MATH 146 Linear Algebra 1, Lecture Notes

by Stephen New

Chapter 1. Concrete Vector Spaces and Affine Spaces

Rings and Fields

1.1 Definition: For a set S we write $S \times S = \{(a, b) | a \in S, b \in S\}$. A **binary operation** on S is a map $* : S \times S \to S$, where for $a, b \in S$ we usually write $*(a, b)$ as $a * b$.

1.2 Definition: A **ring** (with identity) is a set R together with two binary operations $+$ and $\cdot$ (called addition and multiplication), where for $a, b \in R$ we usually write $a \cdot b$ as ab, and two distinct elements 0 and 1, such that

(1) $+$ is associative: $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$,
(2) $+$ is commutative: $a + b = b + a$ for all $a, b \in R$,
(3) 0 is an additive identity: $0 + a = a$ for all $a \in R$,
(4) every element has an additive inverse: for every $a \in R$ there exists $b \in R$ with $a+b = 0$,
(5) $\cdot$ is associative: $(ab)c = a(bc)$ for all $a, b, c \in R$,
(6) 1 is a multiplicative identity: $1 \cdot a = a$ for all $a \in R$, and
(7) $\cdot$ is distributive over $+$: $a(b + c) = ab + ac$ for all $a, b, c \in R$,

A ring R is called **commutative** when
(8) $\cdot$ is commutative: $ab = ba$ for all $a, b \in R$.

For $a \in R$, we say that a is **invertible** (or that a has an **inverse**) when there exists an element $b \in R$ such that $ab = 1 = ba$. A **field** is a commutative ring F such that
(9) every non-zero element has a multiplicative inverse: for every $a \in F$ with $a \neq 0$ there exists $b \in F$ such that $ab = 1$.

An element in a field F is called a **number** or a **scalar**.

1.3 Example: The set of **integers Z** is a commutative ring, but it is not a field because it does not satisfy Property (9). The set of **positive integers** $\mathbf{Z}^+ = \{1, 2, 3, \cdots\}$ is not a ring because $0 \notin \mathbf{Z}^+$ and $\mathbf{Z}^+$ does not satisfy Properties (3) and (4). The set of **natural numbers** $\mathbf{N} = \{0, 1, 2, \cdots\}$ is not a ring because it does not satisfy Property (4). The set of **rational numbers Q**, the set of **real numbers R** and the set of **complex numbers C** are all fields. For $2 \leq n \in \mathbf{Z}$, the set $\mathbf{Z}_n = \{0, 1, \cdots, n - 1\}$ of **integers modulo** n is a commutative ring, and $\mathbf{Z}_n$ is a field if and only if n is prime (in $\mathbf{Z}_1 = \{0\}$ we have $0 = 1$, so $\mathbf{Z}_1$ is not a ring).

1.4 Remark: In a field, we can perform all of the usual arithmetical operations. The next few theorems illustrate this.

1.5 Theorem: *(Uniqueness of Inverse) Let R be a field. Let $a \in R$. Then*
(1) the additive inverse of a is unique: if $a + b = 0 = a + c$ then $b = c$,
(2) if a has an inverse then it is unique: if $ab = 1 = ac$ then $b = c$.

Proof: To prove (1), suppose that $a + b = 0 = a + c$. Then

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

To prove (2), suppose that $a \neq 0$ and that $ab = 1 = ac$. Then

$$b = 1 \cdot b = (ac)b = b(ac) = (ba)c = (ab)c = 1 \cdot c = c\,.$$

1.6 Definition: Let R be a ring and let $a, b \in R$. We write the (unique) additive inverse of a as $-a$, and we write $b - a = b + (-a)$. If a has a multiplicative inverse, we write the (unique) multiplicative inverse of a as a^{-1}. When R is commutative, we also write a^{-1} as $\frac{1}{a}$, and we write $\frac{b}{a} = b \cdot \frac{1}{a}$.

1.7 Theorem: *(Cancellation) Let R be a field. Then for all $a, b, c \in R$, we have*

(1) if $a + b = a + c$ then $b = c$,
(2) if $a + b = a$ then $b = 0$, and
(3) if $a + b = 0$ then $b = -a$.

Let F be a field. Then for all $a, b, c \in F$ we have

(4) if $ab = ac$ then either $a = 0$ or $b = c$.
(5) if $ab = a$ then either $a = 0$ or $b = 1$,
(6) if $ab = 1$ then $b = a^{-1}$, and
(7) if $ab = 0$ then either $a = 0$ or $b = 0$.

Proof: To prove (1), suppose that $a + b = a + c$. Then we have

$$b = 0 + b = -a + a + b = -a + a + c = 0 + c = c\,.$$

Part (2) follows from part (1) since if $a + b = a$ then $a + b = a + 0$, and part (3) follows from part (1) since if $a + b = 0$ then $a + b = a + (-a)$. To prove part (4), suppose that $ab = ac$ and $a \neq 0$. Then we have

$$b = 1 \cdot b = a^{-1}ab = a^{-1}ac = 1 \cdot c = c\,.$$

Note that parts (5), (6) and (7) all follow from part (4).

1.8 Remark: In the above proof, we used associativity and commutativity implicitly. If we wished to be explicit then the proof of part (1) would be as follows. Suppose that $a + b = a + c$. Then we have

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

In the future, we shall often use associativity and commutativity implicitly in our calculations.

1.9 Theorem: *(Multiplication by 0 and -1) Let R be a ring and let $a \in R$. Then*

(1) $0 \cdot a = 0$, and
(2) $(-1)a = -a$.

Proof: We have

$$0a = (0 + 0)a = 0a + 0a.$$

Subtracting $0a$ from both sides (using part 2 of the Cancellation Theorem) gives $0 = 0a$. Also, we have

$$a + (-1)a = (1)a + (-1)a = (1 + (-1))a = 0a = 0,$$

and subtracting a from both sides (part 3 of the Cancellation Theorem) gives $(-1)a = -a$.

The Standard Vector Space

1.10 Definition: Let S be a set. An **n-tuple** on S is a function $a : \{1, 2, \cdots, n\} \to S$. Given an n-tuple a on S, for $k \in \{1, 2, \cdots, n\}$ we write $a_k = a(k)$. The set $\{1, 2, \cdots, n\}$ is called the **index set**, an element $k \in \{1, 2, \cdots, n\}$ is called an **index**, and the element $a_k \in S$ is called the k^{th} **entry** of a. We sometimes write $a = (a_1, a_2, \cdots, a_n)$ but we more often write

$$a = (a_1, a_2, \cdots, a_n)^T = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

to indicate that a is the n-tuple with entries $a_1, a_2, \cdots, a_n$. The set of all n-tuples on S is denoted by S^n, so we have

$$S^n = \Big\{ a = (a_1, a_2, \cdots, a_n)^T \Big| \text{ each } a_i \in S \Big\}.$$

1.11 Definition: For a ring R, we define the **zero element** $0 \in R^n$ to be

$$0 = (0, 0, 0, \cdots, 0)^T$$

or equivalently we define $0 \in R^n$ to be the element with entries $0_i = 0$ for all i. We define the **standard basis elements** $e_1, e_2, \cdots, e_n \in R^n$ to be

$$\begin{aligned} e_1 &= (1, 0, 0, 0 \cdots, 0)^T, \\ e_2 &= (0, 1, 0, 0, \cdots, 0^T, \\ e_3 &= (0, 0, 1, 0, \cdots, 0)^T, \\ &\vdots \\ e_n &= (0, 0, 0, \cdots, 0, 1)^T. \end{aligned}$$

Equivalently, for each index k we define $e_k \in R^n$ to be given by $(e_k)_i = \delta_{ki} = \begin{cases} 1 \text{ if } k = i, \\ 0 \text{ if } k \neq i. \end{cases}$

1.12 Definition: Given $t \in R$, $x = (x_1, x_2, \cdots, x_n)^T \in R^n$ and $y = (y_1, y_2, \cdots, y_n)^T \in R^n$, where R is a ring, we define the product tx and the sum $x + y$ by

$$tx = t \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} t\,x_1 \\ t\,x_2 \\ \vdots \\ t\,x_n \end{pmatrix},$$

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

Equivalently, we can define tx to be the element with entries $(tx)_i = t\,x_i$ for all i, and we can define $x + y$ to be the element with entries $(x + y)_i = x_i + y_i$ for all i.

1.13 Note: For $x_1, x_2, \cdots, x_n \in R$, notice that

$$(x_1, x_2, \cdots, x_n)^T = \sum_{i=1}^{n} x_i e_i = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.$$

1.14 Theorem: *(Basic Properties of R^n) Let R be a ring. Then*
(1) $+$ is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in R^n$,
(2) $+$ is commutative: $x + y = y + x$ for all $x, y \in R^n$,
(3) $0 \in R^n$ is an additive identity $0 + x = x$ for all $x \in R^n$,
(4) every $x \in R^n$ has an additive inverse: for all $x \in R^n$ there exists $y \in R^n$ with $x + y = 0$,
(5) $\cdot$ is associative: $(st)x = s(tx)$ for all $s, t \in R$ and all $x \in R^n$,
(7) $\cdot$ distributes over addition in R: $(s + t)x = sx + tx$, for all $s, t \in R$ and $x \in R^n$,
(8) $\cdot$ distributes over addition in R^n: $t(x + y) = tx + ty$ for all $t \in R$ and $x, y \in R^n$, and
(9) $1 \in R$ acts a multiplicative identity: $1x = x$ for all $x \in R^n$.

Proof: To prove part (4), let $x \in R^n$ and choose $y = (-1)x$. Then for all indices i we have $y_i = ((-1)x)_i = -x_i$ and so $(x + y)_i = x_i - x_i = 0$. Since $(x + y)_i = 0$ for all i, we have $x + y = 0$, as required. To prove part (8), let $t \in R$ and let $x, y \in R^n$. Then for all i we have $\big(t(x + y)\big)_i = t(x + y)_i = t(x_i + y_i) = t\,x_i + t\,y_i = (tx + ty)_i$. Since $\big(t(x + y)\big)_i = (tx + ty)_i$ for all i, we have $t(x + y) = tx + ty$. The other parts can be proven similarly.

1.15 Definition: When F is a field, F^n is called the **standard n-dimensional vector space** over F, and an element of F^n is called a **point** or a **vector**.

1.16 Example: Let $n = 2$ or 3 and let $u, v \in \mathbf{R}^n$. If $u \neq 0$ then the set $\{tu | t \in \mathbf{R}\}$ is the line in $\mathbf{R}^n$ through the points 0 and u, and the set $\{tu | 0 \leq t \leq 1\}$ is the line segment in $\mathbf{R}^2$ between 0 and u. If $u \neq 0$ and v does not lie on the line through 0 and u, then the points 0, u, v and $u + v$ are the veritices of a parallelogram P in $\mathbf{R}^n$, the set $\{su + tv | s \in \mathbf{R}, t \in \mathbf{R}\}$ is the plane which contains P (in that case that $n = 2$, this plane is the entire set $\mathbf{R}^2$), the set $\{su + tv | 0 \leq s \leq 1, 0 \leq t \leq 1\}$ is the set of points inside (and on the edges of) P. As an exercise, describe the sets $\{su + tv | s + t = 1\}$ and $\{su + tv | s \geq 0, t \geq 0, s + t \leq 1\}$.

Vector Spaces and Affine Spaces in F^n

1.17 Definition: Let F be a field. Given a point $p \in F^n$ and a non-zero vector $u \in F^n$, we define the **line** in F^n through p in the direction of u to be the set

$$L = \big\{p + tu \big| t \in F\big\}.$$

Given a point $p \in \mathbf{F}^n$ and two vectors $u, v \in \mathbf{R}^n$ with $u \neq 0$ and $v \neq tu$ for any $t \in F$, we define the **plane** in F^n through p in the direction of u and v to be the set

$$P = \big\{p + su + tv \,\big|\, s, t \in F\big\}.$$

1.18 Remark: We wish to generalize the above definitions by defining higher dimensional versions of lines and planes.

1.19 Note: For a finite set S, the **cardinality** of S, denoted by $|S|$, is the number of elements in S. When we write $S = \{a_1, a_2, \cdots, a_m\}$, we shall always tacitly assume that the elements $a_i \in S$ are all distinct so that $|S| = m$ unless we explicitly indicate otherwise.

1.20 Definition: Let R be a ring, let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\} \subseteq R^n$. A **linear combination** on $\mathcal{A}$ (over R) is an element $x \in R^n$ of the form

$$x = \sum_{i=1}^{m} t_i u_i = t_1 u_1 + t_2 u_2 + \cdots + t_m u_m \quad \text{with each } t_i \in R.$$

The **span** of $\mathcal{A}$ (over R) (also called the **submodule** of R^n **spanned** by $\mathcal{A}$ over R), is the set of all linear combinations on $\mathcal{A}$. We denote the span of $\mathcal{A}$ by $\operatorname{Span}\mathcal{A}$ (or by $\operatorname{Span}_R\mathcal{A}$) so we have

$$\operatorname{Span}\mathcal{A} = \operatorname{Span}_R\mathcal{A} = \Big\{ \sum_{i=1}^{m} t_i u_i \Big| \text{ each } t_i \in R\Big\}.$$

For convenience, we also define $\operatorname{Span}\emptyset = \{0\}$, where $\emptyset$ is the empty set. Given an element $p \in R^n$, we write

$$p + \operatorname{Span}\mathcal{A} = \Big\{p + u \Big| u \in \operatorname{Span}\mathcal{A}\Big\} = \Big\{p + \sum_{i=1}^{m} t_i u_i \Big| \text{each } t_i \in R\Big\}.$$

1.21 Definition: Let F be a field. For a finite set $\mathcal{A} \subseteq F^n$ the set $U = \operatorname{Span}\mathcal{A}$ is called the **vector space** in F^n (or the **subspace** of F^n) spanned by $\mathcal{A}$ (over F). A **vector space** in F^n (or a **subspace** of F^n) is a subset $U \subseteq F^n$ of the form $U = \operatorname{Span}\mathcal{A}$ for some finite subset $\mathcal{A} \subseteq F^n$. Given a point $p \in F^n$ and a finite set $\mathcal{A} \subseteq F^n$, the set $P = p + \operatorname{Span}\mathcal{A}$ is called the **affine space** in F^n (or the **affine subspace** of F^n) through p in the direction of the vectors in $\mathcal{A}$. An **affine space** in F^n (or an **affine subspace** of F^n) is a subset $P \subseteq F^n$ of the form $P = p + U$ for some point $p \in F^n$ and some vector space U in F^n. An element in a subspace of F^n can be called a **point** or a **vector**. An element in an affine subspace of F^n is usually called a **point**.

1.22 Theorem: *(Closure under Addition and Multiplication) Let R be a ring, let $\mathcal{A}$ be a finite subset of R, and let $U = \operatorname{Span}\mathcal{A}$. Then*

(1) U is closed under addition: for all $x, y \in U$ we have $x + y \in U$, and
(2) U is closed under multiplication: for all $t \in R$ and all $x \in U$ we have $tx \in U$.

Proof: Let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\}$, let $t \in R$ and let $x, y \in U = \operatorname{Span}\mathcal{A}$, say $x = \sum_{i=1}^{m} s_i u_i$ and $y = \sum_{i=1}^{m} t_i u_i$. Then $x + y = \sum_{i=1}^{m} (s_i + t_i) u_i \in U$ and $tx = \sum_{i=1}^{m} (t s_i) u_i \in U$.

1.23 Theorem: *Let R be a ring, let $p, q \in R^n$, let $\mathcal{A}$ and $\mathcal{B}$ be finite subsets of R^n, and let $U = \operatorname{Span}\mathcal{A}$ and $V = \operatorname{Span}\mathcal{B}$. Then*

(1) $p + U \subseteq q + V$ if and only if $U \subseteq V$ and $p - q \in V$, and
(2) $p + U = q + V$ if and only if $U = V$ and $p - q \in U$.

Proof: Suppose that $p + U \subseteq q + V$. Since $p = p + 0 \in p + U$, we also have $p \in q + V$, say $p = q + v$ where $v \in V$. Then $p - q = v \in V$. Let $u \in U$. Then we have $p + u \in p + U$ and so $p + u \in q + V$, say $p + u = q + w$ where $w \in V$. Then $u = w - (p - q) = w - v = w + (-1)v \in V$ by closure. Conversely, suppose that $U \subseteq V$ and $p - q \in V$, say $p - q = v \in V$. Let $a \in p + U$, say $a = p + u$ where $u \in U$. Then we have $a = p + u = (q + v) + u = q + (u + v) \in q + V$ by closure, since $u, v \in V$. This proves Part (1), from which Part (2) immediately follows.

1.24 Theorem: *Let $\mathcal{A} = \{u_1, u_2, \cdots, u_l\} \subseteq R^n$ and let $\mathcal{B} = \{v_1, v_2, \cdots, v_m\} \subseteq R^n$, where R is a ring. Then*

(1) $\operatorname{Span}\mathcal{A} \subseteq \operatorname{Span}\mathcal{B}$ *if and only if each* $u_j \in \operatorname{Span}\mathcal{B}$, *and*
(2) $\operatorname{Span}\mathcal{A} = \operatorname{Span}\mathcal{B}$ *if and only if each* $u_j \in \operatorname{Span}\mathcal{B}$ *and each* $v_j \in \operatorname{Span}\mathcal{A}$.

Proof: Note that each $u_j \in \operatorname{Span}\mathcal{A}$ because we can write u_j as a linear combination on $\mathcal{A}$, indeed we have

$$u_j = 0u_1 + 0u_2 + \cdots + 0u_{j-1} + 1u_j + 0u_{j+1} + \cdots + 0u_l = \sum_{i=1}^{l} t_i u_i \text{ with } t_i = \delta_{ij}.$$

It follows that if $\operatorname{Span}\mathcal{A} \subseteq \operatorname{Span}\mathcal{B}$ then we have each $u_j \in \operatorname{Span}\mathcal{B}$. Suppose, conversely, that each $u_j \in \operatorname{Span}\mathcal{B}$, say $u_j = \sum_{i=1}^{m} s_{ji} v_i$. Let $x \in \operatorname{Span}\mathcal{A}$, say $x = \sum_{j=1}^{l} t_j u_j$. Then

$$x = \sum_{j=1}^{l} t_j u_j = \sum_{j=1}^{l} t_j \sum_{i=1}^{m} s_{ji} v_i = \sum_{i=1}^{m} \Big(\sum_{j=1}^{l} t_j s_{ji} \Big) v_i \in \operatorname{Span}\mathcal{B}.$$

This Proves part (1), and Part (2) follows immediately from part (1).

Linear Independence, Bases and Dimension

1.25 Definition: Let R be a ring. For $\mathcal{A} = \{u_1, u_2, \cdots, u_m\} \subseteq R^n$, we say that $\mathcal{A}$ is **linearly independent** (over R) when for all $t_1, t_2, \cdots, t_m \in R$, if $\sum_{i=1}^{m} t_i u_i = 0$ then each $t_i = 0$, and otherwise we say that $\mathcal{A}$ is **linearly dependent**. For convenience, we also say that the empty set $\emptyset$ is linearly independent. For a finite set $\mathcal{A} \subseteq F^n$, when $\mathcal{A}$ is linearly independent and $U = \operatorname{Span} \mathcal{A}$, we say that $\mathcal{A}$ is a **basis** for U.

1.26 Example: Let F be a field. The empty set $\emptyset$ is linearly independent and $\operatorname{Span} \emptyset = \{0\}$ and so $\emptyset$ is a basis for the vector space $\{0\}$ in F^n. If $0 \neq u \in F^n$ then $\{u\}$ is linearly independent and so $\{u\}$ is a basis for $\operatorname{Span}\{u\}$. As an exercise, verify that for $u, v \in F^n$, the set $\{u, v\}$ is linearly independent if and only if $u \neq 0$ and for all $t \in F$ we have $v \neq tu$.

1.27 Example: Verify that the set $\{e_1, e_2, \cdots, e_n\}$ is a basis for F^n. We call it the **standard basis** for F^n.

1.28 Theorem: *Let F be a field and let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\} \subseteq F^n$. Then*
(1) for $1 \leq k \leq m$, we have $u_k \in \operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big)$ if and only if $\operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big) = \operatorname{Span} \mathcal{A}$,
(2) $\mathcal{A}$ is linearly dependent if and only if $u_k \in \operatorname{Span}\big(\mathcal{A} \setminus \{u_m\}\big)$ for some index k.

Proof: Note that if $\operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big) = \operatorname{Span} \mathcal{A}$ then $u_k \in \operatorname{Span} \mathcal{A} = \operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big)$. Suppose, conversely, that $u_k \in \operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big)$, say $u_k = \sum_{i \neq k} s_i u_i$ where each $s_i \in F$. Since $\mathcal{A} \setminus \{u_k\} \subseteq \mathcal{A}$ it is clear that $\operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big) \subseteq \operatorname{Span} \mathcal{A}$. Let $x \in \operatorname{Span} \mathcal{A}$, say $x = \sum_{i=1}^{m} t_i u_i$. Then we have $x = t_k u_k + \sum_{i \neq k} t_i u_i = t_k \sum_{i \neq k} s_i u_i + \sum_{i \neq k} t_i u_i \in \operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big)$. This proves Part (1).

Note that since $\mathcal{A} \setminus \{u_k\} \subseteq \mathcal{A}$, we have $\operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big) \subseteq \operatorname{Span} \mathcal{A}$. Suppose that $\mathcal{A}$ is linearly dependent. Choose coefficients $s_i \in F$, not all equal to zero, so that $\sum_{i=1}^{m} s_i u_i = 0$. Choose an index k so that $s_k \neq 0$. Since $0 = s_k u_k + \sum_{i \neq k} s_i u_i$ we have $u_k = -\sum_{i \neq k} \frac{s_i}{s_k} u_i$. For $x = \sum_{i=1}^{m} t_i u_i \in \operatorname{Span} \mathcal{A}$ we have $x = t_k u_k + \sum_{i \neq k} t_i u_i = -t_k \sum_{i \neq k} \frac{s_i}{s_k} u_i + \sum_{i \neq k} t_i u_i \in \operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big)$. This shows that if $\mathcal{A}$ is linearly dependent then $\operatorname{Span}\big(\mathcal{A} \setminus \{u_k\}\big) = \operatorname{Span} \mathcal{A}$.

1.29 Theorem: *Let F be a field, let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\} \subseteq F^n$, and let $U = \operatorname{Span} \mathcal{A}$. Then $\mathcal{A}$ contains a basis for U.*

Proof: If $\mathcal{A}$ is linearly independent, then $\mathcal{A}$ is a basis for U. Suppose that $\mathcal{A}$ is linearly dependent. Then for some index k we have $u_k \in \operatorname{Span}(\mathcal{A} \setminus \{u_k\})$. Reordering the vectors if necessary, let us assume that $u_m \in \operatorname{Span}(\mathcal{A} \setminus \{u_m\})$. Then we have $\operatorname{Span}\{u_1, u_2, \cdots, u_{m-1}\} = \operatorname{Span}\{u_1, u_2, \cdots, u_m\} = U$. If $\{u_1, u_2, \cdots, u_{m-1}\}$ is linearly independent then it is a basis for U. Otherwise, as above, we can reorder $u_1, u_2, \cdots, u_{m-1}$ if necessary so that $\operatorname{Span}\{u_1, u_2, \cdots, u_{m-2}\} = \operatorname{Span}\{u_1, u_2, \cdots, u_{m-1}\} = U$. Repeating this procedure we will eventually obtain a linearly independent subset $\{u_1, u_2, \cdots, u_k\} \subseteq \mathcal{A}$ with $\operatorname{Span}\{u_1, u_2, \cdots, u_k\} = U$ (if the procedure continues until no vectors are left then we have $k = 0$ and $\{u_1, \cdots, u_k\} = \emptyset$, which is linearly independent).

1.30 Corollary: *For a field F, every subspace of F^n has a basis.*

1.31 Theorem: *Let F be a field, let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\} \subseteq F^n$, let $a_1, a_2, \cdots, a_m \in F$ with $a_k \neq 0$, and let $\mathcal{B} = \{v_1, v_2, \cdots, v_m\}$ where $v_i = u_i$ for $i \neq k$ and $v_k = \sum_{i=1}^{m} a_i u_i$. Then*
(1) $\operatorname{Span}\mathcal{A} = \operatorname{Span}\mathcal{B}$ and
(2) $\mathcal{A}$ is linearly independent if and only if $\mathcal{B}$ is linearly independent.

Proof: For $x = \sum_{i=1}^{m} t_i v_i \in \operatorname{Span}\mathcal{B}$, we have $x = t_k v_k + \sum_{i\neq k} t_i v_i = t_k \sum_{i=1}^{m} a_i u_i + \sum_{i\neq k} t_i u_i$ and so $x \in \operatorname{Span}\mathcal{A}$. This shows that $\operatorname{Span}\mathcal{B} \subseteq \operatorname{Span}\mathcal{A}$.

Suppose $\mathcal{A}$ is linearly independent. Suppose $\sum_{i=1}^{m} t_i v_i = 0$ where each $t_i \in F$. Then

$$0 = t_k v_k + \sum_{i\neq k} t_i v_i = t_k \sum_{i=1}^{m} a_i u_i + \sum_{i\neq k} t_i u_i = t_k a_k u_k + \sum_{i\neq k} (t_k a_i + t_i) u_i.$$

Since $\mathcal{A}$ is linearly independent, all of the coefficients in the above linear combination on $\mathcal{A}$ must be equal to zero, so we have $t_k a_k = 0$ and $t_k a_i + t_i = 0$ for $i \neq k$. Since $t_k a_k = 0$ and $a_k \neq 0$ we have $t_k = 0$ and hence $0 = t_k a_i + t_i = t_i$ for all $i \neq k$. This shows that $\mathcal{B}$ is linearly independent.

