalue of A with unit eigenvector $\vec{v}$.
Then $A\vec{v} = \lambda_1\vec{v}$ and $\|\vec{v}\| = 1$.
Let $W = \{\vec{v}\}^\perp$ with orthonormal basis $\{\vec{z}_1, \dots, \vec{z}_{n-1}\}$.
Thus, $R = \begin{bmatrix}\vec{v} & \vec{z}_1 & \dots & \vec{z}_{n-1}\end{bmatrix}$ is an $n \times n$ unitary matrix. Observe that

$$B = R^H AR = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & b_{n2} & \dots & b_{nn} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & C_{(n-1)\times(n-1)} & \\ 0 & & & \end{bmatrix}$$

and $B^H = (R^H AR)^H = B$. Hence, B is Hermitian and so is C.
Since C is an $(n-1)\times(n-1)$ Hermitian matrix, by the induction hypothesis,
$\exists$ an $(n-1)\times(n-1)$ unitary matrix Q such that $Q^H CQ = \operatorname{diag}(\lambda_2, \dots, \lambda_n)$.
Let $P = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & Q & \\ 0 & & & \end{bmatrix}_{n\times n}$. Then P is an $n \times n$ unitary matrix and

$$P^H BP = P^H R^H ARP = (RP)^H A(RP).$$

Choose $U = RP$. Then $U^H = (RP)^H = P^H R^H = R^{-1}P^{-1} = (RP)^{-1} = U^{-1}$ and

$$U^H AU = P^H BP = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Hence, A is unitarily diagonalizable. □

Example 5.3.4. Diagonalize the Hermitian matrix $A = \begin{bmatrix} 1 & 1-i \\ 1+i & 0 \end{bmatrix}$.

Example 5.3.5. Orthogonally diagonalize the symmetric matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$.

(Given $\lambda = 0, 5, 5$).

Definition. A square matrix A is **normal** if $A^H A = AA^H$.

Clearly, every Hermitian matrix is normal.

Theorem 5.3.8. *A matrix is unitarily diagonalizable if and only if it is normal.*

Proof. It is a consequence of Schur Triangularization Theorem which is beyond the scope of this course. □

Real versus Complex

$(x_1, \ldots, x_n) \in \mathbb{R}^n$	$(x_1, \ldots, x_n) \in \mathbb{C}^n$
length: $\|\vec{x}\|^2 = x_1^2 + \cdots + x_n^2$	$\|\vec{x}\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2$
transpose: $A_{ij}^T = A_{ji}$	Hermitian: $A_{ij}^H = \overline{A}_{ji}$
$(AB)^T = B^T A^T$	$(AB)^H = B^H A^H$
$\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = x_1 y_1 + \cdots + x_n y_n$	$\vec{x} \cdot \vec{y} = \vec{x}^H \vec{y} = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n$
orthogonality: $\vec{x}^T \vec{y} = 0$	$\vec{x}^H \vec{y} = 0$
orthonormal: $P^T P = I_n = PP^T$	unitary: $U^H U = I_n = UU^H$
symmetric matrix: $A^T = A$	Hermitian matrix $A^H = A$
$A = PDP^{-1} = PDP^T$ (real D)	$A = UDU^{-1} = UDU^H$ (real D)
orthogonally diagonalizable	unitarily diagonalizable

5.4 Jordan Forms

Theorem 5.1.9 gives necessary and sufficient conditions for an $n \times n$ matrix to be diagonalizable, namely that it should have n independent eigenvectors. We have also seen square matrices which are *not* diagonalizable. In this section, we discuss the so-called **Jordan canonical form**, a form of matrix to which *every* square matrix is similar.

Definition. Let A be an $n \times n$ matrix. Let λ be an eigenvalue of A with $\operatorname{nullity}(A - \lambda I_n) = \ell$. Assume that λ is of multiplicity m. Then $1 \leq \ell \leq m$.
If $m = 1$, then $\ell = m = 1$.
If $m > 1$ and $\ell < m$, then λ is said to be **defective** and the number $m - \ell > 0$ of missing eigenvector(s) is called the **defect** of λ.

Note that if A has a defective eigenvalue, then A is *not* diagonalizable.

Definition. The **generalized eigenspace** G_λ **corresponding to an eigenvalue** λ **of** A, consists of all vectors $\vec{v}$ such that, for some $k \in \mathbb{N}$, $(A - \lambda I)^k \vec{v} = \vec{0}$, that is,

$$G_\lambda(A) = \{\vec{v} \in F^n : (A - \lambda I)^k \vec{v} = \vec{0} \text{ for some } k \in \mathbb{N}\} = \bigcup_{k \in \mathbb{N}} \operatorname{Nul}(A - \lambda I)^k.$$

Definition. A **length** r **chain of generalized eigenvectors based on the eigenvector** $\vec{v}$ **for** λ is a set $\{\vec{v} = \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ of r linearly independent generalized eigenvectors such that

$$\begin{aligned} (A - \lambda I)\vec{v}_r &= \vec{v}_{r-1}, \\ (A - \lambda I)\vec{v}_{r-1} &= \vec{v}_{r-2}, \\ &\vdots \\ (A - \lambda I)\vec{v}_2 &= \vec{v}_1. \end{aligned}$$

Since $\vec{v}_1$ is an eigenvector, $(A - \lambda I)\vec{v}_1 = \vec{0}$. It follows that

$$(A - \lambda I)^r \vec{v}_r = \vec{0}.$$

We may denote the action of the matrix $A - \lambda I$ on the string of vectors by

$$\vec{v}_r \longrightarrow \vec{v}_{r-1} \longrightarrow \cdots \longrightarrow \vec{v}_2 \longrightarrow \vec{v}_1 \longrightarrow \vec{0}.$$

Now let W be the subspace of G_λ spanned by $\{\vec{v}_1, \ldots, \vec{v}_r\}$. Any vector $\vec{x}$ in W has a representation of the form

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_r\vec{v}_r$$

and

$$\begin{aligned} A\vec{x} &= c_1(A\vec{v}_1) + c_2(A\vec{v}_2) + \cdots + c_r(A\vec{v}_r) \\ &= c_1(\lambda\vec{v}_1) + c_2(\lambda\vec{v}_2 + \vec{v}_1) + \cdots + c_r(\lambda\vec{v}_r + \vec{v}_{r-1}) \\ &= (\lambda c_1 + c_2)\vec{v}_1 + \cdots + (\lambda c_{r-1} + c_r)\vec{v}_{r-1} + \lambda c_r\vec{v}_r. \end{aligned}$$

Thus $A\vec{x}$ is also in W. If $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_r\}$ is a basis for W, then

$$[A\vec{x}]_{\mathcal{B}} = \begin{bmatrix} \lambda c_1 + c_2 \\ \lambda c_2 + c_3 \\ \vdots \\ \lambda c_{r-1} + c_r \\ \lambda c_r \end{bmatrix} = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \cdot & \cdot & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{r-1} \\ c_r \end{bmatrix} = J[\vec{x}]_{\mathcal{B}}$$

where

$$J = J(\lambda; r) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \cdot & \cdot & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}_{r\times r}$$

is called the **Jordan block of size r corresponding to λ**.

Example 5.4.1. Let $A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$. Find generalized eigenspaces of A.

Example 5.4.2. Let $A_1 = \begin{bmatrix} 0 & 1 & 2 \\ -5 & -3 & -7 \\ 1 & 0 & 0 \end{bmatrix}$ and $A_2 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$.
Then A_1 and A_2 have the same characteristic polynomial $(x+1)^3$. Find
(1) the minimal polynomials of A_1 and A_2, and
(2) the generalized eigenspaces of A_1 and A_2.

Theorem 5.4.1. *If a $n \times n$ matrix A has t linearly independent eigenvectors, then it is similar to a matrix J, that is, in Jordan form, with t square blocks on the diagonal:*

$$\textbf{Jordan form} \qquad J = M^{-1}AM = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_t \end{bmatrix}.$$

Each block has one eigenvector, one eigenvalue, and 1s just above the diagonal:

$$\textbf{Jordan block} \qquad J_i = J(\lambda_i, r_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda_i \end{bmatrix}_{r_i\times r_i}.$$

The same λ_i will appear in several blocks, if it has several independent eigenvectors. Moreover, M consists of n generalized eigenvectors which are linearly independent.