Finally note that since $v_k = \sum_{i=1}^{m} a_i u_i = a_k u_k + \sum_{i\neq k} a_i v_i$ with $a_k \neq 0$, it follows that $u_k = \sum_{i=1}^{m} b_i v_i$ where $b_k = \frac{1}{a_k} \neq 0$ and $b_i = -\frac{a_i}{a_k}$ for $i \neq k$. Hence the same arguments used in the previous two paragraphs, with the rôles of $\mathcal{A}$ and $\mathcal{B}$ interchanged, show that $\operatorname{Span}\mathcal{A} \subseteq \operatorname{Span}\mathcal{B}$ and that if $\mathcal{B}$ is linearly independent then so is $\mathcal{A}$.

1.32 Theorem: *Let F be a field, let U be a subspace of F^n and let $\mathcal{A} = \{u_1, u_2, \cdots, u_m\}$ be a basis for U. Let $\mathcal{B} = \{v_1, v_2, \cdots, v_l\} \subseteq U$. Suppose that $\mathcal{B}$ is linearly independent. Then $l \leq m$, if $l = m$ then $\mathcal{B}$ is a basis for U, and if $l < m$ then there exist $m - l$ vectors in $\mathcal{A}$ which, after possibly reordering the vectors u_i we can take to be $u_{l+1}, u_{l+2}, \cdots, u_m$, such that the set $\{v_1, v_2, \cdots, v_l, u_{l+1}, u_{l+2}, \cdots, u_m\}$ is a basis for U.*

Proof: When $l = 0$ so that $\mathcal{B} = \emptyset$, we have $m - l = m$ and we use all m of the vectors in $\mathcal{A}$ to obtain the basis $\{u_1, u_2, \cdots, u_m\}$. Let $l \geq 1$ and suppose, inductively, that for every set $\mathcal{B}_0 = \{v_1, v_2, \cdots, v_{l-1}\} \subseteq U$, if $\mathcal{B}_0$ is linearly independent then we have $l - 1 \leq m$ and we can reorder the vectors u_i so that $\{v_1, v_2, \cdots, v_{l-1}, u_l, u_{l+1}, \cdots, u_m\}$ is a basis for U. Let $\mathcal{B} = \{v_1, v_2, \cdots, v_l\} \subseteq U$. Suppose $\mathcal{B}$ is linearly independent. Let $\mathcal{B}_0 = \{v_1, v_2, \cdots, v_{l-1}\}$. Note that $\mathcal{B}_0$ is linearly independent but $\mathcal{B}_0$ idoes not span U because $v_l \in U$ but $v_l \notin \operatorname{Span}\mathcal{B}_0$. By the induction hypothesis, we have $l - 1 < m$ and we can reorder the vectors u_i so that $\{v_1, v_2, \cdots, v_{l-1}, u_l, u_{l+1}, \cdots, u_m\}$ is a basis for U. Since $v_l \in U$ we can write v_l in the form $v_l = \sum_{i=1}^{l-1} t_i v_i + \sum_{i=l}^{m} s_j u_j$. Note that the coefficients s_j cannot all be equal to zero since $v_l \notin \operatorname{Span}\mathcal{B}_0$. After reordering the vectors u_j we can suppose that $s_l \neq 0$. By Theorem XX, the set $\{v_1, v_2, \cdots, v_{l-1}, v_l, u_{l+1}, \cdots u_m\}$ is a basis fot U (in the case that $l - 1 = m$ this basis is the set $\mathcal{B}$).

1.33 Corollary: *For a vector space U in F^n, any two bases for U have the same number of elements.*

1.34 Definition: For a vector space U in F^n, we define the **dimension** of U, denoted by $\dim U$, to be the number of elements in any basis for U. For an affine space $P = p + U$ in F^n, we define the **dimension** of P to be $\dim P = \dim U$.

Chapter 2. Solving Systems of Linear Equations

Systems of Linear Equations

2.1 Definition: Let R be a ring and let $n \in \mathbf{Z}^+$. A **linear equation** in R^n (or a linear equation in the variables $x_1, x_2, \cdots, x_n$ over R) is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = b$$

with $a = (a_1, a_2, \cdots, a_n)^T \in R^n$ and $b \in R$. The numbers a_i are called the **coefficients** of the equation. A **solution** to the above equation is a point $x = (x_1, x_2, \cdots, x_n)^T \in R^n$ for which the equation holds. The **solution set** of the equation is the set of all solutions. Note that $x = 0$ is a solution if and only if $b = 0$, and in this case we say that the equation is **homogeneous**.

2.2 Example: Let F be a field, let $a = (a_1, a_2, \cdots, a_n)^T \in F^n$, let $b \in F$, and let S be the solution set of the linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$. Show that either $S = \emptyset$ (the empty set) or S is an affine space in F^n.

Solution: If $a = 0$ and $b \neq 0$ then the equation has no solution so we have $S = \emptyset$. If $a = 0$ and $b = 0$ then every $x \in F^n$ is a solution, so $S = F^n$. If $a \neq 0$ then we can choose an index k such that $x_k \neq 0$ and then we have

$$\begin{aligned}
\sum_{i=1}^{n} a_ix_i = b &\iff x_k = \tfrac{b}{a_k} - \sum_{i\neq k} \tfrac{a_i}{a_k}x_i \\
&\iff x = \Big(x_1, x_2, \cdots, x_{k-1}, \tfrac{b}{a_k} - \sum_{i\neq k} \tfrac{a_i}{a_k}\, x_i, x_{k+1}\cdots, x_n\Big)^T \\
&\iff x = p + \sum_{i\neq k} x_iu_i
\end{aligned}$$

where $p = \frac{b}{a_k}\, e_k$ and $u_i = e_i - \frac{a_i}{a_k}\, e_k$.

2.3 Definition: Let R be a ring and let $n, m \in \mathbf{Z}^+$. A **system of m linear equations** in R^n (or in n variables over R) is a set of m equations of the form

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

where each $a_{ij} \in R$ and each $b_j \in R$. The numbers a_{ij} are called the **coefficients** of the system. A **solution** to the above system is a point $x = (x_1, x_2, \cdots, x_n)^T \in R^n$ for which all of the m equations hold. The **solution set** of the system is the set of all solutions. Note that $x = 0$ is a solution if and only if $b = 0$, and in this case we call the system of linear equations **homogeneous**.

2.4 Remark: We shall describe an algorithmic method for solving a given system of liner equations over a field F. We shall see that if the the solution set is not empty then it is an affine space in F^n. The algorithm involves performing the following operations on the equations in the system. These operations do not change the solution set.

(1) $E_i \leftrightarrow E_j$: interchange the i^{th} and j^{th} equations,
(2) $E_i \mapsto t\,E_j$: multiply (both sides of) the i^{th} equation by t where $0 \neq t \in F$, and
(3) $E_i \mapsto E_i + t\,E_j$: add t times (each side of) the j^{th} equation to (the same side of) the i^{th} equation E_i, where $t \in F$.

For now, we illustrate the algorithm in a particular example.

2.5 Example: Consider the system of linear equations over the field $\mathbf{Q}$.

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 7 \\ x_1 + 0x_2 + 2x_3 &= 5 \\ 3x_1 - 4x_2 + 2x_3 &= 3 \end{aligned}$$

Show that the solution set is an affine space in $\mathbf{Q}^3$.

Solution: Performing operations of the above three types, we have

$$\Big(E_1 \leftrightarrow E_2\Big) \begin{array}{r} x_1 + 0x_2 + 2x_3 = 5 \\ 2x_1 - x_2 + 3x_3 = 7 \\ 3x_1 - 4x_2 + 2x_3 = 3 \end{array} \Big(\begin{array}{l} E_2 \mapsto E_2 - 2E_1 \\ E_3 \mapsto E_3 - 3E_1 \end{array}\Big) \begin{array}{r} x_1 + 0x_2 + 2x_3 = 5 \\ 0x_1 - x_2 - x_3 = -3 \\ 0x_1 - 4x_2 - 4x_3 = -12 \end{array}$$

$$\Big(E_2 \mapsto -E_2\Big) \begin{array}{r} x_1 + 0x_2 + 2x_3 = 5 \\ 0x_1 + x_2 + x_3 = 3 \\ 0x_1 - 4x_2 - 4x_3 = -12 \end{array} \Big(E_3 \mapsto E_3 + 4E_2\Big) \begin{array}{r} x_1 + 0x_2 + 2x_3 = 5 \\ 0x_1 + x_2 + x_3 = 3 \\ 0x_1 + 0x_2 + 0x_3 = 0 \end{array}$$

Thus the original system of 3 equations has the same solution set as the system

$$\begin{aligned} x_1 + 2x_3 &= 5 \\ x_2 + x_3 &= 3. \end{aligned}$$

If we let $x_3 = t$, then $x = (x_1, x_2, x_3)^T$ is a solution when $x_1 = 5 - 2t$, $x_2 = 3 - t$ and $x_3 = t$, that is when $x = (5, 3, 0)^T + t(-2, -1, 1)^T$. Thus the solution set is the line through $p = (5, 3, 0)^T$ in the direction of $u = (-2, -1, 1)^T$.

Matrix Notation

2.6 Remark: We wish to introduce some notation which will simplify our discussion of systems of linear equations. In fact the objects that we introduce will turn out to be of interest in their own right.

2.7 Definition: Let F be a field. An $m \times n$ **matrix** over F is an array of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

where each $a_{ij} \in F$. The number a_{ij} is called the (i, j) **entry** of the matrix A, and we write

$$A_{ij} = a_{ij}.$$

The set of all $m \times n$ matrices over F is denoted by $M_{m\times n}(F)$. The set of $n \times n$ matrices is also denoted by $M_n(F)$. Note that

$$F^n = M_{n\times 1}(F) \ , \ M_n(F) = M_{n\times n}(F) \ , \text{ and } F = M_1(F) \, .$$

The j^{th} **column** of the above matrix $A \in M_{m\times n}(F)$ is the vector $(a_{1j}, a_{2j}, \cdots, a_{mj})^T \in F^m$. The j^{th} **row** of A is the vector $(a_{j1}, a_{j2}, \cdots, a_{jn})^T \in F^n$. Given vectors $u_1, u_2, \cdots, u_n \in F^m$, the **matrix with columns** $u_1, u_2, \cdots, u_n$ is the matrix

$$A = \big(u_1, u_2, \cdots, u_n\big) \in M_{m\times n}(F)$$

and given vectors $v_1, v_2, \cdots, v_m \in F^n$, the **matrix with rows** $v_1, v_2, \cdots, v_m$ is the matrix

$$A = \begin{pmatrix} {v_1}^T \\ {v_2}^T \\ \vdots \\ {v_m}^T \end{pmatrix} \in M_{m\times n}(F) \, .$$

Given a matrix $A \in M_{m\times n}(F)$, the **transpose** of A is the matrix $A^T \in M_{n\times m}(F)$ with entries $(A^T)_{ij} = A_{ji} = a_{ji}$, that is

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix} .$$

The $m \times n$ **zero matrix** is the matrix $0 \in M_{m\times n}(F)$ whose entries are all equal to 0. The $n \times n$ **identity** matrix is the matrix $I \in M_n(F)$ with columns $e_1, e_2, \cdots, e_n$. Given a matrix $a = (a_1, a_2, \cdots, a_n) \in M_{1\times n}(F)$ and a vector $x = (x_1, x_2, \cdots, x_n)^T \in F^n$, we define the product $ax \in F$ to be

$$ax = \big(a_1, a_2, \cdots, a_n\big) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} a_i x_i = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

More generally, given a matrix $A \in M_{m\times n}(F)$ with entries $A_{ij} = a_{ij}$, and a vector $x \in F^n$ with entries $x_i \in F$, we define the product $Ax \in F^m$ to be

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

Equivalently, we define Ax to be the vector in F^m with entries

$$(Ax)_j = \sum_{i=1}^{n} a_{ji}x_i.$$

Using this notation, notice that for $a = (a_1, a_2, \cdots, a_n) \in M_{1\times n}(F)$ and $b \in F$, the single linear equation $a_1x_1 + a_2x_2 + \cdots a_nx_n = b$ can be written simply as $ax = b$, and for $A \in M_{m\times n}(F)$ with entries $A_{ij} = a_{ij}$ and $b = (b_1, b_2, \cdots, b_n)^T \in F^n$, the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_k \end{aligned}$$

can be written simply as

$$Ax = b.$$

2.8 Note: For vectors $v_1, v_2, \cdots, v_m \in F^n$ and for $x \in F^n$, if A is the matrix with rows $v_1, v_2, \cdots, v_m$ then the product Ax is defined so that we have

$$Ax = \begin{pmatrix} {v_1}^T \\ {v_2}^T \\ \vdots \\ {v_m}^T \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} {v_1}^T x_1 \\ {v_2}^T x_2 \\ \vdots \\ {v_m}^T x_n \end{pmatrix}.$$

For vectors $u_1, u_2, \cdots, u_n \in F^m$ and $x \in F^n$, if A is the matrix with columns $u_1, u_2, \cdots, u_n$, notice that we have

$$Ax = \big(u_1, u_2, \cdots, u_n\big) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1u_1 + x_2u_2 + \cdots + x_nu_n \in \text{Span}\,\{u_1, u_2, \cdots, u_n\}.$$

2.9 Note: For $0 \in M_{m\times n}(F)$, $I \in M_n(F)$ and $x \in F^n$ we have $0x = 0$ and $Ix = x$.

2.10 Theorem: *(Linearity) For $A \in M_{m\times n}(F)$, we have*
(1) $A(tx) = t\,Ax$ for all $t \in F$ and $x \in F^n$, and
(2) $A(x + y) = Ax + Ay$ for all $x, y \in F^n$.

Proof: For $t \in F$ and for $x, y \in F^n$ we have $\big(A(tx)\big)_j = \sum_{i=1}^{n} A_{ji}(tx)_i = t\sum_{i=1}^{n} A_{ji}x_i = \big(tAx\big)_j$ and $\big(A(x+y)\big)_j = \sum_{i=1}^{n} A_{ji}(x+y)_i = \sum_{i=1}^{n} A_{ji}(x_i+y_i) = \sum_{i=1}^{n} A_{ji}x_i + \sum_{i=1}^{n} A_{ji}y_i = \big(Ax + Ay\big)_j$.

Row Equivalence and Reduced Row Echelon Form

2.11 Definition: Given a matrix $A \in M_{m\times n}(F)$ with entries $A_{ij} = a_{ij}$ and a vector $b \in F^n$, we obtain the system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_n \end{aligned}$$

which we write simply as $Ax = b$. The matrix A is called the **coefficient matrix** of the system, and the matrix $\big(A \,|\, b\big) \in M_{k\times n+1}(F)$ is called the **augmented matrix** of the system. The **solution set** of the equation $Ax = b$ is the set

$$\big\{x \in F^n \big| Ax = b\big\}.$$

2.12 Definition: We noted earlier that we can perform three kinds of operations on the equations in the system without changing the solution set. These correspond to the following three kinds of operations that we can perform on the rows of the augmented matrix without changing the solution set.

(1) $R_i \leftrightarrow R_j$: interchange rows i and j,
(2) $R_i \mapsto t\,R_i$: multiply the i^{th} row by t, where $0 \neq t \in F$, and
(3) $R_i \mapsto R_i + t\,R_j$: add t times the j^{th} row to the i^{th} row, where $t \in F$.

These three kinds of operations will be called **elementary row operations**. Given matrices $A, B \in M_{m\times n}(F)$, we say that A and B are **row equivalent**, and we write $A \sim B$, when B can be obtained by applying a finite sequence of elementary row operations to A.

2.13 Remark: In the next section, we describe an algorithm for finding the solution set to a given matrix equation $Ax = b$. We shall use elementary row operations to construct a sequence of augmented matrices

$$(A|b) = (A_0|b_0) \sim (A_1|b_1) \sim (A_2|b_2) \sim \cdots \sim (A_l, b_l) = (R, c)$$

such that each equation $A_k x = b_k$ has the same solution set as the original equation $Ax = b$, and such that the final matrix $A_l = R$ is in a particularly nice form so that we can easily determine the solution set. We shall find that when the solution set is non-empty, we can write the solutions in the form

$$x = p + t_1u_1 + t_2u_2 + \cdots + t_ru_r.$$

2.14 Definition: A matrix $A \in M_{m\times n}(F)$ is said to be in **reduced row echelon form** when $A = 0$ or there exist column indices $1 \leq j_1 \leq j_2 \leq \cdots \leq j_r \leq n$, where $1 \leq r \leq n$, such that for each row index k with $1 \leq k \leq m$ we have

(1) $A_{k,j_k} = 1$,
(2) for each $j < j_k$ we have $A_{kj} = 0$,
(3) for each $i < k$ we have $A_{i,j_k} = 0$, and
(4) for all $i > r$ and all j we have $A_{ij} = 0$.

The entries $A_{k,j_k} = 1$ are called the **pivots**, the positions (k, j_k) where they occur are called the **pivot positions**, the columns $j_1, j_2, \cdots, j_r$ of A are called the **pivot columns**, and the number of pivots r is called the **rank** of the reduced row echelon matrix A.

2.15 Example: Consider the augmented matrix

$$(A|b) = \left(\begin{array}{ccccccc|c} 1 & -2 & 1 & 0 & -3 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Note that the matrix A is in reduced row echelon form with the pivots in positions $(1,1), (2,4)$ and $(3,6)$. The corresponding system of equations is

$$\begin{gathered} x_1 - 2x_2 + x_3 - 3x_5 + 2x_7 = 4 \\ x_4 + x_5 - x_7 = 1 \\ x_6 - 2x_7 = 3 \end{gathered}$$

The solutions x can be obtained by letting $x_2 = t_1$, $x_3 = t_2$, $x_5 = t_3$ and $x_7 = t_4$, with $t_1, t_2, t_3, t_4 \in F$ arbitrary, and then solving for x_1, x_4 and x_6 to get

$$\begin{gathered} x_1 = 4 + 2t_1 - t_2 + 3t_3 - 2t_4 \\ x_4 = 1 - t_3 + t_4 \\ x_6 = 3 + 2t_4 \end{gathered}$$

Thus the solution set is the set of points of the form

$$x = p + t_1u_1 + t_2u_2 + t_3u_3 + t_4u_4$$

where

$$p = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 1 \\ 0 \\ 3 \\ 0 \end{pmatrix}, \; u_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; u_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; u_3 = \begin{pmatrix} 3 \\ 0 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; u_4 = \begin{pmatrix} -2 \\ 0 \\ 0 \\ 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

We can also write this as

$$x = p + Bt$$

where B is the matrix with columns u_1, u_2, u_3, u_4.

2.16 Note: In general, suppose that $A \in M_{m\times n}(F)$ is in reduced row echelon form with pivot column indices $1 \le j_1 \le j_2 \le \cdots \le j_r \le n$. Let $1 \le l_1 \le l_2 \le \cdots \le l_{n-r} \le n$ be the non-pivot column indices. Write

$$(A \,|\, b) = \left(\begin{array}{c|c} R & c \\ 0 & d \end{array}\right)$$

where $R \in M_{r\times n}(F)$ is the matrix whose rows are the non-zero rows of A and where $c \in F^r$ and $d \in F^{m-r}$. Then the equation $Ax = b$ has a solution if and only if $d = 0$, and in this case, as in the above example, the solution is given by $x = p + Bt$ where $p \in F^n$ and $B \in M_{n\times m-r}(F)$ are given by $p_J = c$, $p_L = 0$, $B_J = -A_L$, and $B_L = I$ where $p_J = (p_{j_1}, p_{j_2}, \cdots, p_{j_r})^T$, $p_L = (p_{l_1}, p_{l_2}, \cdots, p_{l_{m-r}})^T$, B_J is the matrix whose rows are the rows $j_1, j_2, \cdots, j_r$ of B, B_L is the matrix whose rows are the rows $l_1, l_2, \cdots, l_{m-r}$ of B, and A_L is the matrix whose columns are the columns $l_1, l_2, \cdots, l_{m-r}$ of A.

Gauss-Jordan Elimination

2.17 Theorem: *(Gauss-Jordan Elimination) Let $A \in M_{m\times n}(F)$ and let $b \in F^m$, and consider the equation $Ax = b$. If $A = 0$ and $b = 0$ then every $x \in F^n$ is a solution so the solution set is $\mathbf{F}^n$. If $A = 0$ and $b \neq 0$ then there is no solution, so the solution set is $\emptyset$. If $A \neq 0$ then we can perform a series of elementary row operations to obtain a sequence of augmented matrices*

$$(A|b) = (A_0|b_0) \sim (A_1|b_1) \sim (A_2|b_2) \sim \cdots \sim (A_l|b_l) = \left(\begin{array}{c|c} R & c \\ 0 & d \end{array}\right)$$

where A_l is in reduced row echelon form and R is the matrix whose rows are the non-zero rows of A_l. If $d \neq 0$ then the solution set is empty and if $d = 0$ then the solution set is the affine space $x = p + Bt$, as described in Note 2.16 above.

Proof: Suppose that $A \neq 0$. We describe an algorithm to obtain the required sequence of augmented matrices.

Step 1: choose the smallest index $j = j_1$ such that the j^{th} column of A is not zero, then choose the smallest index $i = i_1$ such that $a_{ij} \neq 0$. Perform the row operations $R_i \mapsto \frac{1}{a_{ij}} R_i$ then $R_1 \leftrightarrow R_i$ to obtain a new augmented matrix $(A'|b')$ whose first $j-1$ columns are all zero, with $A'_{1j} = 1$. For each $i > 1$ perform the row operation $R_i \mapsto R_i - A'_{ij} R_1$ to obtain a new augmented matrix $(A_1|b_1)$ whose first $j-1$ columns are zero and whose j^{th} column is e_1.

Step $s+1$: suppose that we have performed the first s steps in the algorithm and have obtained an augmented matrix (A_s, b_s) which is row-equivalent to $(A|b)$ and which has the property that there exist column indices $1 \leq j_1 \leq j_2 \leq \cdots \leq j_s \leq n$ such that for each row index k with $1 \leq k \leq s$ we have

(1) $(A_s)_{k,j_k} = 1$,
(2) for each $j < j_k$ we have $(A_s)_{kj} = 0$,
(3) for each $i < k$ we have $(A_s)_{i,j_k} = 0$, and
(4) for all $i > s$ and all $j \leq j_s$ we have $(A_s)_{ij} = 0$.

If $s = m$ or if $j_s = n$ then we are done because the matrix A_s is already in reduced row echelon form. Suppose that $s < m$ and $j_s < n$. If for all $i > s$ and all $j > j_s$ we have $(A_s)_{ij} = 0$ then we are done because the matrix A_s is already in row echelon form. Suppose that $A_{ij} \neq 0$ for some $i > s$ and some $j > j_s$. Let $j = j_{s+1}$ be the smallest index such that $(A_s)_{ij} \neq 0$ for some $i > s$, then apply elementary row operations, involving only rows i with $i > s$, to the augmented matrix $(A_s|b_s)$ to obtain a matrix $(A'_s|b'_s)$ with $(A'_s)_{s+1,j} = 1$; this can be done for example by choosing an index $i > s$ such that $(A_s)_{ij} \neq 0$ and then performing the row operation $R_i \mapsto \frac{1}{(A_s)_{ij}} R_i$ then (if $i \neq s+1$) the operation $R_{s+1} \leftrightarrow R_i$. Then for each $i \neq s+1$ perform the row operation $R_i \mapsto R_i - (A'_s)_{ij} R_{s+1}$ to the augmented matrix $(A'_s|b_s)$ to obtain the row-equivalent augmented matrix $(A_{s+1}|b_{s+1})$. Verify that the new matrix satisfies the above 4 properties with s replaced by $s+1$.

2.18 Definition: The algorithm for solving the system $Ax = b$ described in the proof of the above theorem is called **Gauss-Jordan elimination**.

2.19 Example: Solve the system $3x+2y+z=4$, $2x+y-z=3$, $x+2y+4z=1$ in $\mathbf{Q}^3$.

Solution: We form the augmented matrix for the system then perform row operations to reduce to reduced row echelon form.

$$(A|b)=\left(\begin{array}{ccc|c}3&2&1&4\\2&1&-1&3\\1&2&4&1\end{array}\right)\ R_1\mapsto R_1-R_3\ \left(\begin{array}{ccc|c}1&1&2&1\\2&1&-1&3\\1&2&4&1\end{array}\right)\ \begin{array}{l}R_2\mapsto R_2-2R_1\\R_3\mapsto R_3-R_1\end{array}$$

$$\left(\begin{array}{ccc|c}1&1&2&1\\0&1&5&-1\\0&1&2&0\end{array}\right)\ \begin{array}{l}R_1\mapsto R_1-R_2\\\\R_3\mapsto R_3-R_2\end{array}\ \left(\begin{array}{ccc|c}1&0&-3&2\\0&1&5&-1\\0&0&-3&1\end{array}\right)\ R_3\mapsto -\tfrac{1}{3}R_3$$

$$\left(\begin{array}{ccc|c}1&0&-3&2\\0&1&5&-1\\0&0&1&-\frac{1}{3}\end{array}\right)\ \begin{array}{l}R_1\mapsto R_1-3R_3\\R_2\mapsto R_2+5R_3\end{array}\ \left(\begin{array}{ccc|c}1&0&0&1\\0&1&0&\frac{2}{3}\\0&0&1&-\frac{1}{3}\end{array}\right)$$

From the final reduced matrix we see that the solution is given by $(x,y,z)=\left(1,\frac{2}{3},-\frac{1}{3}\right)$.

2.20 Remark: Note that in the above solution, at the first step we used the row operation $R_1\mapsto R_1-R_3$ to obtain a pivot in position $(1,1)$, but we could have achieved this in many different ways. For example, we could have used the row operation $R_1\leftrightarrow R_3$ or we could have used $R_1\mapsto\frac{1}{3}R_1$.

2.21 Example: Solve the system $2x+y+3z=1$, $3x+y+5z=2$, $x-y+3z=0$ in $\mathbf{Q}^3$.

Solution: Using Gauss-Jordan elimination, we have

$$(A|b)=\left(\begin{array}{ccc|c}2&1&3&1\\3&1&5&2\\1&-1&3&0\end{array}\right)\sim\left(\begin{array}{ccc|c}1&2&0&1\\3&1&5&2\\1&-1&3&0\end{array}\right)\sim\left(\begin{array}{ccc|c}1&1&2&1\\0&1&5&-1\\0&1&2&0\end{array}\right)$$

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

From the reduced matrix, we see that there is no solution.

2.22 Example: Solve the system $x_1+2x_2+x_3+3x_4=2$, $2x_1+3x_2+x_3+4x_4+x_5=5$, $x_1+3x_2+2x_3+4x_4+x_5=3$ in $\mathbf{Q}^5$.

Solution: Using Gauss-Jordan elimination gives

$$(A|b)=\left(\begin{array}{ccccc|c}1&2&1&3&0&2\\2&3&1&4&1&5\\1&3&2&4&2&3\end{array}\right)\sim\left(\begin{array}{ccccc|c}1&2&1&3&0&2\\0&1&1&2&-1&-1\\0&1&1&1&2&1\end{array}\right)$$

$$\sim\left(\begin{array}{ccccc|c}1&0&-1&-1&2&4\\0&1&1&2&-1&-1\\0&0&0&1&-3&-2\end{array}\right)\sim\left(\begin{array}{ccccc|c}1&0&-1&0&-1&2\\0&1&1&0&5&3\\0&0&0&1&-3&-2\end{array}\right)$$

From the reduced matrix we see that the solution set is the plane given by $x=p+su+tv$ where $p=(2,3,0,-2,0)^T$, $u=(1,-1,1,0,0)^T$ and $v=(1,-5,0,3,1)^T$.

Chapter 3. Matrices and Concrete Linear Maps

The Row Space, Column Space and Null Space of a Matrix

3.1 Definition: Let R be a ring. For a matrix $A \in M_{m\times n}(R)$, the **row span** of A, denoted by $\text{Row}(A)$, is the span of the rows of A, the **column span** of A, denoted by $\text{Col}(A)$, is the span of the columns of A, and the **null set** of A, is the set

$$\text{Null}(A) = \big\{x \in F^n \big| Ax = 0\big\}.$$

When F is a field, $\text{Row}(A)$ and $\text{Col}(A)$ are also called the **row space** and **column space** of A, and we define the **rank** of A and the **nullity** of A are the dimensions

$$\text{rank}(A) = \dim\big(\text{Col}A\big) \quad \text{and} \quad \text{nullity}(A) = \dim\big(\text{Null}A\big).$$

3.2 Note: For $A = (u_1, u_2, \cdots, u_n) \in M_{m\times n}(R)$ and $t \in R^n$ we have $At = \sum_{i=1}^{n} t_i u_i$, so

$$\text{Col}(A) = \big\{At \big| t \in R^n\big\}.$$

3.3 Theorem: *Let F be a field, let $A \in M_{m\times n}(F)$ and let $b \in F^m$. If $x = p$ is a solution to the equation $Ax = b$ then*

$$\big\{x \in F^n \big| Ax = b\big\} = p + \text{Null}(A).$$

Proof: If $Ap = b$ then for $x \in F^n$ we have

$$Ax = b \iff Ax = Ap \iff A(x-p) = 0 \iff (x-p) \in \text{Null}A \iff x \in p + \text{Null}(A).$$

3.4 Note: For $\mathcal{A} = \{u_1, u_2, \cdots, u_n\} \subseteq F^m$ and $A = (u_1, u_2, \cdots, u_n) \in M_{m\times n}(F)$,

$\mathcal{A}$ is linearly independent

$\iff$ for all $t_1, t_2, \cdots, t_n \in F$, if $\sum_{i=1}^{n} t_i u_i = 0$ then each $t_i = 0$

$\iff$ for all $t \in F^n$, if $At = 0$ then $t = 0$

$\iff$ $\text{Null}(A) = \{0\} \iff \text{Null}(R) = \{0\}$

$\iff$ R has a pivot in every column $\iff$ R is of the form $R = \begin{pmatrix} I \\ 0 \end{pmatrix}$, and

$\mathcal{A}$ spans $F^m \iff \text{Col}(A) = F^m$

$\iff$ for every $x \in F^m$ there exists $t \in F^n$ with $At = x$

$\iff$ for every $y \in F^m$ there exists $t \in F^n$ with $Rt = y$

$\iff$ R has a pivot in every row.

3.5 Theorem: *Let F be a field, let $A = (u_1, u_2, \cdots, u_n) \in M_{m\times n}(F)$, and suppose $A \sim R$ where R is in reduced row echelon form with pivots in columns $1 \le j_1 < j_2 < \cdots < j_r \le n$. Then*

(1) the non-zero rows of R form a basis for $\text{Row}(A)$,

(2) the set $\{u_{j_1}, u_{j_2}, \cdots, u_{j_r}\}$ is a basis for $\text{Col}(A)$, *and*

(3) when we solve $Ax = b$ using Gauss-Jordan elimination and write the solution as $x = p + Bt$ as in Note 2.16, the columns of B form a basis for $\text{Null}(A)$.

Proof: First we prove Part (1). By Theorem 1.31, when we perform an elementary row operation on a matrix, the span of the rows is unchanged, and so we have $\text{Row}(A) = \text{Row}(R)$. The nonzero rows of R span $\text{Row}(R)$, so it suffices to show that the nonzero rows of R are linearly independent. Let $1 \leq j_1 < j_2 < \cdots < j_r$ be the indices of the pivot columns in R. Let $v_1, v_2, \cdots, v_r$ be the nonzero rows of R. Because R is in reduced row echelon form, for $1 \leq i \leq r$ and $1 \leq k \leq r$ we have $(v_i)_{j_k} = \delta_{i,k}$. It follows that $\{v_1, v_2, \cdots, v_r\}$ is linearly independent because if $\sum_{i=1}^{r} t_i v_i = 0$ with each $t_i \in F$ then for all k with $1 \leq k \leq r$ we have

$$0 = \big(\sum_{i=1}^{r} t_i v_i\big)_{j_k} = \sum_{i=1}^{r} t_i (v_i)_{j_k} = \sum_{i=1}^{r} t_i \delta_{i,k} = t_k.$$

To prove Part (2), let $1 \leq l_1 < l_2 < \cdots, l_{n-r} \leq n$ be the indices of the non-pivot columns. Let $v_1, v_2, \cdots, v_n \in F^m$ be the columns of R and note that we have $v_{j_i} = e_i$ for $1 \leq i \leq r$. When we use row operations to reduce A to R, the same row operations reduce $A_J = (u_{j_1}, \cdots, u_{j_r})$ to $R_J = (v_{j_1}, \cdots, v_{j_t}) = (e_1, \cdots, e_r) = \begin{pmatrix} I \\ 0 \end{pmatrix}$. This shows that $\{u_{j_1}, \cdots, u_{j_r}\}$ is linearly independent. When we use row operations to reduce A to R, the same row operations will reduce $(A|u_k)$ to $(R|v_k)$, and so the equation $Ax = u_k$ has the same solutions as the equation $Rx = v_k$. Since only the first r columns of R are nonzero, each column v_k can be written as $v_k = \sum_{i=1}^{r} (v_k)_i e_i = \sum_{i=1}^{r} (v_k)_i v_{j_i} = Rt$ where $t \in R^n$ is given by $t_J = v_k$ and $t_L = 0$. Since $Ax = u_k$ and $Rx = v_k$ have the same solutions, we also have $u_k = At = \sum_{i=1}^{r} (v_k)_i u_{j_i} \in \text{Span}\{u_{j_1}, u_{j_2}, \cdots, u_{j_r}\}$. This shows that $\text{Col}(A) = \text{Span}\{u_1, u_2, \cdots, u_n\} = \text{Span}\{u_{j_1}, \cdots, u_{j_r}\}$.

Since the solution set to the equation $Ax = b$ is the set

$$\{x \in \mathbf{R}^n | Ax = b\} = p + \text{Col}(B) = p + \text{Null}(A)$$

we must have $\text{Col}(B) = \text{Null}(A)$. Since (as in Note 2.16) we have $B_L = I$, it follows that the columns of B are linearly independent using the same argument that we used in Part (1) to show that the nonzero rows of R are linearly independent. This proves Part (3).

3.6 Corollary: *Let F be a field, let $A \in M_{m\times n}(F)$, suppose that A is row equivalent to a reduced row echelon matrix which has r pivots. Then*

$$\text{rank}(A) = \dim(\text{Row}A) = \dim(\text{Col}A) = r \text{ , and}$$
$$\text{nullity}(A) = \dim(\text{Null}A) = n - r.$$

3.7 Corollary: *Let F be a field, let $A \in M_{m\times n}(F)$ and suppose that A is row equivalent to a row reduced echelon matrix R.*

(1) The rows of A are linearly independent $\iff$ the columns of A span F^m $\iff$ $\text{rank}(A) = m$ $\iff$ R has a pivot in every row.

(2) The rows of A span F^n $\iff$ the columns of A are linearly independent $\iff$ $\text{rank}(A) = n$ $\iff$ R has a pivot in every column $\iff$ R is of the form $R = \begin{pmatrix} I \\ 0 \end{pmatrix}$.

(3) The rows of A form a basis for $\mathbf{R}^n$ $\iff$ the columns of A form a basis for F^m $\iff$ $\text{rank}(A) = m = n$ $\iff$ $R = I$.

Matrices and Linear Maps

3.8 Definition: Let R be a ring. A map $L : R^n \to R^m$ is called **linear** when
(1) $L(x+y) = L(x) + L(y)$ for all $x, y \in R^n$, and
(2) $L(tx) = t\,L(x)$ for all $x \in R^n$ and all $t \in R$.

3.9 Note: Given a matrix $A \in M_{m\times n}(R)$, the map $L : \mathbf{R}^n \to R^m$ given by $L(x) = Ax$ is linear.

3.10 Theorem: *Let $L : R^n \to R^m$ be linear. There exists a unique matrix $A \in M_{m\times n}(R)$ such that $L(x) = Ax$ for all $x \in R^n$, namely the matrix $A = \big(L(e_1), L(e_2), \cdots, L(e_n)\big)$.*

Proof: Let $L : R^n \to R^m$ and let $A = (u_1, u_2, \cdots, u_n) \in M_{m\times n}(R)$. If $L(x) = Ax$ for all $x \in R$ then for each index k we have $u_k = Ae_k = L(e_k)$. Conversely, suppose that $u_k = L(e_k)$ for every index k. Then for all $x \in R^n$ we have

$$L(x) = L\big(\sum_{i=1}^{n} x_i e_i\big) = \sum_{i=1}^{n} x_i\, L(e_i) = \sum_{i=1}^{n} x_i u_i = Ax\,.$$

3.11 Notation: Often, we shall not make a notational distinction between the matrix $A \in M_{m\times n}(R)$ and its corresponding linear map $A : R^n \to R^m$ given by $A(x) = Ax$. When we do wish to make a distinction, we shall use the following notation. Given a matrix $A \in M_{m\times n}(R)$ we let $L_A : R^n \to R^n$ be the linear map given by

$$L_A(x) = Ax \quad \text{for all } x \in R^n$$

and given a linear map $L : R^n \to R^m$ we let $[L]$ be the corresponding matrix given by

$$[L] = \big(L(e_1), L(e_2), \cdots, L(e_n)\big) \in M_{m\times n}(R).$$

3.12 Definition: For a linear map $L : R^n \to R^m$, the **kernel** (or the **null set** of L is the set

$$\mathrm{Ker}(L) = \mathrm{Null}(L) = L^{-1}(0) = \{x \in R^n | L(x) = 0\}$$

and the **image** (or the **range** of L) is the set

$$\mathrm{Image}(L) = \mathrm{Range}(L) = L(R^n) = \{L(x) | x \in R^n\}.$$

We also use the same terminology for a matrix $A \in M_{m\times n}(R)$ when we think of the matrix as a linear map, so when $A = [L]$ we have $\mathrm{Ker}(L) = \mathrm{Null}(L) = \mathrm{Ker}(A) = \mathrm{Null}(A)$ and $\mathrm{Image}(L) = \mathrm{Range}(L) = \mathrm{Image}(A) = \mathrm{Range}(A) = \mathrm{Col}(A)$. When F is a field and $L : F^n \to F^m$ is linear, we define the **rank** and the **nullity** of L to be the dimensions

$$\mathrm{rank}(L) = \dim\big(\mathrm{Range}(L)\big) \quad \text{and} \quad \mathrm{nullity}(L) = \dim\big(\mathrm{Null}(L)\big).$$

3.13 Theorem: *Let R be a ring and let $L : R^n \to R^m$ be a linear map. Then*
(1) L is surjective if and only if $\mathrm{Range}(L) = F^m$, *and*
(2) L is injective if and only if $\mathrm{Null}(L) = \{0\}$.

Proof: Part (1) is obvious, so we only prove Part (2). Note that since L is linear we have $L(0) = L(0 \cdot 0) = 0\,L(0) = 0$ and so $0 \in \mathrm{Null}(L)$. Suppose that L is injective. Then for $x \in R^n$ we have $x \in \mathrm{Null}(L) \Longrightarrow L(x) = 0 \Longrightarrow L(x) = L(0) \Longrightarrow x = 0$ so $\mathrm{Null}(L) = \{0\}$. Conversely, suppose that $\mathrm{Null}(L) = \{0\}$. Then for $x, y \in R^n$ we have

$$L(x) = L(y) \Longrightarrow L(x-y) = 0 \Longrightarrow (x-y) \in \mathrm{Null}(L) = \{0\} \Longrightarrow x - y = 0 \Longrightarrow x = y$$

and so L is injective.

3.14 Example: The **identity map** on R^n is the map $I : R^n \to R^n$ given by $I(x) = x$ for all $x \in R^n$, and it corresponds to the identity matrix $I \in M_n(R)$ with entries $I_{i,j} = \delta_{i,j}$. The **zero map** $O : R^n \to R^m$ given by $O(x) = 0$ for all $x \in R^n$ corresponds to the zero matrix $O \in M_{m\times n}(R)$ with entries $O_{i,j} = 0$ for all i, j.

3.15 Note: Given linear maps $L, M : R^n \to R^m$ and $K : R^m \to R^l$ and given $t \in R$, the maps $(L + M) : R^n \to R^n$, tL and $KL : R^n \to R^l$ given by $(L + M)(x) = L(x) + M(x)$, $(tL)(x) = t\,L(x)$ and $KL : R^n \to R^l$ are all linear. For example, to see that KL is linear, note that for $x, y \in R^n$ and $t \in R$ we have

$$\begin{aligned}(KL)(x + y) &= K\big(L(x + y)\big) = K\big(L(x) + L(y)\big)\\ &= K\big(L(x)\big) + K\big(L(y)\big) = (KL)(x) + (KL)(y)\ , \text{ and}\\ KL(tx) &= K\big(L(tx)\big) = K\big(t\,L(x)\big) = t\,K\big(L(x)\big) = t(KL)(x).\end{aligned}$$

3.16 Definition: Given $A, B \in M_{m\times n}(R)$ we define $A + B \in M_{m\times n}(R)$ to be the matrix such that $(A + B)(x) = Ax + Bx$ for all $x \in R^n$. Given $A \in M_{m\times n}(R)$ and $t \in R$, we define $tA \in M_{m\times n}(R)$ to be the matrix such that $(tA)(x) = t\,Ax$ for all $x \in R^n$. Given $A \in M_{l\times m}(R)$ and $B \in M_{m\times n}(R)$ we define $AB \in M_{l\times n}(R)$ to be the matrix such that $(AB)x = A(Bx)$ for all $x \in R^n$.

3.17 Note: From the above definitions, it follows immediately that for all matrices A, B, C of appropriate sizes and for all $s, t \in R$, we have

(1) $(A + B) + C = A + (B + C)$,
(2) $A + B = B + A$,
(3) $O + A = A = A + O$,
(4) $A + (-A) = 0$,
(5) $(AB)C = A(BC)$,
(6) $IA = A = AI$,
(7) $OA = O$ and $AO = O$,
(8) $(A + B)C = AC + BC$ and $A(B + C) = AB + AC$,
(9) $s(tA) = (st)A$,
(10) if R is commutative then $A(tB) = t(AB)$,
(11) $(s + t)A = sA + tA$ and $t(A + B) = tA + tB$, and
(12) $0A = O$, $1A = A$ and $(-1)A = -A$.

In particular, the set $M_n(R)$ is a ring under addition and multiplication of matrices.

3.18 Theorem: *For $A, B \in M_{m\times n}(R)$ and $t \in R$, the matrices $A + B$ and tA are given by $(A + B)_{i,j} = A_{i,j} + B_{i,j}$ and $(tA)_{i,j} = t\,A_{i,j}$. For $A = (u_1, u_2, \cdots, u_l)^T \in M_{l\times m}(R)$ and $B = (v_1, v_2, \cdots, v_n) \in M_{m\times n}(R)$, the matrix AB is given by*

$$(AB)_{j,k} = {v_j}^T u_k = \sum_{i=1}^{m} A_{j,i}B_{i,k}\,.$$

Proof: For $A, B \in M_{m\times n}(R)$, the k^{th} column of $(A+B)$ is equal to $(A+B)e_k = Ae_k + Be_k$ which is the sum of the k^{th} columns of A and B. It follows that $(A + B)_{j,k} = A_{j,k} + B_{j,k}$ for all j, k. Similarly for $t \in R$, the k^{th} column of tA is equal to $(tA)e_k = t\,Ae_k$ which is t times the k^{th} column of A.

Now let $A = (u_1, \cdots, u_l)^T \in M_{l\times m}(R)$ and $B = (v_1, \cdots, v_n) \in M_{m\times n}(R)$. The k^{th} column of (AB) is equal to $(AB)e_k = A(Be_k) = Av_k$, so the (j, k) entry of AB is equal to

$$(AB)_{j,k} = {v_j}^T u_k = \big(A_{j,1}, A_{j,2}, \cdots, A_{j,m}\big)$$

The Transpose and the Inverse

3.19 Definition: For a linear map $L : R^n \to R^m$ the **transpose** of L is the map $L^T : R^m \to R^n$ such that $[L^T] = [L]^T$.

3.20 Note: When R is a ring, for $A \in M_{m\times n}(R)$ we have $\text{Row}(A) = \text{Col}(A^T)$ and $\text{Col}(A) = \text{Row}(A^T)$. When F is a field, for $A \in M_{m\times n}(F)$ we have $\text{rank}(A) = \text{rank}(A^T)$ and for a linear map $L : R^n \to R^m$ we have $\text{rank}(L) = \text{rank}(L^T)$.

3.21 Definition: For linear maps $L : R^n \to R^m$ and $M : R^m \to R^n$, when $LM = I$ where $I : R^m \to R^m$ we say that L is a **left inverse** of M and that M is a **right inverse** of L, and when $LM = I$ and $ML = I$ we say that L and M are (two-sided) **inverses** of each other. When $L : R^n \to R^m$ has a (two-sided) inverse $M : R^m \to R^n$ we say that L is **invertible**. We use the same terminology for matrices $A \in M_{m\times n}(R)$ and $B \in M_{n\times m}(R)$.

3.22 Theorem: *Let R be a ring, let $A \in M_{m\times n}(R)$ and $B \in M_{n\times m}(R)$. If B is a left inverse of A and C is a right inverse of A then $B = C$. A similar result holds for linear maps $L : R^n \to R^m$ and $K, M : R^m \to R^n$.*

Proof: Suppose that $BA = I$ and that $AC = I$. Then

$$B = BI = B(AC) = (BA)C = IC = C\,.$$

3.23 Theorem: *Let R be a commutative ring.*
(1) For $A, B \in M_{m\times n}(R)$ and $t \in R$ we have

$$(A^T)^T = A\ ,\ (A+B)^T = A^T + B^T \ \text{ and } \ (tA)^T = t\,A^T.$$

A similar result holds for linear maps $L, M : \mathbf{R}^n \to R^m$.
(2) If $A \in M_{l\times m}(R)$ and $B \in M_{m\times n}(R)$ then

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

A similar result holds for linear maps $L : R^l \to R^m$ and $M : R^m \to R^n$.
(3) For invertible matrices $A, B \in M_n(R)$ and for an invertible element $t \in R$ we have

$$(A^{-1})^{-1}\ ,\ (tA)^{-1} = \tfrac{1}{t}\,A^{-1} \ \text{ and } \ (AB)^{-1} = B^{-1}A^{-1}.$$

A similar result holds for invertible linear maps $L, M : R^n \to R^n$.

Proof: We leave the proof of Part (1) as an exercise. To prove Part (2), suppose that R is commutative and let $A \in M_{l\times m}(R)$ and $B \in M_{m\times n}(R)$. Then for all indices j, k we have

$$(AB)^T{}_{j,k} = (AB)_{k,j} = \sum_{i=1}^{m} A_{k,i}B_{i,j} = \sum_{i=1}^{m} B_{i,j}A_{k,i} = \sum_{i=1}^{m} B^T j, iA^T{}_{i,k} = (B^TA^T)_{j,k}.$$

To prove Part (3), let $A, B \in M_n(R)$ be invertible matrices and let $t \in R$ be an invertible element. Because $AA^{-1} = I$ and $A^{-1}A = I$, it follows that $(A^{-1})^{-1} = A$. Because $(tA)\big(\frac{1}{t}A\big) = \big(t \cdot \frac{1}{t}\big)\,AA^{-1} = 1 \cdot I = I$ and similarly $\big(\frac{1}{t}A\big)\big(tA\big) = I$, it follows that $(tA)^{-1} = \frac{1}{t}\,A$. Because $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$ and similarly $B^{-1}A^{-1})(AB) = I$, it follows that $(AB)^{-1} = B^{-1}A^{-1}$.

3.24 Theorem: *Let F be a field and let $A \in M_{m\times n}(F)$,*

(1) A is surjective $\iff$ A has a right inverse matrix,
(2) A is injective $\iff$ A has a left inverse matrix,
(3) if A is bijective then $n = m$ and A has a (two-sided) inverse matrix, and
(4) when $n = m$, A is bijective $\iff$ A is surjective $\iff$ A is injective.

A similar result holds for a linear map $L : R^n \to R^m$.

Proof: We prove Part (1). Suppose first that A has a right inverse matrix, say $AB = I$ with $B \in M_{n\times n}(F)$. Then given $y \in F^m$ we can choose $x \in F^n$ to get

$$Ax = A(By) = (AB)y = Iy = y.$$

Thus A is surjective. Conversely, suppose that A is surjective. For each index $k \in \{1, 2, \cdots, m\}$, choose $u_k \in F^n$ so that $Au_k = e_k$, and then let $B = (u_1, u_2, \cdots, u_m) \in M_{n\times m}(F)$. Then we have

$$AB = A(u_1, u_2, \cdots, u_m) = \big(Au_1, Au_2, \cdots, Au_m\big) = (e_1, e_2, \cdots, e_m) = I\,.$$

To prove Part (2), suppose first that A has a left inverse matrix, say $BA = I$ with $B \in M_{n\times n}(F)$. Then for $x \in F^n$ we have

$$Ax = 0 \Longrightarrow B(Ax) = 0 \Longrightarrow (BA)x = 0 \Longrightarrow Ix = 0 \Longrightarrow x = 0$$

and so $\text{Null}(A) = \{0\}$. Thus A is injective. Conversely, suppose that A is injective. Then $\text{Null}(A) = \{\}$, so the columns of A are linearly independent, hence the rows of A span F^n, equivalently the columns of A^T span F^n, hence $\text{Range}(A^T) = F^n$ and so A^T is surjective. Since A^T is surjective, we can choose $C \in M_{m\times n}(F)$ so that $A^TC = I$. Let $B = C^T$ so that $A^TB^T = I$. Transpose both sides to get $BA = I^T = I$. Thus the matrix B is a left inverse of A.

Parts (3) and (4) follow easily from Parts (1) and (2) together with previous results (namely Note 3.4, Corollary 3.7 and Theorems 3.13 and 3.22).

3.25 Note: To obtain a right inverse of a given matrix $A \in M_{m\times n}(F)$ using the method described in the proof of Part (1) of the above theorem, we can find vectors $u_1, u_2, \cdots, u_m \in F^n$ such that $Au_k = e_k$ for each index k by reducing each of the augmented matrices $(A|e_k)$. Since the same row operations which are used to reduce $(A|e_1)$ to the form $(R|u_1)$, (with R in reduced echelon form) will also reduce each of the augmented matrices $(A|e_k)$ to the form $(R|e_k)$, we can solve all of the equations $Au_k = e_k$ simultaneously by reducing the matrix $(A|I) = (A|e_1, e_2, \cdots, e_m)$ to the form $(R|u_1, u_2, \cdots, u_m)$.

3.26 Example: Let $A = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 4 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in M_3(\mathbf{Q})$. Find A^{-1}.

Solution: We have

$$(A|I) = \left(\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & -2 & -3 & -2 & 1 & 0 \\ 0 & -2 & -2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 3 & 2 & 1 & 0 & 0 \\ 0 & 1 & \frac{3}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & -2 & -2 & -1 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & -1 & \frac{5}{2} \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & -\frac{3}{2} \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array}\right)$$

and so A^{-1} is equal to the matrix whichh appears on the right of the final matrix above.

Chapter 4. Determinants

Permutations

4.1 Definition: A **group** is a set G together with an element $e \in G$, called the **identity** element, and a binary operation $* : G \times G \to G$, where for $a, b \in G$ we write $*(a, b)$ as $a * b$ or often simply as ab, such that

(1) $*$ is associative: $(ab)c = a(bc)$ for all $a, b, c \in G$,
(2) e is an identity: $ae = a = ea$ for all $a \in G$, and
(3) every $a \in G$ has an inverse: for every $a \in G$ there exists $b \in G$ with $ab = e = ba$.

A group G is called **abelian** when
(4) $*$ is commutative: $ab = ba$ for all $a, b \in G$.

4.2 Note: Let G be a group. Note that the identity element $e \in G$ is the unique element that satisfies Axiom (2) in the above definition because if $u \in G$ has the property that $ua = a = au$ for all $a \in G$, then in the case that $a = e$ we obtain $e = ue = e$. Also note given $a \in G$ the element b which satisfies Axiom (3) above is unique because if $ab = e$ and $ca = e$ then we have $b = eb = (ca)b = c(ab) = ce = e$.