Remark. Theorem 5.4.1 says that every $n \times n$ matrix A has n linearly independent generalized eigenvectors. These n generalized eigenvectors may be arranged in chains, with the sum of the lengths of the chains associated with a given eigenvalue λ equal to the multiplicity of λ. But the structure of these chains depends on the defect of λ, and can be quite complicated. For instance, a multiplicity-four-eigenvalue can correspond to

- Four length 1 chain (defect 0);
- Two length 1 chains and a length 2 chain (defect 1);
- Two length 2 chains (defect 2);
- A length 1 chain and a length 3 chain (defect 2);
- A length 4 chain (defect 3).

Observe that, in each of these cases, the length of the longest chain is at most $d+1$ where d is the defect of the eigenvalue. Consequently, once we have found all the ordinary eigenvectors corresponding to a multiple eigenvalue λ, and therefore know the defect d of λ, we can begin with the equation

$$(A - \lambda I)^{d+1}\vec{u} = \vec{0} \tag{5.4.1}$$

to start building the chains of generalized eigenvectors corresponding to λ.

Algorithm: Begin with a nonzero solution $\vec{u}_1$ of Eq. (5.4.1) and successively multiply by the matrix $A - \lambda I$ until the zero vector is obtained. If

$$\begin{aligned}(A - \lambda I)\vec{u}_1 &= \vec{u}_2 \neq \vec{0}\\ (A - \lambda I)\vec{u}_2 &= \vec{u}_3 \neq \vec{0}\\ &\vdots\\ (A - \lambda I)\vec{u}_{k-1} &= \vec{u}_k \neq \vec{0}\end{aligned}$$

but $(A - \lambda I)\vec{u}_k = \vec{0}$, then we get the string of k generalized eigenvectors

$$\vec{u}_1 \longrightarrow \vec{u}_2 \longrightarrow \cdots \longrightarrow \vec{u}_k.$$

Example 5.4.3. Let $A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -2 & 2 & -3 & 1 \\ 2 & -2 & 1 & -3 \end{bmatrix}$ with the characteristic polynomial $x(x+2)^3$.
Find the chains of generalized eigenvectors corresponding to each eigenvalues and the Jordan form of A.

Example 5.4.4. Let $A = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 3 \\ 4 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$. Find the minimal polynomial of A and chain(s) of generalized eigenvectors and the Jordan form of A.

Example 5.4.5. Write down the Jordan form of the following matrices.

(1) $\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (2) $\begin{bmatrix} 3 & 5 & 0 & 0 \\ 0 & 3 & 6 & 0 \\ 0 & 0 & 4 & 7 \\ 0 & 0 & 0 & 4 \end{bmatrix}$ (3) $\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 5 & 0 \\ 0 & 0 & 4 & 6 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

Let $N(r) = J(0;r)$ denote an $r \times r$ matrix that has 1's immediately above the diagonal and zero elsewhere. For example,

$$N(2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, N(3) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, N(4) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ etc.}$$

Then $J(\lambda;r) = \lambda I + N(r)$, or in abbreviated $J = \lambda I + N$.

Suppose that $f(x)$ is a polynomial of degree s. Then the Taylor expansion around a point c from calculus gives us

$$f(c+x) = f(c) + f'(c)x + \frac{f''(c)}{2!}x^2 + \cdots + \frac{f^{(s)}(c)}{s!}x^s,$$

where $f', f'', \ldots, f^{(s)}$ represent successive derivatives of f. In terms of matrices I and N, we have

$$\begin{aligned} f(J) = f(\lambda I + N) &= f(\lambda I) + f'(\lambda I)N + \frac{f''(\lambda I)N^2}{2!} + \cdots + \frac{f^{(s)}(\lambda I)N^s}{s!} \\ &= f(\lambda)I + f'(\lambda)N + \frac{f''(\lambda)}{2!}N^2 + \cdots + \frac{f^{(s)}(\lambda)}{s!}N^s \\ &= \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdot & \cdot & \\ & f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & & \cdot \\ & & \cdot & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \frac{f''(\lambda)}{2!} \\ & & & & \cdot & f'(\lambda) \\ & & & & & f(\lambda) \end{bmatrix}_{r \times r} \end{aligned}$$

because the entries of N^k that are k steps above the diagonal are 1's and all the other entries are zeros.