4.3 Definition: Let G be a group. Given $a \in G$, the unique element $b \in G$ such that $ab = e = ba$ is called the **inverse** of a and is denoted by a^{-1} (unless the operation in G is addition denoted by $+$, in which case the inverse of a is also called the **negative** of a and is denoted by $-a$). We write $a^0 = e$ and for $k \in \mathbf{Z}^+$ we write $a^k = aa \cdots a$ (where the product involves k copies of a) and $a^{-k} = (a^k)^{-1}$.

4.4 Note: In a group G, we have the cancellation property: for all $a, b, c \in G$, if $ab = ac$ (or if $ca = ba$) then $b = c$. Indeed, if $ab = ac$ then

$$b = eb = (a^{-1}a)b = a^{-1}(ab) = a^{-1}(ac) = (a^{-1}a)c = ec = c.$$

4.5 Example: If R is a ring under addition and multiplication then R is also an abelian group under addition. The identity element is 0 and the inverse of $a \in R$ is $-a$. For example $\mathbf{Z}_n$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{C}$ are all abelian groups under addition.

4.6 Example: If R is a ring under addition and multiplication then the set

$$R^* = \big\{a \in R \big| a \text{ is invertible}\big\}$$

is a group under multiplication. The identity element is 1 and the inverse of $a \in R$ is a^{-1}. For example $\mathbf{Z}^* = \{1, -1\}$, $\mathbf{Q}^* = \mathbf{Q} \setminus \{0\}$, $\mathbf{R}^* = \mathbf{R} \setminus \{0\}$ and $\mathbf{C}^* = \mathbf{C} \setminus \{0\}$ are all abelian groups under multiplication. For $n \in \mathbf{Z}^+$, the **group of units modulo** n is the group

$$U_n = {\mathbf{Z}_n}^* = \big\{a \in \mathbf{Z}_n \big| \gcd(a, n) = 1\big\}.$$

The group of units U_n is an abelian group under multiplication modulo n. When R is a ring (usually commutative), the **general linear group** $GL_n(R)$ is the group

$$GL_n(R) = \big\{A \in M_n(R) \big| A \text{ is invertible}\big\}.$$

When $n \geq 2$, the general linear group $GL_n(R)$ is a non-abelian group under matrix multiplication.

4.7 Definition: Let X be a set. The **group of permutations** of X is the group

$$\mathrm{Perm}\,(X) = \big\{ f : X \to X \big| f \text{ is bijective} \big\}$$

under composition. The identity element is the identity map $I : X \to X$ given by $I(x) = x$ for all $x \in X$. For $n \in \mathbf{Z}^+$, the n^{th} **symmetric group** is the group

$$S_n = \mathrm{Perm}\,\big(\{1, 2, \cdots, n\}\big).$$

4.8 Definition: When $a_1, a_2, \cdots, a_l$ are distinct elements in $\{1, 2, \cdots, n\}$ we write

$$\alpha = (a_1, a_2, \cdots, a_l)$$

for the permutation $\alpha \in S_n$ given by

$$\alpha(a_1) = a_2 \ , \ \alpha(a_2) = a_3 \ , \ \cdots \ , \ \alpha(a_{l-1}) = a_l \ , \ \alpha(a_l) = a_1$$
$$\alpha(k) = k \text{ for all } k \notin \{a_1, a_2, \cdots, a_l\}\,.$$

Such a permutation is called a **cycle of length** l or an l**-cycle**.

4.9 Note: We make several remarks.

(1) We have $e = (1) = (2) = \cdots = (n)$.
(2) We have $(a_1, a_2, \cdots, a_l) = (a_2, a_3, \cdots, a_l, a_1) = (a_3, a_4, \cdots, a_l, a_1, a_2) = \cdots$.
(3) An l-cycle with $l \geq 2$ can be expressed *uniquely* in the form $\alpha = (a_1, a_2, \cdots, a_l)$ with $a_1 = \min\{a_1, a_2, \cdots, a_l\}$.
(4) If $\alpha = (a_1, a_2, \cdots, a_l)$ then $\alpha^{-1} = (a_l, a_{l-1}, \cdots, a_2, a_1) = (a_1, a_l, \cdots, a_3, a_2)$.
(5) If $n \geq 3$ then we have $(12)(23) = (123)$ and $(23)(12) = (132)$ so S_n is not abelian.

4.10 Definition: In S_n, given cycles α_i with $\alpha_i = (a_{i,1}, a_{i,2}, \cdots, a_{i,l_i})$, we say that the cycles α_i are **disjoint** when all the elements $a_{i,j} \in \{1, 2, \cdots, n\}$ are distinct.

4.11 Theorem: *(Cycle Notation) Every $\alpha \in S_n$ can be written as a product of disjoint cycles. Indeed every $\alpha \neq e$ can be written uniquely in the form*

$$\alpha = (a_{1,1}, \cdots, a_{1,l_1})(a_{2,1}, \cdots, a_{2,l_2}) \cdots (a_{m,1}, \cdots, a_{m,l_m})$$

with $m \geq 1$, each $l_i \geq 2$, the elements $a_{i,j}$ all distinct, each $a_{i,1} = \min\{a_{i,1}, a_{i,2}, \cdots, a_{i,l_i}\}$ and $a_{1,1} < a_{2,1} < \cdots < a_{m,1}$.

Proof: Let $e \neq \alpha \in S_n$ where $n \geq 2$. To write α in the given form, we must take $a_{1,1}$ to be the smallest element $k \in \{1, 2, \cdots, n\}$ with $\alpha(k) \neq k$. Then we must have $a_{1,2} = \alpha(a_{1,1})$, $a_{1,3} = \alpha(a_{1,2}) = \alpha^2(a_{1,1})$, and so on. Eventually we must reach l_1 such that $a_{1,1} = \alpha^{l_1}(a_{1,1})$, indeed since $\{1, 2, \cdots, n\}$ is finite, eventually we find $\alpha^i(a_{1,1}) = \alpha^j(a_{1,1})$ for some $1 \leq i < j$ and then $a_{1,1} = \alpha^{-i}\alpha^i(a_{1,1}) = \alpha^{-i}\alpha^j(a_{1,1}) = \alpha^{j-i}(a_{1,1})$. For the smallest such l_1 the elements $a_{1,1}, \cdots, a_{1,l_1}$ will be disjoint since if we had $a_{1,i} = a_{1,j}$ for some $1 \leq i < j \leq l_1$ then, as above, we would have $\alpha^{j-i}(a_{11}) = a_{11}$ with $1 \leq j - i < l_1$. This gives us the first cycle $\alpha_1 = (a_{1,1}, a_{1,2}, \cdots, a_{1,l_1})$.

If we have $\alpha = \alpha_1$ we are done. Otherwise there must be some $k \in \{1, 2, \cdots, n\}$ with $k \notin \{a_{1,1}, a_{1,2}, \cdots, a_{1,l_1}\}$ such that $\alpha(k) \neq k$, and we must choose $a_{2,1}$ to be the smallest such k. As above we obtain the second cycle $\alpha_2 = (a_{2,1}, a_{2,2}, \cdots, a_{2,l_2})$. Note that α_2 must be disjoint from α_1 because if we had $\alpha^i(a_{2,1}) = \alpha^j(a_{1,1})$ for some i, j then we would have $a_{2,1} = \alpha^{-i}\alpha^i(a_{2,1}) = \alpha^{-i}\alpha^j(a_{1,1}) = \alpha^{j-i}(a_{1,1}) \in \{a_{1,1}, \cdots, a_{1,l_1}\}$.

At this stage, if $\alpha = \alpha_1\alpha_2$ we are done, and otherwise we continue the procedure.

4.12 Definition: When a permutation $e \neq \alpha \in S_n$ is written in the unique form of the above theorem, we say that α is written in **cycle notation**. We usually write e as $e = (1)$.

4.13 Theorem: *(Even and Odd Permutations) In S_n, with $n \geq 2$,*

(1) every $\alpha \in S_n$ is a product of 2-cycles,
(2) if $e = (a_1, b_1)(a_2, b_2) \cdots (a_l, b_l)$ then l is even, that is $l = 0 \bmod 2$, and
(3) if $\alpha = (a_1, b_1)(a_2, b_2) \cdots (a_l, b_l) = (c_1, d_1)(c_2, d_2) \cdots (c_m, d_m)$ then $l = m \bmod 2$.

Solution: To prove part (1), note that given $\alpha \in S_n$ we can write α as a product of cycles, and we have

$$(a_1, a_2, \cdots, a_l) = (a_1, a_l)(a_1, a_{l-1}) \cdots (a_1, a_2)\,.$$

We shall prove part (2) by induction. First note that we cannot write e as a single 2-cycle, but we can write e as a product of two 2-cycles, for example $e = (1, 2)(1, 2)$. Fix $l \geq 3$ and suppose, inductively, that for all $k < l$, if we can write e as a product of k 2-cycles the k must be even. Suppose that e can be written as a product of l 2-cycles, say $e = (a_1, b_1)(a_2, b_2) \cdots (a_l, b_l)$. Let $a = a_1$. Of all the ways we can write e as a product of l 2-cycles, in the form $e = (x_1, y_1)(x_2, y_2) \cdots (x_l, y_l)$, with $x_i = a$ for some i, choose one way, say $e = (r_1, s_1)(r_2, s_2) \cdots (r_l, s_l)$ with $r_m = a$ and $r_i, s_i \neq a$ for all $i < m$, with m being as large as possible. Note that $m \neq l$ since for $\alpha = (r_1, s_1) \cdots (r_l, s_l)$ with $r_l = a$ and $r_i, s_i \neq a$ for $i < l$ we have $\alpha(s_l) = a \neq s_l$ and so $\alpha \neq e$. Consider the product $(r_m, s_m)(r_{m+1}, s_{m+1})$. This product must be (after possibly interchanging r_{m+1} and s_{m+1}) of one of the forms

$$(a, b)(a, b)\ ,\ (a, b)(a, c)\ ,\ (a, b)(b, c)\ ,\ (a, b)(c, d)$$

where a, b, c, d are distinct. Note that

$$\begin{aligned}
&(a, b)(a, c) = (a, c, b) = (b, c)(a, b),\\
&(a, b)(b, c) = (a, b, c) = (b, c)(a, c),\ \text{and}\\
&(a, b)(c, d) = (c, d)(a, b)\,,
\end{aligned}$$

and so in each of these three cases we could rewrite e as a product of l 2-cycles with the first occurrence of a being farther to the right, contradicting the fact that we chose m to be as large as possible. Thus the product $(r_m, s_m)(r_{m+1}, s_{m+1})$ is of the form $(a, b)(a, b)$. By cancelling these two terms, we can write e as a product of $(l - 2)$ 2-cycles. By the induction hypothesis, $(l - 2)$ is even, and so l is even.

Finally, to prove part (3), suppose that $\alpha = (a_1, b_1) \cdots (a_l, b_l) = (c_1, d_1) \cdots (c_m, d_m)$. Then we have

$$e = \alpha\, \alpha^{-1} = (a_1, b_1) \cdots (a_l, b_l)(c_m, d_m) \cdots (c_1, d_1).$$

By part (2), $l + m$ is even, and so $l = m \bmod 2$.

4.14 Definition: For $n \geq 2$, a permutation $\alpha \in S_n$ is called **even** if it can be written as a product of an even number of 2-cycles. Otherwise α can be written as a product of an odd number of 2-cycles, and then it is called **odd**. We define the **sign** (or the **parity**) of $\alpha \in S_n$ to be

$$(-1)^\alpha = \begin{cases} 1 \text{ if } \alpha \text{ is even,} \\ -1 \text{ if } \alpha \text{ is odd.} \end{cases}$$

4.15 Note: Note that $(-1)^e = 1$ and that for $\alpha, \beta \in S_n$, we have $(-1)^{\alpha\beta} = (-1)^\alpha (-1)^\beta$ and $(-1)^{\alpha^{-1}} = (-1)^\alpha$. Also note that when α is an l-cycle we have $(-1)^\alpha = (-1)^{l-1}$ because $(a_1, a_2, \cdots, a_l) = (a_1, a_2)(a_2, a_3) \cdots (a_{l-1}, a_l)$.

Multilinear Maps

4.16 Notation: Let R be a commutative ring. For positive integers $n_1, n_2, \cdots, n_k$, let

$$\prod_{i=1}^{k} R^{n_i} = \big\{(u_1, u_2, \cdots, u_k)\big|\text{each } u_i \in R^{n_i}\big\}.$$

Note that

$$M_{n\times k}(R) = \prod_{i=1}^{k} R^{n} = \big\{(u_1, u_2, \cdots, u_k)\big|\text{each } u_i \in R^{n}\big\}.$$

4.17 Definition: For a map $L : \prod_{i=1}^{k} R^{n_i} \to R^m$, we say that L is **k-linear** when for each index $j \in \{1, 2, \cdots, k\}$ and for all $u_i, v, w \in R^{n_j}$ and all $t \in R$ we have

$$\begin{aligned} L(u_1, \cdots, u_{j-1}, v + w, u_{j+1}, \cdots, u_n) &= L(u_1, \cdots, u_{j-1}, v, u_{j+1}, \cdots, u_n) \\ &\quad + L(u_1, \cdots, u_{j-1}, w, u_{j+1}, \cdots, u_n) \text{ , and} \\ L(u_1, \cdots, u_{j-1}, tv, u_{j+1}, \cdots, u_n) &= t\, L(u_1, \cdots, u_{j-1}, u_j, u_{j+1}, \cdots, u_n). \end{aligned}$$

For a k-linear map $L : M_{n\times k}(R) = \prod_{i=1}^{k} R^{n} \to R^m$ we say that L is **symmetric** when for each index $j \in \{1, 2, \cdots, k-1\}$ and for all $u_i, v, w \in R^n$ we have

$$L(u_1, \cdots, u_{j-1}, v, w, u_{j+2}, \cdots, u_n) = L(u_1, \cdots, u_{j-1}, w, v, u_{j+2}, \cdots, u_n)$$

or equivalently when for every permutation $\sigma \in S_k$ and all $u_i \in R^n$ we have

$$L(u_1, u_2, \cdots, u_k) = L(u_{\sigma(1)}, u_{\sigma(2)}, \cdots, u_{\sigma(n)}),$$

and we say that L is **skew-symmetric** when for each index $j \in \{1, 2, \cdots, k-1\}$ and for all $u_i, v, w \in R^n$ we have

$$L(u_1, \cdots, u_{j-1}, v, w, u_{j+2}, \cdots, u_n) = -\,L(u_1, \cdots, u_{j-1}, w, v, u_{j+2}, \cdots, u_n)$$

or equivalently when for every permutation $\sigma \in S_k$ and all $u_i \in R^n$ we have

$$L(u_1, u_2, \cdots, u_k) = (-1)^{\sigma} L(u_{\sigma(1)}, u_{\sigma(2)}, \cdots, u_{\sigma(n)}),$$

and we say that L is **alternating** when for each index $j \in \{1, 2, \cdots, k-1\}$ and for all $u_i, v \in R^n$ we have

$$L(u_1, \cdots, u_{j-1}, v, v, u_{j+2}, \cdots, u_n) = 0.$$

4.18 Example: As an exercise, show that for every matrix $A \in M_{m\times n}(R)$, the map $L : R^n \times R^m \to R$ given by $L(x, y) = y^T A\, x$ is 2-linear and, conversely, that given any 2-linear map $L : R^n \times R^m \to R$ there exists a unique matrix $A \in M_{m\times n}(R)$ such that $L(x, y) = y^T A\, x$ for all $x \in R^n$ and $y \in R^m$.

4.19 Theorem: *Let R be a commutative ring. Let $L : M_{n\times k} = \prod_{i=1}^{k} R^n \to R^m$ be k-linear. Then*

(1) if L is alternating then L is skew-symmetric,
(2) if L is alternating then for all indices $i, j \in \{1, 2, \cdots, k\}$ with $i < j$ and for all $u_i, v \in R^n$ we have $L(u_1, \cdots, u_{i-1}, v, u_{i+1}, \cdots, u_{j-1}, v, u_{j+1}, \cdots, u_n) = 0$, and
(3) if $2 \in R^*$ and L is skew-symmetric then L is alternating.

Proof: To prove Part (1), we suppose that L is alternating. Then for $j \in \{1, 2, \cdots, k-1\}$ and $u_i, v, w \in R^n$ we have

$$\begin{aligned} 0 &= L(u_1, \cdots, u_{j-1}, v+w, v+w, u_{j+2}, \cdots, u_n) \\ &= L(u_1, \cdots, v, v, \cdots, u_n) + L(u_1, \cdots, v, w, \cdots, u_n) \\ &\quad + L(u_1, \cdots, w, v, \cdots, u_n) + L(u_1, \cdots, w, w, \cdots, u_n) \\ &= L(u_1, \cdots, v, w, \cdots, u_n) + L(u_1, \cdots, w, v, \cdots, u_n) \end{aligned}$$

and so $L(u_1, \cdots, v, w, \cdots, u_n) = -\,L(u_1, \cdots, w, v, \cdots, u_n)$, hence L is skew-symmetric.

To prove Part (2) we again suppose that L is alternating. Then, as shown immediately above, L is also skew-symmetric and so for indices $i, j \in \{1, 2, \cdots, k\}$ with $i < j$ and for $u_i, v \in R^n$, in the case that $j > i+1$ we have

$$\begin{aligned} L(u_1, \cdots, u_{i-1}, &v, w, u_{i+2}, \cdots, u_{j-1}, v, u_{j+1}, \cdots, u_n) \\ &= -\,L(u_1, \cdots, u_{i-1}, v, v, u_{i+2}, \cdots, u_{j-1}, w, u_{j+1} \cdots, u_n) = 0. \end{aligned}$$

Finally, to prove Part (3), suppose that $2 \in R^*$ and that L is skew-symmetric. Then for an index $j \in \{1, 2, \cdots, k-1\}$ and for $u_i, v \in R^n$ we have

$$L(u_1, \cdots, u_{j-1}, v, v, u_{j+2}, \cdots, u_n) = -\,L(u_1, \cdots, u_{j-1}, v, v, u_{j+2}, \cdots, u_n)$$

and so $2\,L(u_1, \cdots, u_{j-1}, v, v, u_{j+2}, \cdots, u_n) = 0$. Since $2 \in R^*$ we can multiply both sides by 2^{-1} to get $L(u_1, \cdots, u_{j-1}, v, v, u_{j+2}, \cdots, u_k) = 0$.

4.20 Theorem: *Let R be a commutative ring. Given $c \in R$ there exists a unique alternating n-linear map $L : M_n(R) = \prod_{i=1}^{n} R^n \to R$ such that $L(I) = L(e_1, e_2, \cdots, e_n) = c$. This unique map L is given by*

$$L(A) = c \cdot \sum_{\sigma \in S_n} (-1)^\sigma A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n} \text{ , that is}$$

$$L(u_1, u_2, \cdots, u_n) = c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} (u_2)_{\sigma(2)} \cdots (u_n)_{\sigma(n)}.$$

Proof: First we prove uniqueness. Suppose that $L : M_n(R) = \prod_{i=1}^{n} R^n \to R$ is alternating and n-linear with $L(I) = c$. Then for all $u_i \in R^n$ we have

$$\begin{aligned} L(u_1, u_2, \cdots, u_n) &= L\Big(\sum_{i_1=1}^{n} (u_1)_{i_1} e_{i_1} \,,\, \sum_{i_2=1}^{n} (u_2)_{i_2} e_{i_2} \,, \cdots, \sum_{i_n=1}^{n} (u_n)_{i_n} e_{i_n} \Big) \\ &= \sum_{i_1, i_2, \cdots, i_n = 1}^{n} (u_1)_{i_1} (u_2)_{i_2} \cdots (u_n)_{i_n} L(e_{i_1}, e_{i_2}, \cdots, e_{i_n}). \end{aligned}$$

Note that because L is alternating, whenever we have $e_{i_j} = e_{i_k}$ for some $j \neq k$, we obtain $L(e_{i_1}, e_{i_2}, \cdots, e_{i_n}) = 0$, and so the only nonzero terms in the above sum occur when

$i_1, i_2, \cdots, i_n$ are distinct, so there is a permutation $\sigma \in S_n$ with $i_j = \sigma(j)$ for all j. Thus

$$\begin{aligned} L(u_1, u_2, \cdots, u_n) &= \sum_{\sigma \in S_n} (u_1)_{\sigma(1)} (u_2)_{\sigma(2)} \cdots (u_n)_{\sigma(n)} L(e_{\sigma(1)}, e_{\sigma(2)}, \cdots, e_{\sigma(n)}) \\ &= \sum_{\sigma \in S_n} (u_1)_{\sigma(1)} (u_2)_{\sigma(2)} \cdots (u_n)_{\sigma(n)} (-1)^\sigma L(e_1, e_2, \cdots, e_n) \\ &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} (u_2)_{\sigma(2)} \cdots (u_n)_{\sigma(n)} \end{aligned}$$

This proves that there is a unique such map L and that it is given by the required formula.

To prove existence, it suffices to show that the map $L : M_n(R) = \prod_{i=1}^{n} R^n \to R$ given by the formula

$$L(u_1, u_2, \cdots, u_n) = c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} (u_2)_{\sigma(2)} \cdots (u_n)_{\sigma(n)}.$$

is n-linear and alternating with $L(I) = c$. Note that this map L is n-linear because

$$\begin{aligned} L(u_1, \cdots, v + w, \cdots, u_n) &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} \cdots (v + w)_{\sigma(j)}, \cdots (u_n)_{\sigma(n)} \\ &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} \cdots v_{\sigma(j)} \cdots (u_n)_{\sigma(n)} + c \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} \cdots w_{\sigma(j)} \cdots (u_n)_{\sigma(n)} \\ &= L(u_1, \cdots, v, \cdots, u_n) + L(u_1, \cdots, w, \cdots, u_n) \end{aligned}$$

and similarly $L(u_1, \cdots, tv, \cdots, u_n) = t\, L(u_1, \cdots, v, \cdots, u_n)$.

Note that L is alternating because, given indices $i, j \in \{1, 2, \cdots, n\}$ with $i < j$, when $u_i = u_j = v$ we have

$$\begin{aligned} L(u_1, \cdots, v, \cdots, v, \cdots, u_n) &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (u_1)_{\sigma(1)} \cdots v_{\sigma(i)} \cdots v_{\sigma(j)} \cdots (u_n)_{\sigma(n)} \\ &= c \cdot \sum_{\sigma \in S_n, \sigma(i) < \sigma(j)} (-1)^\sigma (u_1)_{\sigma(1)} \cdots v_{\sigma(i)} \cdots v_{\sigma(j)} \cdots (u_n)_{\sigma(n)} \\ &\quad + c \cdot \sum_{\tau \in S_n, \tau(i) > \tau(j)} (-1)^\tau (u_1)_{\tau(1)} \cdots v_{\tau(i)} \cdots v_{\tau(j)} \cdots (u_n)_{\tau(n)}. \end{aligned}$$

This is equal to 0 because the term in the first sum labeled by $\sigma \in S_n$ with $\sigma(i) < \sigma(j)$ can be paired with the term in the second sum labeled by $\tau = \sigma(i, j)$ (where $\sigma(i, j)$ denotes the composite of σ with the 2-cycle (i, j)), and then the sum of the two terms in each pair is equal to 0 because $(-1)^\tau = -(-1)^\sigma$.

Finally note that

$$\begin{aligned} L(e_1, e_2, \cdots, e_n) &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma (e_1)_{\sigma(1)} (e_2)_{\sigma(2)} \cdots (e_n)_{\sigma(n)} \\ &= c \cdot \sum_{\sigma \in S_n} (-1)^\sigma \delta_{1,\sigma(1)} \delta_{2,\sigma(2)} \cdots \delta_{n,\sigma(n)} = c \end{aligned}$$

because the only nonzero term in the sum occurs when $\sigma = e$.

The Determinant

4.21 Definition: Let R be a commutative ring. The unique alternating n-linear map $\det : M_n(R) \to R$ with $\det(I) = 1$ is called the **determinant** map. For $A \in M_n(R)$, the **determinant** of A, denoted by $|A|$ or by $\det(A)$, is given by

$$|A| = \det(A) = \sum_{\sigma \in S_n} (-1)^\sigma A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n}.$$

4.22 Example: As an exercise, find an explicit formula for the determinant of a 2×2 matrix and for the determinant of a 3×3 matrix.

4.23 Note: Given $c \in R$, according to the above theorem, the unique alternating n-linear map $L : M_n(R) \to R$ with $L(I) = c$ is given by $L(A) = c\,|A|$.

4.24 Theorem: *Let R be a commutative ring and let $A, B \in M_n(R)$. Then*
(1) $|A^T| = |A|$, and
(2) $|AB| = |A|\,|B|$.

Proof: To prove Part (1) note that

$$\begin{aligned}
|A| &= \sum_{\sigma \in S_n} (-1)^\sigma A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n} \\
&= \sum_{\sigma \in S_n} (-1)^\sigma A_{1,\sigma^{-1}(1)} A_{2,\sigma^{-1}(2)} \cdots A_{n,\sigma^{-1}(n)} \\
&= \sum_{\tau \in S_n} (-1)^\tau A_{1,\tau(1)} A_{2,\tau(2)} \cdots A_{n,\tau(n)} \\
&= \sum_{\tau \in S_n} (-1)^\tau (A^T)_{\tau(1),1} (A^T)_{\tau(2),2} \cdots (A^T)_{\tau(n),n} = |A^T|.
\end{aligned}$$

To prove Part (2), fix a matrix $A \in M_n(R)$ and define $L : M_r(R) \to R$ by $L(B) = |AB|$. Note that L is n-linear because

$$\begin{aligned}
L(u_1, \cdots, v + w, \cdots, u_n) &= \big|A(u_1, \cdots, v + w, \cdots, u_n)\big| \\
&= \big|(Au_1, \cdots, A(v + w), \cdots, Au_n)\big| \\
&= \big|(Au_1, \cdots, Av + Aw, \cdots, Au_n)\big| \\
&= \big|(Au_1, \cdots, Av, \cdots, Au_n)\big| + \big|(Au_1, \cdots, Aw, \cdots, Au_n)\big| \\
&= \big|A(u_1, \cdots, v, \cdots, u_n)\big| + \big|A(u_1, \cdots, w, \cdots, u_n)\big| \\
&= L(u_1, \cdots, v, \cdots, u_n) + L(u_1, \cdots, w, \cdots, u_n).
\end{aligned}$$

and similarly $L(u_1, \cdots, tv, \cdots, u_n) = t\,L(u_1, \cdots, v, \cdots, u_n)$. Note that L is alternating because

$$\begin{aligned}
L(u_1, \cdots, v, v, \cdots, u_n) &= \big|A(u_1, \cdots, v, v, \cdots, u_n)\big| \\
&= \big|(Au_1, \cdots, Av, Av, \cdots, Au_n)\big| = 0.
\end{aligned}$$

Note that $L(I) = \big|AI\big| = |A|$. Thus, by Theorem 4.20 (see Note 4.23) it follows that $L(B) = L(I)|B| = |A|\,|A|$.

4.25 Definition: Let R be a commutative ring and let $A \in M_n(R)$. We say that A is **upper triangular** when $A_{j,k} = 0$ for all $j > k$, and we say that A is **lower-triangular** when $A_{j,k} = 0$ for all $j < k$.

4.26 Theorem: *Let R be a commutative ring and let $A, B \in M_n(R)$.*

(1) If B is obtained from A by performing an elementary column operation then $|B|$ is obtained from $|A|$ as follows.
(a) if we use $C_k \leftrightarrow C_l$ with $k \neq l$ then $|B| = -|A|$,
(b) if we use $C_k \mapsto t\,C_k$ with $t \in R$ then $|B| = t\,|A|$, and
(c) if we use $C_k \mapsto C_k + tC_l$ with $t \in R$ and $k \neq l$ then $|B| = |A|$.
The same rules apply when B is obtained from A using an elementary row operation.

(2) If A is either upper-triangular or lower-triangular then $|A| = \prod_{i=1}^{n} A_{i,i}$.

Proof: If B is obtained from A using the column operation $C_k \leftrightarrow C_l$ with $k \neq l$ then $|B| = -|A|$ because the determinant map is skew-symmetric. If B is obtained from A using $C_k \mapsto t\,C_k$ with $t \in R$ then $|B| = t|A|$ because the determinant map is linear. Suppose B is obtained from A using $C_k \mapsto C_k + t\,C_l$ where $t \in R$ and $k \neq l$. Write $A = (u_1, u_2 \cdots, u_n)$ with each $u_i \in R^n$. Then since the determinant map is n-linear and alternating we have

$$\begin{aligned} |B| &= \big|(u_1, \cdots, u_k + tu_l, \cdots, u_l, \cdots, u_n)\big| \\ &= \big|(u_1, \cdots, u_k, \cdots, u_l, \cdots, u_n)\big| + t\,\big|(u_1, \cdots, u_l, \cdots, u_l, \cdots, u_n)\big| \\ &= |A| + t \cdot 0 = |A|. \end{aligned}$$

This proves Part (1) in the case of column operations. The same rules apply when using row operations because $|A^T| = |A|$.

To prove Part (2), suppose that A is upper-triangular (the case that A is lower-triangular is similar). We claim that for every $\sigma \in S_n$ with $\sigma \neq e$ we have $\sigma(i) > i$, hence $A_{\sigma(i),i} = 0$, for some $i \in \{1, 2, \cdots, n\}$. Suppose, for a contradiction, that $\sigma \neq e$ and $\sigma(i) \leq i$ for all indices i. Let k be the largest index for which $\sigma(k) < k$. Then we have $\sigma(i) = i > k$ for all $i > k$ and $\sigma(k) < k$ and $\sigma(i) \leq i < k$ for all $i < k$. This implies that there is no index i for which $\sigma(i) = k$, but this is not possible since σ is surjective. This proves the claim. Thus $|A| = \sum_{\sigma \in S_n} (-1)^{\sigma} A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n} = \prod_{i=1}^{n} A_{i,i}$ because the only nonzero term in the above sum occurs when $\sigma = e$.

4.27 Example: The above theorem gives us a method that we can use to calculate determinants. For example, using only row operations of the form $R_k \to R_k + t\,R_l$ we have

$$\begin{vmatrix} 1 & 3 & 2 & 4 \\ 2 & 4 & 1 & 2 \\ 3 & 5 & 4 & 1 \\ 1 & 1 & 5 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 & 4 \\ 0 & -2 & -3 & -6 \\ 0 & -4 & -2 & -11 \\ 0 & -2 & 3 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 & 4 \\ 0 & 2 & -1 & 5 \\ 0 & -4 & -2 & -11 \\ 0 & -2 & 3 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 & 4 \\ 0 & 2 & -1 & 5 \\ 0 & 0 & -4 & -1 \\ 0 & 0 & 2 & 4 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 3 & 2 & 4 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & 2 & 11 \\ 0 & 0 & 2 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 & 4 \\ 0 & 2 & -1 & 5 \\ 0 & 0 & 2 & 11 \\ 0 & 0 & 0 & -7 \end{vmatrix} = -28.$$

4.28 Definition: Let R be a commutative ring and let $A \in M_n(R)$ with $n \geq 2$. We write $A^{(j,k)}$ to denote the $(n-1) \times (n-1)$ matrix which is obtained by removing the j^{th} row and the k^{th} column of A. The **cofactor matrix** of A is the matrix $\text{Cof}(A) \in M_n(R)$ with entries

$$\text{Cof}(A)_{k,l} = (-1)^{k+l} \big|A^{(l,k)}\big|.$$

4.29 Theorem: *Let R be a commutative ring and let $A \in M_n(R)$ with $n \geq 2$.*
(1) For each $k \in \{1, 2, \cdots, n\}$ we have

$$|A| = \sum_{j=1}^{n} (-1)^{j+k} A_{j,k} \big| A^{(j,k)} \big| = \sum_{j=1}^{n} (-1)^{k+j} A_{k,\,j} \big| A^{(k,\,j)} \big|.$$

(2) We have

$$A \cdot \mathrm{Cof}(A) = |A| \cdot I = \mathrm{Cof}(A) \cdot A.$$

(3) A is invertible in $M_n(R)$ if and only if $|A|$ is invertible in R, and in this case we have $|A^{-1}| = |A|^{-1}$ and

$$A^{-1} = \tfrac{1}{|A|} \mathrm{Cof}(A).$$

(4) If A is invertible then the unique solution to the equation $Ax = b$ is the element $x \in R^n$ with entries

$$x_k = \frac{|B_k|}{|A|}$$

where B_k is the matrix obtained by replacing the k^{th} column of A by b.

Proof: We have

$$\begin{aligned}
|A| &= \sum_{\sigma \in S_n} (-1)^{\sigma} A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n} \\
&= \sum_{j=1}^{n} \sum_{\sigma \in S_n, \sigma(k)=j} (-1)^{\sigma} A_{\sigma(1),1} \cdots A_{\sigma(k-1),k-1} A_{j,k} A_{\sigma(k+1),k+1} \cdots A_{\sigma(n),n} \\
&= \sum_{j=1}^{n} A_{j,k} \sum_{\sigma \in S_n, \sigma(k)=j} (-1)^{\sigma} {A^{(j,k)}}_{\tau(1),1} \cdots {A^{(j,k)}}_{\tau(k-1),k-1} {A^{(j,k)}}_{\tau(k),k} \cdots {A^{(j,k)}}_{\tau(k-1),k-1}
\end{aligned}$$

where $\tau = \tau(\sigma) \in S_{n-1}$ is the permutation defined as follows:

$$\text{if } i < k \;\; \tau(i) = \begin{cases} \sigma(i) & \text{if } \sigma(i) < j, \\ \sigma(i) - 1 & \text{if } \sigma(i) > j, \end{cases} \quad \text{and if } i > k \;\; \tau(i-1) = \begin{cases} \sigma(i) & \text{if } \sigma(i) < j, \\ \sigma(i) - 1 & \text{if } \sigma(i) > j, \end{cases}$$

or equivalently, τ is the composite

$$\tau = (n, n-1, \cdots, j+1, j)\, \sigma\, (k, k+1, \cdots, n-1, n).$$

Note that $(-1)^{\tau} = (-1)^{n-j} (-1)^{\sigma} (-1)^{n-k}$ and so we have $(-1)^{\sigma} = (-1)^{j+k} (-1)^{\tau}$. Thus

$$\begin{aligned}
|A| &= \sum_{j=1}^{n} A_{j,k} \sum_{\tau \in S_{n-1}} (-1)^{j+k} (-1)^{\tau} {A^{(j,k)}}_{\tau(1),1} \cdots {A^{(j,k)}}_{\tau(n-1),n-1} \\
&= \sum_{j=1}^{n} (-1)^{j+k} A_{j,k} \big| A^{(j,k)} \big|.
\end{aligned}$$

The proof that $|A| = \sum_{j=1}^{n} (-1)^{k+j} A_{k,j} \big| A^{(k,j)} \big|$ is similar (or it follows from the formula $|A^T| = |A|$). This completes the proof of Part (1).

To prove Part (2) we note that

$$\Big(\mathrm{Cof}(A)\cdot A\Big)_{k,l} = \sum_{j=1}^{n} \mathrm{Cof}(A)_{k,j}A_{j,l} = \sum_{j=1}^{n}(-i)^{k+j}A_{j,l}\left|A^{(j,k)}\right|.$$

By Part (1), the sum on the right is equal to the determinant of the matrix $B^{(k,l)} \in M_n(R)$ which is obtained from A by replacing the k^{th} column of A by a copy of its l^{th} column. Since $B^{(k,l)}$ is equal to A when $k = l$, and $B^{(k,l)}$ has two equal columns when $k \neq l$ we have

$$\Big(\mathrm{Cof}(A)\cdot A\Big)_{k,l} = \left|B^{(k,l)}\right| = \left\{\begin{array}{c} |A| \text{ if } k = l, \\ 0 \text{ if } k \neq l \end{array}\right\} = |A|\cdot\delta_{k,l}.$$

This proves that $\mathrm{Cof}(A)\cdot A = |A|\cdot I$. A similar proof shows that $A\cdot\mathrm{Cof}(A) = |A|\cdot I$.

If A is invertible in $M_n(R)$, then we have $|A|\,|A^{-1}| = |A\cdot A^{-1}| = |I| = 1$ and similarly $|A^{-1}|\,|A| = 1$ and so $|A|$ and $|A^{-1}|$ are invertible in R with $|A^{-1}| = |A|$. Conversely, the formulas in Part (2) show that if $|A|$ is invertible in R then A is invertible in $M_n(R)$ with $A^{-1} = \frac{1}{|A|}\mathrm{Cof}(A)$. This proves Part (3).

Part (4) now follows from Parts (1) and (3). Indeed if A is invertible then the solution to $Ax = b$ is given by $x = A^{-1}b$ and so

$$\begin{aligned} x_k = \big(A^{-1}b\big)_k &= \tfrac{1}{|A|}\big(\mathrm{Cof}(A)\,b\big)_k = \tfrac{1}{|A|}\sum_{j=1}^{n}\mathrm{Cof}(A)_{k,j}b_j \\ &= \tfrac{1}{|A|}\sum_{j=1}^{n}(-1)^{k+j}b_j\left|A^{(j,k)}\right| = \tfrac{1}{|A|}\,|B_k| \end{aligned}$$

where B_k is the matrix obtained by replacing the k^{th} column of A by b.

4.30 Definition: For a matrix $A \in M_n(R)$, the first of the two sums in Part (1) of the above theorem is called the **cofactor expansion** of $|A|$ **along the** k^{th} **column** of A, and the second sum is called the **cofactor expansion** of $|A|$ **along the** k^{th} **row** of A.

4.31 Example: Using row operations of the form $R_k \mapsto R_k + t\,R_l$, together with cofactor expansions along various columns, we have

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

$$= -\begin{vmatrix} 4 & 4 & 4 \\ 2 & 1 & 2 \\ 2 & 4 & 6 \end{vmatrix} = -\begin{vmatrix} -4 & 0 & -4 \\ 2 & 1 & 2 \\ -6 & 0 & -2 \end{vmatrix} = -\begin{vmatrix} -4 & -4 \\ -6 & -2 \end{vmatrix} = -\begin{vmatrix} 8 & 0 \\ -6 & -2 \end{vmatrix} = 16.$$

4.32 Example: From the formula in Part (2), if $ad - bc \neq 0$ then we have

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

As an exercise, find a similar formula for the inverse of a 3×3 matrix.

4.33 Corollary: *Let R be a commutative ring and let $A \in M_n(R)$. If A is invertible in $M_n(R)$ then $|A|$ is invertible in R and we have $|A^{-1}| = |A|^{-1}$.*

Chapter 5. The Dot and Cross Products in $\mathbf{R}^n$

5.1 Definition: Let F be a field. For vectors $x, y \in F^n$ we define the **dot product** of x and y to be

$$x \cdot y = y^T x = \sum_{i=1}^{n} x_i y_i \in F\,.$$

5.2 Theorem: *(Properties of the Dot Product) For all $x, y, z \in \mathbf{R}^n$ and all $t \in \mathbf{R}$ we have*
(1) (Bilinearity) $(x+y)\cdot z = x\cdot z + y\cdot z\ ,\ (tx)\cdot y = t(x\cdot y)$
$x\cdot(y+z) = x\cdot y + x\cdot z\ ,\ x\cdot(ty) = t(x\cdot y)$,
(2) (Symmetry) $x\cdot y = y\cdot x$, *and*
(3) (Positive Definiteness) $x\cdot x \geq 0$ *with* $x\cdot x = 0$ *if and only if* $x = 0$.

Proof: The proof is left as an exercise.

5.3 Definition: For a vector $x \in \mathbf{R}^n$, we define the **length** (or **norm**) of x to be

$$|x| = \sqrt{x\cdot x} = \sqrt{\sum_{i=1}^{n} {x_i}^2}.$$

We say that x is a **unit vector** when $|x| = 1$.

5.4 Theorem: *(Properties of Length) Let $x, y \in \mathbf{R}^n$ and let $t \in \mathbf{R}$. Then*
(1) (Positive Definiteness) $|x| \geq 0$ *with* $|x| = 0$ *if and only if* $x = 0$,
(2) (Scaling) $|tx| = |t||x|$,
(3) $|x \pm y|^2 = |x|^2 \pm 2(x\cdot y) + |y|^2$.
(4) (The Polarization Identities) $x\cdot y = \frac{1}{2}\big(|x+y|^2 - |x|^2 - |y|^2\big) = \frac{1}{4}\big(|x+y|^2 - |x-y|^2\big)$,
(5) (The Cachy-Schwarz Inequality) $|x\cdot y| \leq |x|\,|y|$ *with* $|x\cdot y| = |x|\,|y|$ *if and only if the set $\{x, y\}$ is linearly dependent, and*
(6) (The Triangle Inequality) $|x+y| \leq |x| + |y|$.

Proof: We leave the proofs of Parts (1), (2) and (3) as an exercise, and we note that (4) follows immediately from (3). To prove part (5), suppose first that $\{x, y\}$ is linearly dependent. Then one of x and y is a multiple of the other, say $y = tx$ with $t \in \mathbf{R}$. Then

$$|x\cdot y| = |x\cdot(tx)| = |t(x\cdot x)| = |t|\,|x|^2 = |x|\,|tx| = |x|\,|y|.$$

Suppose next that $\{x, y\}$ is linearly independent. Then for all $t \in \mathbf{R}$ we have $x + ty \neq 0$ and so

$$0 \neq |x+ty|^2 = (x+ty)\cdot(x+ty) = |x|^2 + 2t(x\cdot y) + t^2|y|^2.$$

Since the quadratic on the right is non-zero for all $t \in \mathbf{R}$, it follows that the discriminant of the quadratic must be negative, that is

$$4(x\cdot y)^2 - 4|x|^2|y|^2 < 0.$$

Thus $(x\cdot y)^2 < |x|^2|y|^2$ and hence $|x\cdot y| < |x|\,|y|$. This proves part (5).

Using part (5) note that

$$|x+y|^2 = |x|^2 + 2(x\cdot y) + |y|^2 \leq |x+y|^2 + 2|x\cdot y| + |y|^2 \leq |x|^2 + 2|x|\,|y| + |y|^2 = \big(|x| + |y|\big)^2$$

and so $|x+y| \leq |x| + |y|$, which proves part (6).

5.5 Definition: For points $a, b \in \mathbf{R}^n$, we define the **distance** between a and b to be

$$\operatorname{dist}(a, b) = |b - a|.$$

5.6 Theorem: *(Properties of Distance) Let $a, b, c \in \mathbf{R}^n$. Then*

(1) (Positive Definiteness) $\operatorname{dist}(a, b) \geq 0$ *with* $\operatorname{dist}(a, b) = 0$ *if and only if* $a = b$,
(2) (Symmetry) $\operatorname{dist}(a, b) = \operatorname{dist}(b, a)$, *and*
(3) (The Triangle Inequality) $\operatorname{dist}(a, c) \leq \operatorname{dist}(a, b) + \operatorname{dist}(b, c)$.

Proof: The proof is left as an exercise.

5.7 Definition: For nonzero vectors $0 \neq x, y \in \mathbf{R}^n$, we define the **angle** between x and y to be

$$\theta(x, y) = \cos^{-1}\left(\frac{x \bullet y}{|x|\,|y|}\right) \in [0, \pi].$$

Note that $\theta(x, y) = \frac{\pi}{2}$ if and only if $x \bullet y = 0$. For vectors $x, y \in \mathbf{R}^n$, we say that x and y are **orthogonal** when $x \bullet y = 0$.

5.8 Theorem: *(Properties of Angle) Let $0 \neq x, y \in \mathbf{R}^n$. Then*

(1) $\theta(x, y) \in [0, \pi]$ *with* $\begin{cases} \theta(x, y) = 0 \text{ if and only if } y = tx \text{ for some } t > 0, \text{ and} \\ \theta(x, y) = \pi \text{ if and only if } y = tx \text{ for some } t < 0, \end{cases}$
(2) (Symmetry) $\theta(x, y) = \theta(y, x)$,
(3) (Scaling) $\theta(tx, y) = \theta(x, ty) = \begin{cases} \theta(x, y) & \text{if } 0 < t \in \mathbf{R}, \\ \pi - \theta(x, y) & \text{if } 0 > t \in \mathbf{R}, \end{cases}$
(4) (The Law of Cosines) $|y - x|^2 = |x|^2 + |y|^2 - 2|x|\,|y| \cos\theta(x, y)$,
(5) (Pythagoras' Theorem) $\theta(x, y) = \frac{\pi}{2}$ *if and only if* $|y - x|^2 = |x|^2 + |y|^2$, *and*
(6) (Trigonometric Ratios) if $(y - x) \bullet x = 0$ *then* $\cos\theta(x, y) = \frac{|x|}{|y|}$ *and* $\sin\theta(x, y) = \frac{|y-x|}{|y|}$.

Proof: The Law of Cosines follows from the identity $|y - x|^2 = |y|^2 - 2(y \bullet x) + |x|^2$ and the definition of $\theta(x, y)$. Pythagoras' Theorem is a special case of the Law of Cosines. We Prove Part (6). Let $0 \neq x, y \in \mathbf{R}^n$ and write $\theta = \theta(x, y)$. Suppose that $(y - x) \bullet x = 0$. Then we have $y \bullet x - x \bullet x = 0$ so that $x \bullet y = |x|^2$, and so we have

$$\cos\theta = \frac{x \bullet y}{|x|\,|y|} = \frac{|x|^2}{|x|\,|y|} = \frac{|x|}{|y|}\,.$$

Also, by Pythagoras' Theorem we have $|x|^2 + |y - x|^2 = |y|^2$ so that $|y|^2 - |x|^2 = |y - x|^2$, and so

$$\sin^2\theta = 1 - \cos^2\theta = 1 - \frac{|x|^2}{|y|^2} = \frac{|y|^2 - |x|^2}{|y|^2} = \frac{|y - x|^2}{|y|^2}\,.$$

Since $\theta \in [0, \pi]$ we have $\sin\theta \geq 0$, and so taking the square root on both sides gives

$$\sin\theta = \frac{|y - x|}{|y|}\,.$$

5.9 Definition: For points $a, b, c \in \mathbf{R}^n$ with $a \neq b$ and $b \neq c$ we define

$$\angle abc = \theta(b - a, c - b).$$

Orthogonal Complement and Orthogonal Projection in $\mathbf{R}^n$

5.10 Definition: Let F be a field and let U, V and W be subspaces of F^n. Recall that

$$U+V=\big\{u+v\big|u\in U, v\in V\big\}$$

is a subspace of F^n. We say that W is the **internal direct sum** of U with V, and we write $W=U\oplus V$, when $W=U+V$ and $U\cap V=\{0\}$. As an exercise, show that $W=U\oplus V$ if and only if for every $x\in W$ there exist unique vectors $u\in U$ and $v\in V$ with $x=u+v$.

5.11 Definition: Let $U\subseteq\mathbf{R}^n$ be a subspace. We define the **orthogonal complement** of U in $\mathbf{R}^n$ to be

$$U^\perp=\big\{x\in\mathbf{R}^n\big|x\bullet u=0 \text{ for all } u\in U\big\}.$$

5.12 Theorem: *(Properties of the Orthogonal Complement) Let $U\subseteq\mathbf{R}^n$ be a subspace, let $S\subseteq U$ and let $A\in M_{k\times n}(\mathbf{R})$. Then*
(1) If $U=\text{Span}\,(S)$ then $U^\perp=\big\{x\in\mathbf{R}^n\big|x\cdot u=0$ for all $u\in S\big\}$,
(2) $(\text{Row}A)^T=\text{Null}A$.
(3) $U^\perp$ is a vector space,
(4) $\dim(U)+\dim(U^\perp)=n$
(5) $U\oplus U^\perp=\mathbf{R}^n$,
(6) $(U^\perp)^\perp=U$,
(7) $(\text{Null}A)^\perp=\text{Row}A$.

Proof: To prove part (1), let $T=\big\{x\in\mathbf{R}^n\big|x\bullet u=0 \text{ for all } u\in S\big\}$. Note that $U^\perp\subseteq T$. Let $x\in T$. Let $u\in U=\text{Span}\,(S)$, say $u=\sum_{i=1}^{n}t_iu_i$ with each $t_i\in\mathbf{R}$ and each $u_i\in S$. Then $x\bullet u=x\bullet\sum_{i=1}^{n}t_iu_i=\sum_{i=1}^{n}t_i(x\bullet u_i)=0$. Thus $x\in U^\perp$ and so we have $T\subseteq U^\perp$.

To prove part (2), let $v_1,v_2,\cdots,v_n$ be the rows of A. Note that $Ax=\begin{pmatrix}x\bullet v_1\\ \vdots\\ x\bullet v_n\end{pmatrix}$ so we have $x\in\text{Null}A\iff x\bullet v_i=0$ for all $i\iff x\in\text{Span}\,\{v_1,v_2,\cdots,v_k\}^\perp=(\text{Row}A)^\perp$ by part (1).

Part (3) follows from Part (2) since we can choose the matrix A so that $U=\text{Row}(A)$ and then we have $U^\perp=\text{Null}(A)$ which is a vector space in $\mathbf{R}^n$.

Part (4) also follows from part (2) since if we choose A so that $\text{Row}A=U$ then we have $\dim(U)+\dim(U^\perp)=\dim\text{Row}A+\dim(\text{Row}A)^\perp=\dim\text{Row}A+\dim\text{Null}A=n$.

To prove part (5), in light of part (4), it suffices to show that $U\cap U^\perp=\{0\}$. Let $x\in U\cap U^\perp$. Since $x\in U^\perp$ we have $x\bullet u=0$ for all $u\in U$. In particular, since $x\in U$ we have $x\bullet x=0$, and hence $x=0$. Thus $U\cap U^\perp=\{0\}$ and so $U\oplus U^\perp=\mathbf{R}^n$.

To prove part (6), let $x\in U$. By the definition of $U^\perp$ we have $x\bullet v=0$ for all $v\in U^\perp$. By the definition of $(U^\perp)^\perp$ we see that $x\in(U^\perp)^\perp$. Thus $U\subseteq(U^\perp)^\perp$. By part (4) we know that $\dim U+\dim U^\perp=n$ and also that $\dim U^\perp+\dim(U^\perp)^\perp=n$. It follows that $\dim U=n-\dim U^\perp=\dim(U^\perp)^\perp$. Since $U\subseteq(U^\perp)^\perp$ and $\dim U=\dim(U^\perp)^\perp$ we have $U=(U^\perp)^\perp$, as required.

By parts (3) and (6) we have $(\text{Null}A)^\perp=\big((\text{Row}A)^\perp\big)^\perp=\text{Row}A$, proving part (7).

5.13 Definition: For a subspace $U \subseteq \mathbf{R}^n$ and a vector $x \in \mathbf{R}^n$, we define the **orthogonal projection** of x onto U, denoted by $\mathrm{Proj}_U(x)$, as follows. Since $\mathbf{R}^n = U \oplus U^\perp$, we can choose unique vectors $u, v \in \mathbf{R}^n$ with $u \in U$, $v \in U^\perp$ and $x = u + v$. We then define

$$\mathrm{Proj}_U(x) = u.$$

Note that since $U = (U^\perp)^\perp$, for u and v as above we have $\mathrm{Proj}_{U^\perp}(x) = v$. When $y \in \mathbf{R}^n$ and $U = \mathrm{Span}\,\{y\}$, we also write $\mathrm{Proj}_y(x) = \mathrm{Proj}_U(x)$ and $\mathrm{Proj}_{y^\perp}(x) = \mathrm{Proj}_{U^\perp}(x)$.