Example 5.4.6. Compute $J(\lambda;4)^2$, $J(\lambda;3)^{10}$ and $J(\lambda;2)^s$.

Remark. If $J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_t \end{bmatrix}$ is in a Jordan form, then $J^s = \begin{bmatrix} J_1^s & & \\ & \ddots & \\ & & J_t^s \end{bmatrix}$.

Example 5.4.7. Compute J^s for $J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

Example 5.4.8. Given a square matrix A, use the Jordan form of A, to determine its minimal polynomial.

Solution. Let J be the Jordan form of A. Since $f(A) = Mf(J)M^{-1}$, $f(A) = \underline{\mathbf{0}}$ if and only if $f(J) = \underline{\mathbf{0}}$. Also, if $J(\lambda;r)$ is a Jordan block, then $f(J(\lambda;r))$ is a Jordan block of $f(J)$. We must thus find a polynomial such that, for every Jordan block $J(\lambda;r)$ of J, $f(J(\lambda;r)) = \underline{\mathbf{0}}$ holds.

But we derived a formula for $f(J(\lambda;r))$, and it equals the zero matrix if and only if $f(\lambda), f'(\lambda), \ldots, f^{(r-1)}(\lambda)$ are all zero. Thus, $f(x)$ and its first $r-1$ derivatives must vanish at $x = \lambda$; in other words, $(x-\lambda)^r$ must be a factor of $f(x)$.

Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of A and m_i the "maximum size" of the Jordan blocks corresponding to the eigenvalue λ_i. Hence, we obtain

$$f(x) = (x - \lambda_1)^{m_1} \ldots (x - \lambda_k)^{m_k}$$

is the minimal polynomial of A. □

Example 5.4.9. Find the minimal polynomial of the following matrices.

(1) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (2) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (3) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

(4) $\begin{bmatrix} 3 & 1 & & & & & \\ & 3 & 1 & & & & \\ & & 3 & & & & \\ & & & 3 & & & \\ & & & & 8 & 1 & \\ & & & & & 8 & \\ & & & & & & 5 \end{bmatrix}$ (5) $\begin{bmatrix} 3 & 1 & & & & & \\ & 3 & & & & & \\ & & 3 & 1 & & & \\ & & & 3 & & & \\ & & & & 8 & 1 & \\ & & & & & 8 & \\ & & & & & & 5 \end{bmatrix}$

Exercises for Chapter 5.

1. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & b & c \end{bmatrix}$. Find a, b, c so that $\det(A - \lambda I_3) = 9\lambda - \lambda^3$.
2. Let $T : V \to V$ be a linear operator.
   A subspace U of V is T-**invariant** if $T(U) \subseteq U$, i.e., $\forall \vec{u} \in U, T(\vec{u}) \in U$.
   (a) Show that $\ker T$ and $\operatorname{im} T$ are T-invariant.
   (b) If U and W are T-invariant, prove that $U \cap W$ and $U + W$ are also T-invariant.
   (c) Show that the eigenspace $E_\lambda(T)$ is T-invariant.
3. Show that A and A^T have the same eigenvalues.
4. Show that if $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A, then $\lambda_1^m, \ldots, \lambda_k^m$ are eigenvalues of A^m for all $m \geq 1$. Moreover, each eigenvector of A is an eigenvector of A^m.
5. Let A and B be $n \times n$ matrices over a field F. If $I - AB$ is invertible, prove that $I - BA$ is invertible and $(I - BA)^{-1} = I + B(I - AB)^{-1}A$.
6. Show that if A and B are the same size, then AB and BA have the same eigenvalues.
7. Determine all 2×2 diagonalizable matrices A with nonzero repeated eigenvalue a, a.
8. Let V be the space of all real-valued continuous functions. Define $T : V \to V$ by