5.14 Theorem: *Let $U \subseteq \mathbf{R}^n$ be a subspace and let $x \in \mathbf{R}^n$. Then $\mathrm{Proj}_U(x)$ is the unique point in U which is nearest to x.*

Proof: Let $u, v \in \mathbf{R}^n$ with $u \in U$, $v \in V$ and $u + v = x$ so that $\mathrm{Proj}_U(x) = u$. Let $w \in U$ with $w \neq u$. Since $v \in U^\perp$ and $u, w \in U$ we have $v \cdot u = v \cdot w = 0$ and so $v \cdot (w - u) = v \cdot w - v \cdot u = 0$. Thus we have

$$\begin{aligned} |x - w|^2 = |u + v - w|^2 = |v - (w - u)|^2 &= \big(v - (w - u)\big) \cdot \big(v - (w - u)\big) \\ &= |v|^2 - 2\, v \cdot (w - u) + |w - u|^2 = |v|^2 + |w - u|^2 = |x - u|^2 + |w - u|^2 . \end{aligned}$$

Since $w \neq u$ we have $|w - u| > 0$ and so $|x - w|^2 > |x - u|^2$. Thus $|x - w| > |x - u|$, that is $\mathrm{dist}(x, w) > \mathrm{dist}(x, u)$, so u is the vector in U nearest to x, as required.

5.15 Theorem: *For any matrix $A \in M_{n\times l}(\mathbf{R})$ we have $\mathrm{Null}(A^TA) = \mathrm{Null}(A)$ and $\mathrm{Col}(A^TA) = \mathrm{Col}(A^T)$ so that $\mathrm{nullity}(A^TA) = \mathrm{nullity}(A)$ and $\mathrm{rank}(A^TA) = \mathrm{rank}(A)$.*

Proof: If $x \in \mathrm{Null}(A)$ then $Ax = 0$ so $A^TAx = 0$ hence $x \in \mathrm{Null}(A^TA)$. This shows that $\mathrm{Null}(A) \subseteq \mathrm{Null}(A^TA)$. If $x \in \mathrm{Null}(A^TA)$ then we have $A^TAx = 0$ which implies that $|Ax|^2 = (Ax)^T(Ax) = x^TA^TAx = 0$ and so $Ax = 0$. This shows that $\mathrm{Null}(A^TA) \subseteq \mathrm{Null}(A)$. Thus we have $\mathrm{Null}(A^TA) = \mathrm{Null}(A)$. It then follows that

$$\mathrm{Col}(A^T) = \mathrm{Row}(A) = \mathrm{Null}(A)^\perp = \mathrm{Null}(A^TA)^\perp = \mathrm{Row}(A^TA) = \mathrm{Col}\big((A^TA)^T\big) = \mathrm{Col}(A^TA).$$

5.16 Theorem: *Let $A \in M_{n\times l}(\mathbf{R})$, let $U = \mathrm{Col}(A)$ and let $x \in \mathbf{R}^n$. Then*
(1) the matrix equation $A^TA\,t = A^Tx$ has a solution $t \in \mathbf{R}^l$, and for any solution t we have

$$\mathrm{Proj}_U(x) = At,$$

(2) if $\mathrm{rank}(A) = l$ then A^TA is invertible and

$$\mathrm{Proj}_U(x) = A(A^TA)^{-1}A^Tx.$$

Proof: Note that $U^\perp = (\mathrm{Col}A)^\perp = \mathrm{Row}(A^T)^\perp = \mathrm{Null}(A^T)$. Let $u, v \in \mathbf{R}^n$ with $u \in U$, $v \in U^\perp$ and $u + v = x$ so that $\mathrm{Proj}_U(x) = u$. Since $u \in U = \mathrm{Col}A$ we can choose $t \in \mathbf{R}^l$ so that $u = At$. Then we have $x = u + v = At + v$. Multiply by A^T to get $A^T = A^TAt + A^Tv$. Since $v \in U^\perp = \mathrm{Null}(A^T)$ we have $A^Tv = 0$ so $A^TA\,t = A^Tx$. Thus the matrix equation $A^TA\,t = A^Tx$ does have a solution $t \in \mathbf{R}^l$.

Now let $t \in \mathbf{R}^l$ be any solution to $A^TA\,t = A^tx$. Let $u = At$ and $v = x - u$. Note that $x = u + v$, $u = At \in \mathrm{Col}(A) = U$, and $A^Tv = A^T(x - u) = A^T(x - At) = A^Tx - A^TA\,t = 0$ so that $v \in \mathrm{Null}(A^T) = U^\perp$. Thus $\mathrm{Proj}_U(x) = u = At$, proving part (1).

Now suppose that $\mathrm{rank}(A) = l$. Since $A^TA \in M_{l\times l}(\mathbf{R})$ with $\mathrm{rank}(A^TA) = \mathrm{rank}(A) = l$, the matrix A^TA is invertible. Since A^TA is invertible, the unique solution $t \in \mathbf{R}^l$ to the matrix equation $A^TA\,t = A^Tx$ is the vector $t = (A^TA)^{-1}A^Tx$, and so from Part (1) we have $\mathrm{Proj}_U(x) = At = A(A^TA)^{-1}A^Tx$, proving Part (2).

The Volume of a Parallelotope

5.17 Definition: Given vectors $u_1, u_2, \cdots, u_k \in \mathbf{R}^n$, we define the **parallelotope** on $u_1, \cdots, u_k$ to be the set

$$P(u_1, \cdots, u_k) = \Big\{ \sum_{j=1}^{k} t_i u_i \Big| 0 \le t_i \le 1 \text{ for all } i \Big\}.$$

We define the **volume** of this parallelotope, denoted by $V(u_1, \cdots, u_k)$, recursively by $V(u_1) = |u_1|$ and

$$V(u_1, \cdots, u_k) = V(u_1, \cdots, u_{k-1}) \big| \text{Proj}_{U^\perp}(u_k) \big|$$

where $U = \text{Span}\{u_1, \cdots, u_{k-1}\}$.

5.18 Theorem: *Let $u_1, \cdots, u_k \in \mathbf{R}^n$ and let $A = (u_1, \cdots, u_k) \in M_{n\times k}(\mathbf{R})$. Then*

$$V(u_1, \cdots, u_n) = \sqrt{\det(A^T A)}.$$

Proof: We prove the theorem by induction on k. Note that when $k = 1$, $u_1 \in \mathbf{R}^n$ and $A = u_1 \in M_{n\times 1}(\mathbf{R})$, we have $V(u_1) = |u_1| = \sqrt{u_1 \bullet u_1} = \sqrt{{u_1}^T u_1} = \sqrt{A^T A}$, as required. Let $k \ge 2$ and suppose, inductively, that when $A = (u_1, \cdots, u_{k-1}) \in M_{n\times k-1}$ we have $\det(A^T A) > 0$ and $V(u_1, \cdots, u_{k-1}) = \sqrt{\det(A^T A)}$. Let $B = (u_1, \cdots, u_k) = (A, u_k)$. Let $U = \text{Span}\{u_1, \cdots, u_{k-1}\} = \text{Col}(A)$. Let $v = \text{Proj}_U(u_k)$ and $w = \text{Proj}_{U^\perp}(u_k)$. Note that $v \in U = \text{Col}(A)$ and $w \in U^\perp = \text{Null}(A^T)$. Then we have $u_k = v + w$ so that $B = (A, v + w)$. Since $v \in \text{Col}(A)$, the matrix B can be obtained from the matrix (A, w) by performing elementary column operations of the type $C_k \mapsto C_k + tC_i$. Let E be the product of the elementary matrices corresponding to these column operations, and note that $B = (A, v + w) = (A, w)E$. Since the row operations $C_k \mapsto C_k + tC_i$ do not alter the determinant, E is a product of elementary matrices of determinant 1, so we have $\det(E) = 1$. Since $\det(E) = 1$ and $w \in \text{Null}(A^T)$ we have

$$\begin{aligned} \det(B^T B) &= \det\Big(E^T (A, w)^T (A, w) E\Big) = \det\left(\begin{pmatrix} A^T \\ w^T \end{pmatrix} \begin{pmatrix} A & w \end{pmatrix}\right) \\ &= \det\begin{pmatrix} A^T A & A^T w \\ w^T A & w^T w \end{pmatrix} = \begin{pmatrix} A^T A & 0 \\ 0 & |w|^2 \end{pmatrix} = \det(A^T A)\, |w|^2. \end{aligned}$$

By the induction hypothesis, we can take the square root on both sides to get

$$\sqrt{\det(B^T B)} = \sqrt{\det(A^T A)}\, |w| = V(u_1, \cdots, u_{k-1})\ |w| = V(u_1, \cdots, u_k).$$

The Cross Product in $\mathbf{R}^n$

5.19 Definition: Let F be a field. For $n \geq 2$ we define the **cross product**

$$X : \prod_{k=1}^{n-1} F^n \to F^n$$

as follows. Given vectors $u_1, u_2, \cdots, u_{n-1} \in F^n$, we define $X(u_1, u_2, \cdots, u_{n-1}) \in F^n$ to be the vector with entries

$$X(u_1, u_2, \cdots, u_{n-1})_j = (-1)^{n+j} |A^{(j)}|$$

where $A^{(j)} \in M_{n-1}(F)$ is the matrix obtained from $A = (u_1, u_2, \cdots, u_{n-1}) \in M_{n \times n-1}(F)$ by removing the j^{th} row. Given a vector $u \in F^2$ we write $X(u)$ as $u^\times$, and given two vectors $u, v \in F^3$ we write $X(u, v)$ as $u \times v$.

5.20 Example: Given $u \in F^2$ we have

$$u^\times = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^\times = \begin{pmatrix} -u_2 \\ u_1 \end{pmatrix}.$$

Given $u, v \in F^3$ we have

$$u \times v = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} \\ -\begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} \\ \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}.$$

5.21 Note: Because the determinant is n-linear, alternating and skew-symmetric, it follows that the cross product is $(n-1)$-linear, alternating and skew-symmetric. Thus for $u_i, v, w \in F^n$ and $t \in F$ we have

(1) $X(u_1, \cdots, v + w, \cdots, u_{n-1}) = X(u_1, \cdots, v, \cdots, u_{n-1}) + X(u_1, \cdots, w, \cdots, u_{n-1})$,
(2) $X(u_1, \cdots, t\, u_k, \cdots, u_{n-1}) = t\, X(u_1, \cdots, u_k, \cdots, u_{n-1})$,
(3) $X(u_1, \cdots, u_k, \cdots, u_l, \cdots, u_{n-1}) = -X(u_1, \cdots, u_l, \cdots, u_k, \cdots, u_{n-1})$.

5.22 Definition: Recall that for $u_1, \cdots, u_n \in \mathbf{R}^n$, the set $\{u_1, \cdots, u_n\}$ is a basis for $\mathbf{R}^n$ if and only if $\det(u_1, \cdots, u_n) \neq 0$. For an ordered basis $\mathcal{A} = (u_1, \cdots, u_n)$, we say that $\mathcal{A}$ is **positively oriented** when $\det(u_1, \cdots, u_n) > 0$ and we say that $\mathcal{A}$ is **negatively oriented** when $\det(u_1, \cdots, u_n) < 0$.

5.23 Theorem: *Let $u_1, \cdots, u_{n-1}, v_1, \cdots, v_{n-1}, w \in \mathbf{R}^n$. Then*

(1) $X(u_1, \cdots, u_{n-1}) \bullet w = \det(u_1, \cdots, u_{n-1}, w)$,
(2) $X(u_1, \cdots, u_{n-1}) = 0$ if and only if $\{u_1, \cdots, u_{n-1}\}$ is linearly dependent.
(3) When $w = X(u_1, \cdots, u_{n-1}) \neq 0$ we have $\det(u_1, \cdots, u_{n-1}, w) > 0$ so that the n-tuple $(u_1, \cdots, u_{n-1}, w)$ is a positively oriented basis for $\mathbf{R}^n$,
(4) $X(u_1, \cdots, u_{n-1}) \bullet X(v_1, \cdots, v_{n-1}) = \det(A^T B)$ where $A = (u_1, \cdots, u_{n-1}) \in M_{n \times n-1}(\mathbf{R})$ and $B = (v_1, \cdots, v_{n-1}) \in M_{n \times n-1}(\mathbf{R})$, and
(5) $\left|X(u_1, \cdots, u_{n-1})\right|$ is equal to the volume of the parallelipiped on $u_1, \cdots, u_{n-1}$.

Proof: I may include a proof later.

6.1 Definition: Let R be a commutative ring. A **module** over R (or an R-**module**) is a set U together with an element $0 \in U$ and two operations $+ : U \times U \to U$ and $* : R \times U \to U$, where we write $+(x, y)$ as $x + y$ and $*(t, x)$ as $t * x$, $t \cdot x$ or as tx, such that

(1) $+$ is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in U$,
(2) $+$ is commutative: $x + y = y + x$ for all $x, y \in U$,
(3) 0 is an additive identity: $x + 0 = x$ for all $x \in U$,
(4) every element has an additive inverse: for all $x \in U$ there exists $y \in U$ with $x + y = 0$,
(5) $*$ is associative: $(st)x = s(tx)$ for all $s, t \in R$ and all $x \in U$,
(6) 1 is a multiplicative identity: $1 \cdot x = x$ for all $x \in U$,
(7) $*$ is distributive over $+$ in R: $(s + t)x = sx + tx$ for all $s, t \in R$ and all $x \in U$, and
(8) $*$ is distributive over $+$ in U: $t(x + y) = tx + ty$ for all $t \in R$ and all $x, y \in U$.

When F is a field, a module over F is also called a **vector space** over F.

6.2 Note: In an R-module U, the zero element is unique, the additive inverse of a given element $x \in U$ and we denote it by $-x$, and we have additive cancellation.

6.3 Definition: Let R be a commutative ring and let W be an R-module. A **submodule** of W over R is a subset $U \subseteq W$ which is also an R-module using the (restrictions of) the same operations used in W. Note that for a subset $U \subseteq W$, the operations on W restrict to well-defined operations on U if and only if

(1) U is closed under $+$: for all $x, y \in U$ we have $x + y \in U$, and
(2) U is closed under $*$: for all $t \in R$ and all $x \in U$ we have $tx \in U$.

When the operations do restrict as above, U is a submodule of W if and only if

(3) U contains the zero element: $0 \in U$, and
(4) U is closed under negation: for all $x \in U$ we have $-x \in U$.

When F is a field and W is a vector space over F, a submodule of W is also called a **subspace** of W over F.

6.4 Example: Let R be a ring. Then R^n, R^ω and R^∞ are all R-modules, where

$$R^n = \{f : \{1, 2, \cdots, n\} \to R\} = \{(a_1, a_2, \cdots, a_n) \big| \text{each } a_k \in R\},$$
$$R^\omega = \{f : \mathbf{Z}^+ \to R\} = \{(a_1, a_2, a_3, \cdots) \big| \text{ each } a_k \in R\} \text{ and}$$
$$R^\infty = \{f : \mathbf{Z}^+ \to R \big| f(k) = 0 \text{ for all but finitely many } k \in \mathbf{Z}^+\}$$
$$= \{(a_1, a_2, a_3, \cdots) \big| \text{ each } a_k \in R \text{ with } a_k = 0 \text{ for all but finitely many } k \in \mathbf{Z}^+\}$$

6.5 Example: When R is a ring, $M_{m\times n}(R)$ is an R-modulue.

6.6 Example: For two sets A and B, we denote the set of all functions $f : A \to B$ by B^A or by $\text{Func}(A, B)$. When A is a set and R is a ring, we define operations on $\text{Func}(A, R)$ by $(tf)(x) = t\, f(x)$, $(f + g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$ for all $x \in X$. The set $\text{Func}(A, R)$ is a ring under addition and multiplication and also an R-module under addition and multiplication by $t \in R$.

6.7 Example: Let R be a ring. Recall that a **polynomial** ,with coefficients in R, is an expression of the form $f(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ where $n \in \mathbf{N}$ and each

$c_k \in R$. We denote the set of all such polynomials by $R[x]$ or by $P(R)$. We consider two polynomials to be equal when their coefficients are all equal, so for $f(x) = \sum_{k=0}^{n} a_k x^k$ and $g(x) = \sum_{k=0}^{n} b_k x^k$ we have $f = g$ when $a_k = b_k$ for all k. In particular, we have $f = 0$ when $a_k = 0$ for all indices k. When $0 \neq f \in R[x]$, the **degree** of f, denoted by $\deg(f)$, is the largest $n \in \mathbf{N}$ for which $c_n = 0$. The degree of the zero polynomial is $-\infty$. For $n \in \mathbf{N}$, we denote the set of all polynomials of degree at most n by $P_n(R)$. A (formal) **power series**, with coefficients in R, is an expression of the form $f(x) = \sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 + c_2 x^2 + \cdots$. We denote the set of all such power series by $R[[x]]$. Thus we have

$$P_n(R) = \Big\{ f(x) = \sum_{k=0}^{n} c_k x^k \Big| \text{ each } c_k \in R \Big\},$$

$$R[x] = P(R) = \bigcup_{n \in \mathbf{N}} P_n(R) = \Big\{ f(x) = \sum_{k=0}^{n} c_k x^k \Big| \, n \in \mathbf{N} \text{ and each } c_k \in R \Big\} \text{ and}$$

$$R[[x]] = \Big\{ \sum_{k=0}^{\infty} c_k x^k \Big| \text{ each } c_k \in R \Big\}.$$

We define operations on these sets as follows. Given $f(x) = \sum_{ikge0} a_k x^k$ and $g(x) = \sum_{k \geq 0} b_k x^k$ (where the sums may be finite or infinite) we define tf, $f + g$ and fg by

$$(tf)(x) = \sum_{k \geq o} t\, a_k x^k \ , \ (f+g)(x) = \sum_{ikge0} (a_k + b_k) x^k \ , \text{ and}$$

$$(fg)(x) = \sum_{i,j \geq 0} (a_i b_j) x^{i+j} = \sum_{k \geq 0} c_k x^k \ \text{ with } c_k = \sum_{i=0}^{k} a_i b_{k-i}.$$

The sets $P_n(R)$, $R[x] = P(R)$ and $R[[x]]$ are all rings under addition and multiplication, and they are all R-modules under addition and multiplication by elements $t \in R$.

6.8 Definition: Let R be a commutative ring. An **algebra** over R (or an **R-algebra**) is a set U with an element $0 \in U$ together with three operations $+ : U \times U \to U$, $* : U \times U \to U$ and $* : R \times U \to U$ such that U is an abelian group under $+$, such that the two multiplication operations satisfy $(xy)z = x(yz)$, $(x+y)z = xz + yz$, $x(y+z) = xy + xz$ and $t(xy) = (tx)y$ for all $t \in R$ and all $x, y, z \in U$.

6.9 Example: When R is a commutative ring and A is a set, R^n, R^∞, R^ω, $M_n(R)$, $\mathrm{Func}(A, R)$, $P_n(R)$, $R[x]$, $R[[x]]$ and R^A are all R-algebras using their usual operations.

6.10 Theorem: *Let R be a ring, and let W be an R-module. Let A be a set, and for each $\alpha \in A$ let U_α be a submodule of W over R. Then $\bigcap_{\alpha \in A} U_\alpha$ is an R-module.*

Proof: I may include a proof later.

6.11 Definition: Let R be a ring, let W be an R-module, and let $S \subseteq U$. A **linear combination** of the set S (or of the elements in S) (over R) is an element $w \in W$ of the form $w = \sum_{i=1}^{n} t_i u_i$ for some $n \in \mathbf{N}$, $t_i \in R$ and $u_i \in U$ (we allow the case $n = 0$ to include the empty sum, which we take to be equal to 0). The **span** of S (over R), denoted by $\mathrm{Span}\,(S)$ or by $\mathrm{Span}_R(S)$, is the set of all such linear combinations, that is

$$\mathrm{Span}\,(S) = \Big\{ \sum_{i=1}^{n} t_i u_i \Big| n \in \mathbf{N}, t_i \in R, u_i \in S \Big\}.$$

When $U = \mathrm{Span}\,(S)$, we also say that S **spans** U or that U is **spanned by** S.

The submodule of W **generated by** S, denoted by $\langle S \rangle$, is the smallest submodule of W containing S, that is

$$\langle S \rangle = \bigcap \big\{ U \subseteq W \big| U \text{ is a submodule of } W \text{ with } S \subseteq U \big\}.$$

6.12 Theorem: *Let R be a ring, let W be an R-module, and let $S \subseteq W$. Then*

$$\langle S \rangle = \mathrm{Span}\,(S).$$

Proof: I may include a proof later.

6.13 Definition: Let R be a ring, let U be an R-module, and let $S \subseteq U$. We say that S is **linearly independent** (over R) when for all $n \in \mathbf{Z}^+$, for all $t_i \in R$, and for all distinct elements $u_i \in S$, if $\sum_{i=1}^{n} t_i u_i = 0$ then $t_i = 0$ for all indices i. Otherwise we say that S is **linearly dependent**. We say that S is a **basis** for U when S spans U and S is linearly independent.

6.14 Note: A set S is a basis for an R-module U if and only if every element $x \in U$ can be written uniquely (up to the order of the terms in the sum) in the form $x = \sum_{i=1}^{n} t_i u_i$ where $n \in \mathbf{N}$, $0 \neq t_i \in F$ and $u_1, u_2, \cdots, u_n$ are distinct elements in S.

6.15 Example: For $e_k \in R^n$ given by $(e_k)_i = \delta_{k,i}$, the set $\mathcal{S} = \{e_1, e_2, \cdots, e_n\}$ is a basis for R^n, and we call it the **standard basis** for R^n. For $e_k \in R^\infty$ given by $(e_k)_i = \delta_{k,i}$, the set $\mathcal{S} = \{e_1, e_2, e_3, \cdots\}$ is a basis for R^∞, which we call the **standard basis** for R^∞. It is not immediately obvious whether the R-module R^ω has a basis.

6.16 Example: For each $k, l \in \{1, 2, \cdots, n\}$, let $E_{k,l} \in M_n(R)$ denote the matrix with $(E_{k,l})_{i,j} = \delta_{k,i}\delta_{l,j}$ (so the (k, l) entry is equal to 1 and all other entries are equal to 0). Then the set $\mathcal{S} = \big\{ E_{k,l} \big| k, l \in \{1, 2, \cdots, n\} \big\}$ is a basis for $M_{m \times n}(R)$, which we call the **standard basis** for $M_{m \times n}(R)$.

6.17 Example: For a ring R, the set $\mathcal{S} = \mathcal{S}_n = \big\{1, x, x^2, \cdots, x^n\big\}$ is the **standard basis** for $P_n(R)$, and the set $\mathcal{S} = \big\{1, x, x^2, x^3, \cdots\big\}$ is the **standard basis** for $P_n(R) = R[x]$. It is not immediately obvious whether the R-module $R[[x]]$ has a basis.

Ordered Bases and the Coordinate Map

6.18 Definition: Let R be a ring and let W be an R-module. Let $\mathcal{A} = (u_1, u_2, \cdots)$ be an ordered n-tuple of elements in W. A **linear combination** of $\mathcal{A}$ (over R) is an element $x \in W$ of the form $x = \sum_{i=1}^{n} t_i u_i$ with each $t_i \in R$. The **span** of U (over R) is the set

$$\text{Span}\,(\mathcal{A}) = \Big\{ \sum_{i=1}^{n} t_i u_i \Big| t \in R^n \Big\}.$$

When $U = \text{Span}\,(\mathcal{A})$ we say that $\mathcal{A}$ **spans** U or that U is **spanned by** $\mathcal{A}$. We say that $\mathcal{A}$ is **linearly independent** (over R) when for all $t \in R^n$, if $\sum_{i=1}^{n} t_i u_i = 0$ then $t = 0$. We say that $\mathcal{A}$ is an **ordered basis** for U when $\mathcal{A}$ is linearly independent and spans U.

6.19 Note: Let R be a ring, let W be an R-module, and let $u_1, u_2, \cdots, u_n \in W$. Then the span of the n-tuple $(u_1, u_2, \cdots, u_n)$ is equal to the span of there set $\{u_1, u_2, \cdots, u_n\}$, and the n-tuple $(u_1, u_2, \cdots, u_n)$ is linearly independent if and only of the elements $u_1, u_2, \cdots, u_n$ are distinct and the set $\{u_1, u_2, \cdots, u_n\}$ is linearly independent. For a submodule $U \subseteq W$, the n-tuple $(u_1, \cdots, u_n)$ is a basis for U if and only if the elements u_i are distinct and the set $\{u_1, \cdots, u_n\}$ is a basis for U.

6.20 Definition: Let R be a ring, let U be an R-module, and let $\mathcal{A} = (u_1, u_2, \cdots, u_n)$ be an ordered basis for U. Given an element $x \in U$, since $\mathcal{A}$ spans U we can write x as a linear combination $x = \sum_{i=1}^{n} t_i u_i$ with $t \in R^n$, and since $\mathcal{A}$ is linearly independent the element $t \in R^n$ is unique. We denote the unique element $t \in R^n$ such that $x = \sum_{i=1}^{n} t_i u_i$ by $[x]_{\mathcal{A}}$. Thus for $x \in U$ and $t \in R^n$ we have

$$t = [x]_{\mathcal{A}} \iff x = \sum_{i=1}^{n} t_i u_i \,.$$

The elements $t_1, t_2, \cdots, t_n \in R$ are called the **coordinates** of x with respect to the ordered basis $\mathcal{A}$. In the case that R is a field, the vector $t = [x]_{\mathcal{A}}$ is called the **coordinate vector** of x with respect to $\mathcal{A}$. The bijective map $\phi_{\mathcal{A}} : U \to R^n$ given by

$$\phi_{\mathcal{A}}(x) = [x]_{\mathcal{A}}$$

is called the **coordinate map** from U to R^n induced by the ordered basis $\mathcal{A}$.

6.21 Theorem: *Let R be a ring, let U be a free R-module, and let $\mathcal{A}$ be a finite ordered basis for U. Then coordinate map $\phi_{\mathcal{A}} : U \to R^n$ is bijective and linear. That is*

$$\phi_{\mathcal{A}}(x + y) = \phi_{\mathcal{A}}(x) + \phi_{\mathcal{A}}(y) \quad \textit{and} \quad \phi_{\mathcal{A}}(tx) = r\,\phi_{\mathcal{A}}(x)$$

for all $x, y \in U$ and all $r \in R$.

Proof: I may include a proof later.

The Dimension of Finite Dimensional Vector Spaces

6.22 Note: Let R be a ring and let U be a free R-module. If $\mathcal{A}$ and $\mathcal{B}$ are bases for U with $\mathcal{A} \subseteq \mathcal{B}$, then we must have $\mathcal{A} = \mathcal{B}$ because if we had $\mathcal{A} \subsetneqq \mathcal{B}$ then we could choose $v \in \mathcal{B} \setminus \mathcal{A}$ then write v as a linear combination $v = \sum_{i=1}^{n} t_i u_i$ with each $u_i \in \mathcal{A}$, but then we would would have $0 = 1 \cdot v - \sum_{i=1}^{n} t_i u_i$ which is a linear combination of elements in $\mathcal{B}$ with coefficients not all equal to zero.

6.23 Theorem: *Let R be a ring and let U be a free R-module. Let $\mathcal{A}$ and $\mathcal{B}$ be two bases for U over R. Then $\mathcal{A}$ and $\mathcal{B}$ are either both finite or both infinite.*

Proof: If $U = \{0\}$ then $\mathcal{A} = \mathcal{B} = \emptyset$. Suppose that $U \neq \{0\}$. Suppose that one of the two bases is finite, say $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$. For each index i, write $u_i = \sum_{j=1}^{m_i} t_{i,j} v_{i,j}$ with $t_{i,j} \in R$ and $v_{i,j} \in \mathcal{B}$. Let $\mathcal{C} = \{v_{i,j} \big| 1 \leq i \leq n\,,\, 1 \leq j \leq m_i\}$. Note that $\mathcal{C}$ is finite, and $\mathcal{C}$ is linearly independent because $\mathcal{C} \subseteq \mathcal{B}$ and $\mathcal{B}$ is linearly independent, and $\mathcal{C}$ spans U because given $x \in U$ we can write $x = \sum_{i=1}^{n} t_i u_i$ and then we have $x = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (t_i s_{i,j}) v_{i,j} \in \text{Span}\,(\mathcal{C})$. Since $\mathcal{B}$ and $\mathcal{C}$ are both bases for U and $\mathcal{C} \subseteq \mathcal{B}$ we have $\mathcal{C} = \mathcal{B}$, so $\mathcal{B}$ is finite.

6.24 Theorem: *Let F be a field and let U be a vector space over F. Let $\mathcal{A}$ and $\mathcal{B}$ be finite bases for U over F. Then $|\mathcal{A}| = |\mathcal{B}|$.*

Proof: If $U = \{0\}$ then $\mathcal{A} = \mathcal{B} = \emptyset$. Suppose that $U \neq \{0\}$. Let $n = |\mathcal{A}|$ and say $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$. Replace the set $\mathcal{A}$ by the ordered n-tuple $\mathcal{A} = (u_1, u_2, \cdots, u_n)$. Let $\phi = \phi_{\mathcal{A}} : U \to F^n$ and consider the set $\phi(\mathcal{B}) = \{\phi(v) \big| v \in \mathcal{B}\} \subseteq F^n$. Note that $\phi(\mathcal{B})$ spans F^n because given $t \in F^n$ we can let $x = \sum_{i=1}^{n} t_i u_i$ so that $[x]_{\mathcal{A}} = t$, then we can write $x = \sum_{i=1}^{m} s_i v_i$ with $s_i \in R$ and $v_i \in \mathcal{B}$, and then we have $t = \phi(x) = \phi\Big(\sum_{i=1}^{m} s_i v_i\Big) = \sum_{i=1}^{m} s_i \phi(v_i) \in \text{Span}\,\big(\phi(\mathcal{B})\big)$. Also, we claim that $\phi(\mathcal{B})$ is linearly independent in F^n. Suppose that $\sum_{i=1}^{m} s_i y_i = 0$ where the y_i are distinct elements in F^n and each $s_i \in F$. Choose elements $x_i \in \mathcal{B}$ so that $\phi(x_i) = y_i$, and note that the elements x_i will be distinct because ϕ is bijective. Then we have $0 = \sum_{i=1}^{m} s_i y_i = \sum_{i=1}^{m} s_i \phi(x_i) = \phi\Big(\sum_{i=1}^{m} s_i x_i\Big)$. Since ϕ is injective it follows that $\sum_{i=1}^{m} s_i x_i = 0$. Since the elements x_i are distinct elements in $\mathcal{B}$ and $\mathcal{B}$ is linearly independent, it follows that every $s_i = 0$. Thus $\phi(\mathcal{B})$ is linearly independent, as claimed. Since $\phi(\mathcal{B})$ spans F^n it follows that $\big|\phi(\mathcal{B})\big| \geq n$, and since $\phi(\mathcal{B})$ is linearly independent it follows that $\big|\phi(\mathcal{B})\big) \leq n$, and so we have $\big|\phi(\mathcal{B})\big| = n = |\mathcal{A}|$. Since ϕ is bijective we have $|\mathcal{B}| = \big|\phi(\mathcal{B})\big| = |\mathcal{A}|$.

6.25 Definition: Let U be a vector space over a field F. When U has a finite basis, we say that U is **finite dimensional** (over F) and we define the **dimension** of U to be $\dim(U) = |\mathcal{A}|$ where $\mathcal{A}$ is any basis for U.

The Existence of a Basis for a Vector Space

6.26 Definition: Let S be a nonempty set of sets. A **chain** in S is a nonempty subset $T \subseteq S$ with the property that for all $A, B \in T$, either $A \subseteq B$ or $B \subseteq A$. For a subset $T \subseteq S$, an **upper bound** for T in S is an element $B \in S$ such that $A \subseteq B$ for all $A \in T$. A **maximal** element in S is an element $B \in S$ such that there is no $A \in S$ with $B \subseteq A$.

6.27 Theorem: *(Zorn's Lemma) Let S be a nonempty set. Suppose that every chain in S has an upper bound in S. Then S has a maximal element.*

Proof: We take Zorn's Lemma to be an axiom, which means that we accept it as true without proof.

6.28 Note: Let F be a field and let U be a vector space over F. Let $\mathcal{A}$ be a linearly independent subset of U and let $v \in U$. Then $\mathcal{A} \cup \{v\}$ is linearly independent if and only if $v \notin \text{Span}(\mathcal{A})$ We leave the proof of this result as an exercise. Note that the analogous result does not hold when U is a module over a ring R.

6.29 Theorem: *Every vector space has a basis. Indeed, every linearly independent set in a vector space is contained in a basis for the vector space.*

Proof: Let F be a field and let U be a vector space over F. Let $\mathcal{A}$ be a subset of U which is linearly independent over F. Let S be the collection of all linearly independent sets $\mathcal{B} \subseteq U$ with $\mathcal{A} \subseteq \mathcal{B}$. Note that $S \neq \emptyset$ because $\mathcal{A} \in S$. Let T be a chain in S. We claim that $\bigcup T \in S$. Since $T \neq \emptyset$ we can choose an element $\mathcal{B}_\prime \in T$ and then we have $\mathcal{A} \subseteq \mathcal{B}_\prime \subseteq \bigcup T$. Since for every $\mathcal{B} \in T$ we have $\mathcal{B} \subseteq U$ it follows that $\bigcup T \subseteq U$. It remains to show that $\bigcup T$ is linearly independent. Suppose that $\sum_{i=1}^{n} t_i u_i = 0$ where the u_i are distinct elements in $\bigcup T$ and $t_i \in F$. For each index i, since $u_i \in \bigcup T$ we can choose $\mathcal{B}_i \in T$ with $u_i \in \mathcal{B}_i$. Since T is a chain, for all indices i and j, either $\mathcal{B}_i \subseteq \mathcal{B}_j$ or $\mathcal{B}_| \subseteq \mathcal{B}_i$. It follows that we can choose an index k so that $\mathcal{B}_i \subseteq B_k$ for all indices i. Then we have $u_i \in \mathcal{B}_k$ for all i. Since the u_i are distinct elements in $\mathcal{B}_k$ with $\sum_{i=1}^{n} t_i u_i = 0$ and since $\mathcal{B}_\|$ is linearly independent it follows that $t_i = 0$ for every i. This shows that $\bigcup T$ is linearly independent, and so $\bigcup T \in S$, as claimed. Since $\bigcup T \in S$ it follows that T has an upper bound in S since for every $\mathcal{B} \in T$ we have $\mathcal{B} \subseteq \bigcup T$. By Zorn's Lemma, it follows that S has a maximal element. Let $\mathcal{B}$ be a maximal element in S. We claim that $\mathcal{B}$ is a basis for U. Since $\mathcal{B} \in S$ we know that $\mathcal{A} \subseteq \mathcal{B} \subseteq U$ and that $\mathcal{B}$ is linearly independent. Note also that $\mathcal{B}$ spans U because if we had $\text{Span}(\mathcal{B}) \subsetneqq U$ then we could choose $w \in U$ with $w \notin \text{Span}(\mathcal{B})$ and then $\mathcal{B} \cup \{w\}$ would be linearly independent by the above Note, but then $\mathcal{B}$ would not be maximal in S.

6.30 Example: When F is a field and A is any set, the vector spaces F^ω, $F[[x]]$ and $\text{Func}(A, F)$ all have bases. It is not easy to construct an explicit basis for any of these vector spaces.

6.31 Example: There exists a basis for $\mathbf{R}$ as a vector space over $\mathbf{Q}$, but it is not easy to construct an explicit basis.

Some Cardinal Arithmetic

6.32 Definition: Let S be a set of nonempty sets. A **choice function** on S is a function $f : S \to \bigcup S$ such that $f(A) \in A$ for every $A \in S$.

6.33 Theorem: *(The Axiom of Choice) Every set of nonempty sets has a choice function.*

Proof: A proof can be found in Ehsaan Hossain's Tutorial Lecture Notes.

6.34 Corollary: *Let A be a set. For each $\alpha \in A$, let X_α be a nonempty set. Then there exists a function $f : A \to \bigcup\limits_{\alpha \in A} X_\alpha$ with $f(\alpha) \in X_\alpha$ for all $\alpha \in A$.*

Proof: Let $S = \{X_\alpha | \alpha \in A\}$. Note that $\bigcup S = \bigcup\limits_{\alpha \in A} X_\alpha$. Let $g : S \to \bigcup S$ be a choice function for S, so we have $g(X_\alpha) \in X_\alpha$ for all $\alpha \in A$. Define the map $f : A \to \bigcup\limits_{\alpha \in A} X_\alpha$ by $f(\alpha) = g(X_\alpha)$ to obtain $f(\alpha) \in X_\alpha$ for all $\alpha \in A$, as required..

6.35 Theorem: *Let A and B be nonempty sets and let $f : A \to B$. Then*

(1) f is injective if and only if f has a left inverse, and
(2) f is surjective if and only if f has a right inverse.

Proof: The proofs can be found in last term's MATH 147 Lecture Notes. We remark that the proof of Part (1) does not require the Axiom of Choice but the proof of Part (2) does.

6.36 Definition: *For sets A and B, we say that A and B are* **equipotent** *(or that A and B have the same cardinality), and we write $|A| = |B|$ when there exists a bijection $f : A \to B$. We say that the cardinality of A is* **less than or equal to** *the cardinality of B, and we write $|A| \leq |B|$, when there exists an injective map $f : A \to B$. Note that by the above theorem, we have $|A| \leq |B|$ if and only if there exists a surjective map $fgB \to A$.*

6.37 Note: *It follows immediately from the above definitions that for all sets A, B and C we have*

(1) $|A| = |A|$,
(2) if $|A| = |B|$ then $|B| = |A|$, and
(3) if $|A| = |B|$ and $|B| = |C|$ then $|A| = |C|$, and also
(4) $|A| \leq |A|$, and
(5) if $|A| \leq |B|$ and $|B| \leq |C|$ then $|A| \leq |C|$.

Properties 1, 2 and 3 imply that equipotence is an equivalence relation on the class of all sets. Properties 4 and 5 are two of the 4 properties which appear in the definition a total order. The other 2 properties also hold, but they require proof. The third property is known as the Cantor-Schroeder-Bernstein Theorem, and we state it below. After that, we state and prove the fourth property.

6.38 Theorem: *(The Cantor-Schroeder-Bernstein Theorem) Let A and B be sets. If $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$.*

Proof: A proof can be found in last term's MATH 147 Lecture Notes (we remark that it does not require the Axiom of Choice).

6.39 Theorem: *Let A and B be sets. Then either $|A| \leq |B|$ or $|B| \leq |A|$.*

Proof: If $A = \emptyset$ we have $|A| \leq |B|$. If $B = \emptyset$ we have $|B| \leq |A|$. Suppose that $|A| \neq \emptyset$ and $B \neq \emptyset$. Let S be the set of all (graphs of) injective functions $f : X \to B$ with $X \subseteq A$. Note that $S \neq \emptyset$ since we can choose $a \in A$ and $b \in B$ and define $f : \{a\} \to B$ by $f(a) = b$. Let T be a chain in S. Note that $\bigcup T \in S$, as in the proof of the Axiom of Choice found in Ehsaan Hossain's notes, and so T has an upper bound in S. By Zorn's Lemma, it follows that S has a maximal element. Let (the graph of) $g : X \to B$ be a maximal element in S. Note that either $X = A$ or $g(X) = B$ since if we had $X \subsetneqq A$ and $g(X) \subsetneqq B$ then we could choose an element $a \in A \setminus X$ and an element $b \in B \setminus f(A)$ and then extend g to the injective map $h : X \cup \{a\} \to B$ defined by $h(x) = g(x)$ for $x \in X$ and $h(a) = b$, contradicting the maximality of g. In the case that $X = A$, since the map $g : A \to B$ is injective we have $|A| \leq |B$. In the case that $f(X) = B$, the map $g : X \to B$ is surjective so we have $|B| \leq |X| \leq |A|$.

6.40 Note: Recall that for sets X and Y, the set of all functions $f : Y \to X$ is denoted by X^Y. As an exercise, verify that given sets A_1, A_2, B_1 and B_2 with $|A_1| = |A_2|$ and $|B_1| = |B_2|$, we have

(1) $|A_1 \times B_1| = |A_2 \times B_2|$,
(2) $\left|{A_1}^{B_1}\right| = \left|{A_2}^{B_2}\right|$, and
(3) if $A_1 \cap B_1 = \emptyset$ and $A_2 \cap B_2 = \emptyset$ then $|A_1 \cup B_1| = |A_2 \cup B_2|$.

6.41 Definition: (Cardinal Arithmetic) For sets A, B and X, we write $|X| = |A|\,|B|$ when $|X| = |A \times B|$, $|X| = |A|^{|B|}$ when $|X| = |A^B|$, and $|X| = |A| + |B|$ when $|X| = |A' \cup B'|$ for disjoint sets A' and B' with $|A'| = |A|$ and $|B'| = |B|$ (for example the sets $A' = A \times \{1\}$ and $B' = B \times \{2\}$).

6.42 Note: Let B be an infinite set and let $n \in \mathbf{Z}^+$, and let $S_n = \{1, 2, \cdots, n\}$. As an exercise, show that there are injective maps $f : B \to B \times S_n$ and $g : B \times S_n \to B \times B$ so that $|B| \leq |B \times S_n| \leq |B \times B|$, then use the Cantor-Schroeder-Bernstein Theorem to show that if $|B| = |B \times B|$ then we have $|B \times S_n| = |B|$.

6.43 Theorem: *Let A be an infinite set. Then $|A \times A| = |A|$.*

Proof: Let S be the set of all (graphs of) bijective functions $f : X \times X \to X$ where X is an infinite subset of A. Note that $S \neq \emptyset$ because, as proven in last term's MATH 147 Lecture notes, we can choose a countable subset $X \in A$ and a bijection $f : X \times X \to X$. Let T be a chain in S. Note that $\bigcup T \in S$ as in the proof of the Axiom of Choice found in Ehsaan Hossain's notes, and so T has an upper bound in S. By Zorn's Lemma, S has a maximal element. Let (the graph of) $g : B \times B \to B$ be a maximal element in S. Note that since $g : B \times B \to B$ is bijective we have $|B \times B| = |B|$. By the previous theorem, either $|B| \leq |A \setminus B|$ or $|A \setminus B| \leq |B|$. We claim that $|A \setminus B| \leq |B|$.

Suppose, for a contradiction, that $|B| \leq |A \setminus B|$. Choose $C \subseteq A \setminus B$ with $|C| = |B|$. Note that the set $(B \cup C) \times (B \cup C)$ is the disjoint union

$$(B \cup C) \times (B \cup C) = (B \times B) \cup (B \times C) \cup (C \times B) \cup (C \times C)$$

and we have

$$\begin{aligned} \big|(B \times C) \cup (C \times B) \cup (C \times C)\big| &= |B \times C| + |C \times B| + |C \times C| \\ &= |B \times B| + |B \times B| + |B \times B| = |B| + |B| + |B| = |B \times \{1, 2, 3\}| = |B| = |C| \end{aligned}$$

and so the maximal bijective map $g : B \times B \to B$ can be extended to a bijective map $h : (B \cup C) \times (B \cup C) \to (B \cup C)$ contradicting the maximality of g. Thus the case in which $|B| \leq |A \setminus B|$ cannot arise, and so we must have $|A \setminus B| \leq |B|$, as claimed.

Since $|A \setminus B| \leq |B|$ we have

$$|A| = \big|(A \setminus B) \cup B\big| = |A \setminus B| + |B| \leq |B| + |B| = |B \times \{1, 2\}| = |B|$$

and hence

$$|A \times A| = |B \times B| = |B| = |A|.$$

6.44 Corollary: *Let A and B be sets.*

(1) If A is nonempty and B is infinite and $|A| \leq |B|$, then $|A|\,|B| = |B|$.
(2) If B is infinite and $|A| \leq |B|$ then $|A| + |B| = |B|$.

Proof: The proof is left as an exercise.

6.45 Corollary: *Let A be an infinite set. For each $u \in A$ let B_u be a finite set. Then*

$$\Big|\bigcup_{u \in A} B_u\Big| \leq |A|.$$

Proof: For each $u \in A$, choose a surjective map $f_u : A \to B_u$. Define $f : A \times A \to \bigcup\limits_{u \in A} B_u$ by $g(u, v) = f_u(v)$. Note that g is surjective and so we have

$$\Big|\bigcup_{u \in A} B_u\Big| \geq |A \times A| = |A|.$$

The Dimension of an Infinite Dimensional Vector Space

6.46 Theorem: *Let R be a ring and let U be a free R-module. Then any two infinite bases for U have the same cardinality.*

Proof: Let $\mathcal{A}$ and $\mathcal{B}$ be infinite bases for U. For each $u \in \mathcal{A}$, let $c = c(u) : \mathcal{A} \to R$ be the (unique) function, with $c(u)_v = 0$ for all but finitely many elements $v \in \mathcal{B}$, such that $u = \sum_{v \in \mathcal{B}} c(u)_v \cdot v$, and then let B_u be the set of all elements $v \in \mathcal{B}$ for which $c(u)_v \neq 0$. Note that each set B_u is a nonempty finite subset of $\mathcal{B}$. Let $\mathcal{C} = \bigcup_{u \in \mathcal{A}} B_u$. Note that $\mathcal{C}$ spans U because given any $x \in U$ we can write $x = \sum_{i=1}^{n} t_i u_i$ with each $u_i \in \mathcal{A}$, and then for each index i we can write $u_i = \sum_{j=1}^{m_i} s_{i,j} v_{i,j}$ with each $v_{i,j} \in B_{u_i}$, and then we have $x = \sum_{i,j} (t_i s_{i,j}) v_{i,j} \in \text{Span}\big(\bigcup_{i=1}^{n} B_{u_i}\big) \subseteq \text{Span}\,(\mathcal{C})$. Since $\mathcal{B}$ is linearly independent and $\mathcal{C} \subseteq \mathcal{B}$, it follows that $\mathcal{C}$ is linearly independent. Since $\mathcal{C}$ is linearly independent and spans U, it is a basis for U. Since $\mathcal{C}$ and $\mathcal{B}$ are bases for U with $\mathcal{C} \subseteq \mathcal{B}$ it follows that $\mathcal{C} = \mathcal{B}$ because if we had $\mathcal{C} \subsetneqq \mathcal{B}$ then we could choose $v \in \mathcal{B} \setminus \mathcal{C}$ then write v as a linear combination $v = \sum_{i=1}^{n} t_i v_i$ with each $v_i \in \mathcal{C}$, but then we would have $0 = 1 \cdot v - \sum_{i=1}^{n} t_i v_i$ which is a linear combination of elements in $\mathcal{B}$ with not all coefficients equal to 0. By the above theorem, we have

$$|\mathcal{B}| = |\mathcal{C}| = \Big| \bigcup_{u \in \mathcal{A}} B_u \Big| \leq |\mathcal{A}|.$$

By interchanging the rôles of $\mathcal{A}$ and $\mathcal{B}$ in the above proof, we see that $|\mathcal{A}| \leq |\mathcal{B}|$. Thus we have $|\mathcal{A}| = |\mathcal{B}|$ by the Cantor-Schroeder-Bernstein Theorem.

6.47 Definition: Let R be a ring and let U be a free R-module with an infinite basis. We define the **rank** of R to be $\text{rank}(R) = |\mathcal{A}|$ where $\mathcal{A}$ is any basis for U. When F is a field and U is a vector space over F which has an infinite basis, the rank of U is also called the **dimension** of U.

Chapter 7. Module Homomorphisms and Linear Maps

7.1 Definition: Let R be a ring and let U and V be R-modules. An (R-module) **homomorphism** from U to V is a map $L : U \to V$ such that

$$L(x+y) = L(x) + L(y) \quad \text{and} \quad L(tx) = t\,L(x)$$

for all $x, y \in U$ and all $t \in R$. A bijective homomorphism from U to V is called an **isomorphism** from U to V, a homomorphism from U to U is called an **endomorphism** of U, and an isomorphism from U to U is called an **automorphism** of U. We say that U is **isomorphic** to V, and we write $U \cong V$, when there exists an isomorphism $L : U \to V$. We use the following notation

$$\begin{aligned}
\mathrm{Hom}(U,V) = \mathrm{Hom}_R(U,V) &= \big\{L : U \to V \big| L \text{ is a homomorphism}\big\},\\
\mathrm{Iso}(U,V) = \mathrm{Iso}_R(U,V) &= \big\{L : U \to V \big| L \text{ is an isomorphism}\big\},\\
\mathrm{End}(U) = \mathrm{End}_R(U) &= \big\{L : U \to U \big| L \text{ is an endomorphism}\big\},\\
\mathrm{Aut}(U) = \mathrm{Aut}_R(U) &= \big\{L : U \to U \big| L \text{ is an automorphism}\big\}.
\end{aligned}$$

For $L, M \in \mathrm{Hom}(U,V)$ and $t \in R$ we define $L + M$ and tM by

$$(L+M)(x) = L(x) + M(x) \quad \text{and} \quad (tL)(x) = t\,L(x).$$

Using these operations, if R is commutative then the set $\mathrm{Hom}(U,V)$ is an R-module. For $L \in \mathrm{Hom}(U,V)$, the **image** (or **range** of L and the **kernel** (or **null set**) of L are the sets

$$\begin{aligned}
\mathrm{Image}(L) = \mathrm{Range}(L) = L(U) &= \big\{L(x) \big| x \in U\big\} \quad \text{and}\\
\mathrm{Ker}(L) = \mathrm{Null}(L) = L^{-1}(0) &= \big\{x \in U \big| L(x) = 0\big\}.
\end{aligned}$$

When F is a field and U and V are vector spaces over F, an F-module homomorphism from U to V is also called a **linear map** from U to V.

7.2 Note: For an R-module homomorphism $L : U \to V$ and for $x \in U$ we have $L(0) = 0$ and $L(-x) = -L(x)$, and for $t_i \in R$ and $x_i \in U$ we have $L\big(\sum_{i=1}^{n} t_i x_i\big) = \sum_{i=1}^{n} t_i\,L(x_i)$.

7.3 Definition: When G and H are groups, a map $L : G \to H$ is called a **group homomorphism** when $L(xy) = L(x)L(y)$ for al $x, y \in G$. A **group isomorphism** is a bijective group homomorphism. When R and S are rings, a map $L : R \to S$ is called a **ring homomorphism** when $L(x+y) = L(x) + L(y)$ and $L(xy) = L(x)L(y)$ for all $x, y \in R$. A **ring isomorphism** is a bijective ring homomorphism. When R is a ring and U and V are R-algebras, a map $L : U \to V$ is called an **R-algebra homomorphism** when $L(x+y) = L(x) + L(y)$, $L(xy) = L(x)L(y)$ and $L(tx) = t\,L(x)$ for all $x, y \in U$ and all $t \in R$. An **R-algebra isomorphism** is a bijective R-algebra homomorphism.

7.4 Theorem: *Let R be a ring and let U, V and W be R-modules.*

(1) If $L : U \to V$ and $M : V \to W$ are homomorphisms then so is the composite $ML : U \to W$.
(2) If $L : U \to V$ is an isomorphism, then so is the inverse $L^{-1} : V \to U$.

Proof: Suppose that $L : U \to V$ and $M : V \to W$ are R-module homomorphisms. Then for all $x, y \in U$ and all $t \in R$ we have

$$M(L(x+y)) = M(L(x) + L(y)) = M(L(x)) + M(L(y)) \text{ and}$$
$$M(L(tx)) = M(t\,L(x)) = t\,M(L(x)).$$

Suppose that $L : U \to V$ is an isomorphism. Then given $u, v \in V$ and $t \in R$, if we let $x = L^{-1}(u)$ and $y = L^{-1}(v)$ then we have

$$L^{-1}(u+v) = L^{-1}\big(L(x) + L(y)\big) = L^{-1}\big(L(x+y)\big) = x + y = L^{-1}(u) + L^{-1}(v) \text{ and}$$
$$L^{-1}(tu) = L^{-1}\big(t\,L(x)\big) = L^{-1}\big(L(tx)\big) = tx = t\,L^{-1}(u).$$

7.5 Corollary: *Let R be a ring. Then isomorphism is an equivalence relation on the class of all R-modules. This means that for all R-modules U, V and W we have*

(1) $U \cong U$,
(2) if $U \cong V$ then $V \cong U$, and
(3) if $U \cong V$ and $V \cong W$ then $U \cong W$.