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

   Show that T has no eigenvalues.
9. Prove that if A is invertible and diagonalizable, then A^{-1} is diagonalizable.
10. Let $V = \operatorname{Span}\{1, \sin 2t, \sin^2 t\}$. Let $T : V \to V$ defined by $T(f) = f''$.
    Find all eigenvalues and eigenspaces of D. Is T diagonalizable? Explain.
11. Let $A = [a_{ij}]$ be an $n \times n$ matrix such that for each $i = 1, 2, \ldots, n$, we have $\sum_{j=1}^{n} a_{ij} = 0$.
    Show that 0 is an eigenvalue of A.
12. Let A be an $n \times n$ matrix with characteristic polynomial $(x - \lambda_1)^{d_1} \ldots (x - \lambda_k)^{d_k}$.
    Show that $\operatorname{tr} A = d_1\lambda_1 + \cdots + d_k\lambda_k$.
13. Let A be a 2×2 matrix. Prove that the characteristic polynomial of A is given by

$$x^2 - (\operatorname{tr} A)x + \det A = 0.$$

14. If A and B are 2×2 matrices with determinant one, prove that

$$\operatorname{tr} AB - (\operatorname{tr} A)(\operatorname{tr} B) + \operatorname{tr} AB^{-1} = 0.$$

15. Find the 2×2 matrices with real entries that satisfy the equation

$$X^3 - 3X^2 = \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix}.$$

(*Hint*. Apply the Cayley-Hamilton Theorem.)

16. Let $A = \begin{bmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{bmatrix}$.
Prove that the minimal polynomial of A and the characteristic polynomial of A are the same.
17. A 3×3 matrix A has the characteristic polynomial $x(x-1)(x+2)$.
What is the characteristic polynomial of A^2?
18. Let $V = M_n(F)$ be the vector space of $n \times n$ matrices over a field F. Let A be an $n \times n$ matrix.
Let T_A be the linear operator on V defined by $T_A(B) = AB$.
Show that the minimal polynomial for T_A is the minimal polynomial for A.
19. Let U be an $n \times n$ real orthonormal matrix. Prove that
(a) $|\mathrm{tr}\,(U)| \leq n$, and (b) $\det(U^2 - I_n) = 0$ if n is odd.
20. If $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_n \end{bmatrix}$ with $(\vec{u}_j, \vec{u}_k) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}$, prove that U is unitary.
21. Let A be an $n \times n$ symmetric matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Prove that

$$(A - \lambda_1 I_n) \dots (A - \lambda_k I_n) = \underline{0}.$$

22. Unitarily diagonalize the following matrices.
(a) $\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & i \\ -i & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 2 & i & i \\ -i & 1 & 0 \\ -i & 0 & 1 \end{bmatrix}$
23. Show that every unitarily diagonalizable matrix is normal.
24. Suppose that A is real symmetric and orthonormal. Prove that the only possible eigenvalues of A are ± 1.
25. Show that if a real matrix A is skew-symmetric (i.e., $A^T = -A$), then iA is Hermitian.
26. Prove that if A is unitarily diagonalizable, then so is A^H.
27. Let A be any square real matrix. Show that the eigenvalues of $A^T A$ are all non-negative.
28. Show that the generalized eigenspace G_λ corresponding to an eigenvalue λ of an $n \times n$ matrix A is a subspace of F^n.
29. Suppose the characteristic polynomial of a 4×4 matrix A is $(x-1)^2(x+1)^2$.
(a) Prove that $A^{-1} = 2A - A^3$. (b) Write down all possible Jordan form(s) of A.
30. Let $J = J(\lambda; r)$ be an $r \times r$ Jordan block with λ on its diagonal. Show that J has only one linearly independent eigenvector corresponding to λ.
31. If J is in Jordan form with k Jordan blocks on the diagonal, prove that J has exactly k linearly independent eigenvectors.
32. These Jordan matrices have eigenvalues $0, 0, 0, 0$:

$$J = \begin{bmatrix} 0 & 1 & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix} \quad \text{and} \quad K = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & \\ & & & 0 \end{bmatrix}.$$

For any matrix M, compare JM with MK. If they equal, show that M is not invertible. Then J and K are not similar.
33. Suppose that a square matrix has two eigenvalues $\lambda = 2, 5$, and $n_p(\lambda) = \mathrm{nullity}(A - \lambda I)^p$, $p \in \mathbb{N}$, are as follows:
$n_1(2) = 2$, $n_2(2) = 4$, $n_p(2) = 5$ for $p \geq 3$, and $n_1(5) = 1$, $n_p(5) = 2$ for $p \geq 2$.
Write down the Jordan form of A.
34. If $J = J(0; 5)$ is the 5×5 Jordan block with $\lambda = 0$. Find J^2, count its eigenvectors and write its Jordan form.
35. How many possible Jordan forms are there for a 6×6 matrix with characteristic polynomial $(x-1)^2(x+2)^4$?

36. Let $A = \begin{bmatrix} 2 & a & b \\ 0 & 2 & c \\ 0 & 0 & 1 \end{bmatrix} \in M_3(\mathbb{R})$.
(a) Prove that A is diagonalizable if and only if $a = 0$.
(b) Find the minimal polynomial of A when (i) $a = 0$ (ii) $a \neq 0$.

37. Let $V = \{h(x, y) = ax^2 + bxy + cy^2 + dx + ey + f : a, b, c, d, e, f \in \mathbb{R}\}$ be a subspace of the space of polynomial in two variables x and y over $\mathbb{R}$. Then $\mathcal{B} = \{x^2, xy, y^2, x, y, 1\}$ is a basis for V. Define $T : V \to V$ by

$$(T(h))(x, y) = \frac{\partial}{\partial y}\left(\int h(x, y)\, dx\right).$$

(a) Prove that T is a linear transformation and find $A = [T]_{\mathcal{B}}$.
(b) Compute the characteristic polynomial and the minimal polynomial of A.
(c) Find the Jordan form of A.

38. True or False:
(a) $\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$ and $\begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$ are similar. (b) $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ and $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ are similar.

39. Show that $\begin{bmatrix} a & 1 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$ and $\begin{bmatrix} b & 0 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$ are similar.

40. Write down the Jordan form for the following matrices and find its minimal polynomial.
(a) $\begin{bmatrix} -2 & 1 \\ -1 & -4 \end{bmatrix}$ (b) $\begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 \\ -2 & -2 & -3 \\ 2 & 3 & 4 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 0 & 0 \\ -7 & 9 & 7 \\ 0 & 0 & 2 \end{bmatrix}$
(e) $\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} -2 & 17 & 4 \\ -1 & 6 & 1 \\ 0 & 1 & 2 \end{bmatrix}$ (g) $\begin{bmatrix} -3 & 5 & -5 \\ 3 & -1 & 3 \\ 8 & -8 & 10 \end{bmatrix}$ (h) $\begin{bmatrix} 5 & -1 & 1 \\ 1 & 3 & 0 \\ -3 & 2 & 1 \end{bmatrix}$
(i) $\begin{bmatrix} 1 & -4 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 6 & -12 & -1 & -6 \\ 0 & -4 & 0 & -1 \end{bmatrix}$ (j) $\begin{bmatrix} 2 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ (k) $\begin{bmatrix} -1 & -4 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ (l) $\begin{bmatrix} 1 & 3 & 7 & 0 \\ 0 & -1 & -4 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & -6 & -14 & 1 \end{bmatrix}$

Eigenvalues: (b) $-1, -1, -1$ (c) $1, 1, 1$ (d) $2, 2, 9$ (e) $1, 2, 2$ (f) $2, 2, 2$ (g) $2, 2, 2$ (h) $3, 3, 3$ (i) $-1, -1, 1, 1$ (k) $1, 1, 1, 1$ (l) $1, 1, 1, 1$.