7.6 Corollary: *When R is a commutative ring and U is an R-module,* $\mathrm{End}(U)$ *is a ring under addition and composition, hence also an R-algebra, and* $\mathrm{Aut}(U)$ *is a group under composition.*

7.7 Theorem: *Let $L : U \to V$ be an R-algebra homomorphism.*

(1) If U_0 is a submodule of U then $L(U)$ is a submodule of V. In particular, the image of L is a submodule of V.
(2) If V_0 is a submodule of V then $L^{-1}(V_0)$ is a submodule of U. In particular, the kernel of L is a submodule of U.

Proof: To prove Part (1), let U_0 be a submodule of U. Let $u, v \in L(U_0)$ and let $t \in R$. Choose $x, y \in U_0$ with $L(x) = u$ and $L(y) = v$. Since $x + y \in U_0$ and $L(x+y) = L(x) + L(y) = u + v$, it follows that $u + v \in L(U_0)$. Since $tx \in U_0$ and $L(tx) = t\,L(x) = tu$, it follows that $tu \in L(U_0)$. Thus $L(U_0)$ is closed under the module operations and so it is a submodule of V.

To prove Part (2), let V_0 be a submodule of V. Let $x, y \in L^{-1}(V_0)$ and let $t \in R$. Let $u = L(x) \in V_0$ and $v = L(y) \in V_0$. Since $L(x+y) = L(x) + L(y) = u + v \in V_0$ it follows that $x + y \in L^{-1}(V_0)$. Since $L(tx) = t\,L(x) = Lu \in V_0$ it follows that $tx \in L^{-1}(V_0)$. Thus $L^{-1}(V_0)$ is closed under the module operations and so it is a sub algebra of U.

7.8 Theorem: *Let $L : U \to V$ be an R-module homomorphism. Then*

(1) L is surjective if and only if $\mathrm{Range}(L) = V$, *and*
(2) L is injective if and only if $\mathrm{Ker}(L) = \{0\}$.

Proof: Part (1) is simply a restatement of the definition of subjectivity and does not require proof. To Prove Part (2), we begin by remarking that since $L(0) = 0$ we have $\{0\} \subseteq \mathrm{Ker}(L)$. Suppose L is injective. Then $x \in \mathrm{Ker}(L) \Longrightarrow L(x) = 0 \Longrightarrow L(x) = L(0) \Longrightarrow x = 0$ and so $\mathrm{Ker}(L) = \{0\}$. Suppose, conversely, that $\mathrm{Ker}(L) = \{0\}$. Then $L(X) = L(y) \Longrightarrow L(x) - L(y) = 0 \Longrightarrow L(x-y) = 0 \Longrightarrow x - y \in \mathrm{Ker}(L) = \{0\} \Longrightarrow x - y = 0 \Longrightarrow x = y$ and so L is injective.

7.9 Example: The maps

$$L : P_n(R) \to R^{n+1} \text{ given by } L\big(\sum_{i=0}^{n} a_i x^i\big) = (a_0, a_1, \cdots, a_n)$$

$$L : R[x] \to R^{\infty} \text{ given by } L\big(\sum_{i=0}^{n} a_i x^i\big) = (a_0, a_1, \cdots, a_n, 0, 0, \cdots)$$

$$L : R[[x]] \to R^{\omega} \text{ given by } L\big(\sum_{i=0}^{\infty} a_i x^i\big) = (a_0, a_1, a_2, \cdots)$$

are all R-algebra isomorphisms, so we have $P_n(R) \cong R^{n+1}$, $R[x] \cong R^{\infty}$ and $R[[x]] \cong R^{\omega}$.

7.10 Example: The map $L : M_{m\times n}(R) \to R^{m\cdot n}$ given by

$$L\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} = \big(a_{1,1}, a_{1,2}, \cdots, a_{1,n}, a_{2,1}, \cdots, a_{2,n}, \cdots, a_{m,1}, \cdots, a_{m,n}\big)$$

is an R-module isomorphism, so we have $M_{m\times n}(R) \cong R^{m\cdot n}$.

7.11 Example: Let A and B be sets with $|A| = |B|$, and let $g : A \to B$ be a bijection. Then the map $L : R^A \to R^B$ given by $L(f)(b) = f(g^{-1}(b))$, that is by $L(f) = fg^{-1}$, is an R-module isomorphism, and so we have $R^A \cong R^B$. In particular, if $|A| = n$ then we have $R^A \cong R^{\{1,2,\cdots,n\}} = R^n$, and if $|A| = \aleph_0$ then we have $R^A \cong R^{\{1,2,3,\cdots\}} = R^{\omega}$.

7.12 Example: Let $\mathcal{A} = (u_1, u_2, \cdots, u_n)$ be a finite ordered basis for a free R-module U. Then the map $\phi_{\mathcal{A}} : U \to R^n$ given by $\phi_{\mathcal{A}}(x) = [x]_{\mathcal{A}}$ is an R-module isomorphism, so we have $U \cong R^n$.

If $\mathcal{B} = (v_1, v_2, \cdots, v_n)$ is a finite ordered basis for another free R-module V, then the map $\phi_{\mathcal{B}}^{-1}\phi_{\mathcal{A}} : U \to V$ is an R-molule isomorphism, so we have $U \cong V$.

7.13 Example: Let R be a commutative ring. Let $\phi : \mathrm{Hom}(R^n, R^m) \to M_{m\times n}(R)$ be the map given by $\phi(L) = [L] = \big(L(e_1), L(e_2), \cdots, L(e_n)\big) \in M_{m\times n}(R)$. Recall that the inverse of ϕ is the map $\psi : M_{m\times n}(R) \to \mathrm{Hom}(R^n, R^m)$ given by $\psi(A) = L_A$ where $L_A(a) = Ax$. Note that ψ preserves the R-module operations because

$$\phi(A + B) = L_{A+B} = L_A + L_B = \psi(A) + \psi(B) \text{ and}$$
$$\psi(tA) = L_{tA} = t\, L_A = t\, \psi(A).$$

Thus ϕ and ψ are R-module isomorphisms and we have $\mathrm{Hom}(R^n, R^m) \cong M_{m\times n}(R)$. In the case that $m = n$, we also have

$$\psi(AB) = L_{AB} = L_A L_B = \psi(A)\psi(B)$$

and so the maps ϕ and ψ are in fact R-algebra isomorphisms so we have $\mathrm{End}(R^n) \cong M_n(R)$ as R-algebras. By restricting ϕ and ψ to the invertible elements, we also obtain a group isomorphism $\mathrm{Aut}(R^n) \cong GL_n(R)$.

7.14 Theorem: *Let R be a ring and let U and V be free R-modules. Then $U \cong V$ if and only if there exists a basis $\mathcal{A}$ for U and a basis $\mathcal{B}$ for V with $|\mathcal{A}| = |\mathcal{B}|$.*

Proof: Suppose that $U \cong V$. Let $\mathcal{A}$ be a basis for U and let $L : U \to V$ be an isomorphism. Let $\mathcal{B} = L(\mathcal{A}) = \big\{L(u)\big|u \in \mathcal{A}\big\}$. Since L is bijective we have $|\mathcal{A}| = |\mathcal{B}|$. Note that $\mathcal{B}$ spans V because given $y \in V$ we can choose $x \in U$ with $L(x) = y$, then write $x = \sum_{i=1}^{n} t_i u_i$ with $t_i \in R$ and $u_i \in \mathcal{A}$, and then we have

$$y = L(x) = L\Big(\sum_{i=1}^{n} t_i u_i\Big) = \sum_{i=1}^{n} t_i L(u_i) \in \text{Span}\,(\mathcal{B}).$$

It remains to show that $\mathcal{B}$ is linearly independent. Suppose that $\sum_{i=1}^{n} t_i v_i = 0$ where $t_i \in R$ and the v_i are distinct elements in $\mathcal{B}$. For each index i, choose $u_i \in \mathcal{A}$ with $L(u_i) = v_i$, and note that the elements u_i are distinct because L is bijective. We have

$$0 = \sum_{i=1}^{n} t_i v_i = \sum_{i=1}^{n} t_i L(u_i) = L\Big(\sum_{i=1}^{n} t_i u_i\Big).$$

Because L is injective, it follows that $\sum_{i=1}^{n} t_i u_i = 0$ and then, because $\mathcal{A}$ is linearly independent, it follows that each $t_i = 0$. Thus $\mathcal{B}$ is linearly independent, as required.

Conversely, suppose that $\mathcal{A}$ is a basis for U and $\mathcal{B}$ is a basis for V with $|\mathcal{A}| = |\mathcal{B}|$. Let $g : A \to B$ be a bijection. Define a map $L : U \to V$ as follows. Given $x \in U$, write $x = \sum_{i=1}^{n} t_i u_i$ where $t_i \in R$ and the u_i are distinct elements in $\mathcal{A}$, and then define $L(x) = \sum_{i=1}^{n} t_i g(u_i)$. Note that L is an R-module homomorphism because for $r \in R$ and for $x = \sum_{i=1}^{n} s_i u_i$ and $y = \sum_{i=1}^{n} t_i u_i$ (where we are using the same elements u_i in both sums with some of the coefficients equal to zero), we have

$$L(rx) = L\Big(\sum_{i=1}^{n} (rs_i)u_i\Big) = \sum_{i=1}^{n} rs_i g(u_i) = r\sum_{i=1}^{n} s_i g(u_i) = rL(x)\ \text{, and}$$

$$L(x+y) = L\Big(\sum_{i=1}^{n} (s_i + t_i)u_i\Big) = \sum_{i=1}^{n} (s_i + t_i)g(u_i) = \sum_{i=1}^{n} s_i g(u_i) + \sum_{i=1}^{n} t_i g(u_i) = L(x) + L(y).$$

Also note that L is bijective with inverse $M : V \to U$ given by $M\Big(\sum_{i=1}^{n} t_i v_i\Big) = \sum_{i=1}^{n} t_i g^{-1}(v_i)$, where $t_i \in R$ and the v_i are distinct elements in $\mathcal{B}$.

7.15 Corollary: *Let F be a field and let U and V be vector spaces over F. Then*

$$U \cong V \iff \dim(U) = \dim(V).$$

7.16 Remark: When U and V are modules over a commutative ring R, we have

$$U \cong V \iff \text{rank}(U) = \text{rank}(V),$$

but we have not built up enough machinery to prove this result.

7.17 Theorem: *Let R be a ring, let U be a free R-module, and let V be any R-module. Let $\mathcal{A}$ be basis for U and, for each $u \in \mathcal{A}$, let $v_u \in V$. Then there exists a unique R-module homomorphism $L : U \to V$ with $L(u) = v_u$ for all $u \in \mathcal{A}$.*

Proof: Note that if $L : U \to V$ is an R-module homomorphism with $L(u) = v_u$ for all $u \in U$, then for $t_i \in R$ and $u_i \in \mathcal{A}$ we have

$$L\Big(\sum_{i=1}^{n} t_i u_i\Big) = \sum_{i=1}^{n} t_i L(u_i) = \sum_{i=1}^{n} t_i v_{u_i}.$$

This shows that the map L is unique and must be given by the above formula.

To prove existence, we define $L : U \to V$ by $L\Big(\sum_{i=1}^{n} t_i u_i\Big) = \sum_{i=1}^{n} t_i v_{u_i}$ where $t_i \in R$ and $u_i \in \mathcal{A}$, and we note that L is an R-module homomorphism because for $x = \sum_{i=1}^{n} s_i u_i$ and $y = \sum_{i=1}^{n} t_i u_i$ (using the same elements u_i in both sums) and $r \in R$ we have

$$L(rx) = L\Big(\sum_{i=1}^{n} r s_i u_i\Big) = \sum_{i=1}^{n} r s_i v_{u_i} = r \sum_{i=1}^{n} s_i v_{u_i} = r\, L(x) \text{ , and}$$
$$L(x+y) = L\Big(\sum_{i=1}^{n} (s_i + t_i) u_i\Big) = \sum_{i=1}^{n} (s_i + t_i) v_{u_i} = \sum_{i=1}^{n} s_i v_{u_i} + \sum_{i=1}^{n} t_i v_{u_i} = L(x) + L(y).$$

7.18 Corollary: *Let R be a commutative ring, let U be a free R-module with basis $\mathcal{A}$ and let V be an R-module. Then the map $\phi : \mathrm{Hom}(U, V) \to V^{\mathcal{A}}$, given by $\phi(L)(u) = L(u)$ for all $u \in \mathcal{A}$, is an R-module isomorphism, and so we have $\mathrm{Hom}(U, V) \cong V^{\mathcal{A}}$.*

Proof: The above theorem states that the map ϕ is bijective, and we note that ϕ is an R-module homomorphism because for $L, M \in \mathrm{Hom}(U, V)$ and $t \in R$ we have

$$\phi(L+M)(u) = (L+M)(u) = L(u) + M(u) = \phi(L)(u) + \phi(M)(u) \text{ , and}$$
$$\phi(tL)(u) = (tL)(u) = t\, L(u) = t\phi(L)(u)$$

for all $u \in \mathcal{A}$ hence $\phi(L+M) = \phi(L) + \phi(M)$ and $\phi(tL) = t\phi(L)$.

7.19 Example: For a module U over a commutative ring R, the **dual module** of U is the R-module

$$U^* = \mathrm{Hom}(U, R).$$

By the above corollary, when U is a free module with basis $\mathcal{A}$. we have $U^* \cong R^{\mathcal{A}}$, and if $|\mathcal{A}| = n$ then we have $U^* \cong R^n \cong U$. When U is a vector space over a field F, U^* is called the **dual vector space** of U.

7.20 Definition: Let R be a commutative ring and let U and V be free R-modules with finite bases. Let $\mathcal{A}$ and $\mathcal{B}$ be finite ordered bases for U and V respectively, with $|\mathcal{A}| = n$ and $|\mathcal{B}| = m$. For $L \in \mathrm{Hom}(U, V)$, we define the **matrix** of L with respect to $\mathcal{A}$ and $\mathcal{B}$ to be the matrix

$$[L]_{\mathcal{B}}^{\mathcal{A}} = \left[\phi_{\mathcal{B}} L\, \phi_{\mathcal{A}}^{-1}\right] \in M_{m \times n}(R).$$

When $L \in \mathrm{End}(U)$ we write $[L]_{\mathcal{A}}$ for the matrix $[L]_{\mathcal{A}}^{\mathcal{A}} \in M_n(R)$.

7.21 Theorem: *Let R be a commutative ring. Let $\mathcal{A}$ and $\mathcal{B}$ be finite ordered bases for free R-modules U and V, respectively, with $|\mathcal{A}| = n$ and $|\mathcal{B}| = m$. Let $L \in \mathrm{Hom}(U,V)$. Then*

(1) $[L]_{\mathcal{B}}^{\mathcal{A}}$ is the matrix such hat $[L]_{\mathcal{B}}^{\mathcal{A}}[u]_{\mathcal{A}} = \big[L(u)\big]_{\mathcal{B}}$ for all $u \in U$, and
(2) if $\mathcal{A} = (u_1, u_2, \cdots, u_n)$ and $\mathcal{B} = (v_1, v_2, \cdots, v_m)$ then

$$[L]_{\mathcal{B}}^{\mathcal{A}} = \Big(\big[L(u_1)\big]_{\mathcal{B}}, \big[L(u_2)\big]_{\mathcal{B}}, \cdots, \big[L(u_n)_{\mathcal{B}}\big]\Big) \in M_{m\times n}(R).$$

Proof: Part (1) holds because for $u \in U$ and $x = \phi_{\mathcal{A}}(u) = [u]_{\mathcal{A}}$ we have

$$[L]_{\mathcal{B}}^{\mathcal{A}}[u]_{\mathcal{A}} = \big[\phi_{\mathcal{B}} L\, \phi_{\mathcal{A}}^{-1}\big]\phi_{\mathcal{A}}(u) = \big(\phi_{\mathcal{B}} L\, \phi_{\mathcal{A}}^{-1}\big)\big(\phi_{\mathcal{A}}(u)\big)\phi_{\mathcal{B}}\big(L(u)\big) = \big[L(u)\big]_{\mathcal{B}}.$$

Part (2) follows from Part (1) because for each index k, the k^{th} column of $[L]_{\mathcal{B}}^{\mathcal{A}}$ is

$$[L]_{\mathcal{B}}^{\mathcal{A}}(e_k) = [L]_{\mathcal{B}}^{\mathcal{A}}[u_k]_{\mathcal{A}} = \big[L(u_k)\big]_{\mathcal{B}}.$$

7.22 Theorem: *Let R be a commutative ring. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be finite ordered bases for free R-modules U, V and W, respectively.*

(1) For $L, M \in \mathrm{Hom}(U,V)$ and $t \in R$ we have $\big[L+M\big]_{\mathcal{B}}^{\mathcal{A}} = [L]_{\mathcal{B}}^{\mathcal{A}} + [M]_{\mathcal{B}}^{\mathcal{A}}$ and $\big[t\,L\big]_{\mathcal{B}}^{\mathcal{A}} = t\,[L]_{\mathcal{B}}^{\mathcal{A}}$.
(2) For $L \in \mathrm{Hom}(U,V)$ and $M \in \mathrm{Hom}(V,W)$ we have $\big[ML\big]_{\mathcal{C}}^{\mathcal{A}} = [M]_{\mathcal{C}}^{\mathcal{B}}[L]_{\mathcal{B}}^{\mathcal{A}}$.

Proof: We prove Part (2), leaving the (similar) proof of Part (1) as an exercise. Let $L \in \mathrm{Hom}(U,V)$ andn let $M \in \mathrm{Hom}(V,W)$. Say $|\mathcal{A}| = n$, $|\mathcal{B}| = m$ and $|\mathcal{C}| = l$. Let $x \in R^n$. Choose $u \in U$ with $[u]_{\mathcal{A}} = \phi_{\mathcal{A}}(u) = x$. Then

$$\big[ML\big]_{\mathcal{C}}^{\mathcal{A}}x = \big[ML\big]_{\mathcal{C}}^{\mathcal{A}}[u]_{\mathcal{A}} = \big[M(L(u))\big]_{\mathcal{C}} = \big[M\big]_{\mathcal{C}}^{\mathcal{B}}\big[L(u)\big]_{\mathcal{B}} = \big[M\big]_{\mathcal{C}}^{\mathcal{B}}\big[L\big]_{\mathcal{B}}^{\mathcal{A}}[u]_{\mathcal{A}} = \big[M\big]_{\mathcal{C}}^{\mathcal{B}}\big[L\big]_{\mathcal{B}}^{\mathcal{A}}x.$$

Since $\big[ML\big]_{\mathcal{C}}^{\mathcal{A}}x = \big[M\big]_{\mathcal{C}}^{\mathcal{B}}\big[L\big]_{\mathcal{B}}^{\mathcal{A}}x$ for all $x \in R^n$ it follows that $\big[ML\big]_{\mathcal{C}}^{\mathcal{A}} = \big[M\big]_{\mathcal{C}}^{\mathcal{B}}\big[L\big]_{\mathcal{B}}^{\mathcal{A}}$.

7.23 Corollary: *Let R be a commutative ring. Let $\mathcal{A}$ and $\mathcal{B}$ be finite ordered bases for free R-modules U and V. Then the map $\phi_{\mathcal{B}}^{\mathcal{A}} : \mathrm{Hom}(U,V) \to M_{m\times n}(R)$ given by $\phi_{\mathcal{B}}^{\mathcal{A}}(L) = [L]_{\mathcal{B}}^{\mathcal{A}}$ is an R-module isomorphism, and the map $\phi_{\mathcal{A}} : \mathrm{End}(U) \to M_n(R)$ given by $\phi_{\mathcal{A}}(L) = [L]_{\mathcal{A}}$ is an R-algebra isomorphism which restricts to a group isomorphism $\phi_{\mathcal{A}} : \mathrm{Aut}(U) \to GL_n(R)$.*

7.24 Corollary: *(Change of Basis) Let R be a commutative ring. Let U and V be free R-modules. Let $\mathcal{A}$ and $\mathcal{C}$ be two ordered bases for U with $|\mathcal{A}| = |\mathcal{C}| = n$ and let $\mathcal{B}$ and $\mathcal{D}$ be two ordered bases for V with $|\mathcal{B}| = |\mathcal{D}| = m$. For $L \in \mathrm{Hom}(U,V)$ and $u \in U$ we have*

$$[u]_{\mathcal{C}} = [I_U]_{\mathcal{C}}^{\mathcal{A}}[u]_{\mathcal{A}} \quad \text{and} \quad [L]_{\mathcal{D}}^{\mathcal{C}} = [I_V]_{\mathcal{D}}^{\mathcal{B}}[L]_{\mathcal{B}}^{\mathcal{A}}[I_U]_{\mathcal{A}}^{\mathcal{C}}$$

where I_U and I_V are the identity maps on U and V.

Proof: By Part (1) of Theorem 7.21 we have $[u]_{\mathcal{C}} = \big[I_U(u)\big]_{\mathcal{C}} = [I_U]_{\mathcal{C}}^{\mathcal{A}}[u]_{\mathcal{A}}$ and by Part (2) of Theorem 7.22

$$\big[L\big]_{\mathcal{D}}^{\mathcal{C}} = \big[I_V\, L\, I_U\big]_{\mathcal{D}}^{\mathcal{C}} = [I_V]_{\mathcal{D}}^{\mathcal{B}}[L]_{\mathcal{B}}^{\mathcal{A}}[I_U]_{\mathcal{A}}^{\mathcal{C}}.$$

7.25 Definition: Let $\mathcal{A} = (u_1, u_2, \cdots, u_n)$ and $\mathcal{B} = (v_1, v_2, \cdots, v_n)$ be two finite ordered bases for a module U over a commutative ring R. The matrix

$$[I]_{\mathcal{B}}^{\mathcal{A}} = \big([u_1]_{\mathcal{B}}, [u_2]_{\mathcal{B}}, \cdots, [u_n]_{\mathcal{B}}\big) \in M_n(R)$$

is called the **change of basis matrix** from $\mathcal{A}$ to $\mathcal{B}$. Note that $[I]_{\mathcal{B}}^{\mathcal{A}}$ is invertible with

$$\Big([I]_{\mathcal{B}}^{\mathcal{A}}\Big)^{-1} = [I]_{\mathcal{A}}^{\mathcal{B}}.$$

7.26 Note: Let $\mathcal{A}$ and $\mathcal{B}$ be two finite ordered bases, with $|\mathcal{A}| = |\mathcal{B}|$, for a free module U over a commutative ring R. For $L \in \text{End}(U)$, the Change of Basis Theorem gives

$$[L]_{\mathcal{B}} = [L]_{\mathcal{B}}^{\mathcal{B}} = [I]_{\mathcal{B}}^{\mathcal{A}}[L]_{\mathcal{A}}^{\mathcal{A}}[I]_{\mathcal{A}}^{\mathcal{B}}.$$

If we let $A = [L]_{\mathcal{A}}$ and $B = [L]_{\mathcal{B}}$ and $P = [I]_{\mathcal{B}}^{\mathcal{A}}$ then the formula becomes

$$B = PAP^{-1}.$$

7.27 Note: Given a finite ordered basis $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$ for a free R-module U over a commutative ring R, and given an invertible matrix $P \in GL_n(R)$, if we choose $u_k \in U$ with $[u_k]_{\mathcal{B}} = Pe_k$, then $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ is an ordered basis for U such that $[I]_{\mathcal{B}}^{\mathcal{A}} = P$. Thus every invertible matrix $P \in GL_n(R)$ is equal to a change of basis matrix.

7.28 Definition: Let R be a commutative ring. For $A, B \in M_n(R)$, we say that A and B are **similar**, and we write $A \sim B$, when $B = PAP^{-1}$ for some $P \in GL_n(R)$.

7.29 Note: Let R be a commutative ring, and let $A, B \in M_n(R)$ with $A \sim B$. Choose $P \in GL_n(R)$ so that $B = PAP^{-1}$. Then we have

$$\det(B) = \det(PAP^{-1}) = \det(P)\det(A)\det(P)^{-1} = \det(A).$$

Thus similar matrices have the same determinant.

7.30 Definition: Let F be a field and let U be a finite dimensional vector space over F. For $L \in \text{End}(U)$, we define the **determinant** of L to be

$$\det(L) = \det\left([L]_{\mathcal{A}}\right)$$

where $\mathcal{A}$ is any ordered basis for U.

Chapter 8. Eigenvalues, Eigenvectors and Diagonalization

8.1 Definition: For a square matrix $D \in M_n(R)$ with entries in a ring R, we say that D is a **diagonal** matrix when $D_{k,l} = 0$ whenever $k \neq l$. For $\lambda_1, \lambda_2, \cdots, \lambda_n \in R$, we write $D = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ for the diagonal matrix D with $D_{k,k} = \lambda_k$ for all indices k.

8.2 Definition: Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. We say that that L is **diagonalizable** when there exists an ordered basis $\mathcal{A}$ for U such that $[L]_{\mathcal{A}}$ is diagonal.

8.3 Note: Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. When $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ is an ordered basis for U and $\lambda_1, \lambda_2, \cdots, \lambda_n \in F$, we have

$$[L]_{\mathcal{A}} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \iff \big[L(u_k)\big]_{\mathcal{A}} = \lambda_k e_k \text{ for all } k \iff L(u_k) = \lambda_k u_k \text{ for all } k$$

Thus L is diagonalizable if and only if there exists an ordered basis $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ for U and there exist $\lambda_1, \lambda_2, \cdots, \lambda_n \in F$ such that $L(u_k) = \lambda_k u_k$ for all k.

8.4 Definition: Let $L \in \text{End}(U)$ where U is a vector space over a field F. For $\lambda \in F$, we say that λ is an **eigenvalue** (or a **characteristic value**) of L when there exists a nonzero vector $0 \neq u \in F^n$ such that $L(u) = \lambda u$. Such a vector $0 \neq u \in U$ is called an **eigenvector** (or **characteristic vector**) of L for λ. The **spectrum** of L is the set

$$\text{Spec}(L) = \big\{\lambda \in F \big| \lambda \text{ is an eigenvalue of } L\big\}.$$

For $\lambda \in F$, the **eigenspace** of L for λ is the subspace

$$E_\lambda = \big\{u \in U \big| L(u) = \lambda u\big\} = \big\{u \in U \big| (L - \lambda I)u = 0\big\} = \text{Ker}(L - \lambda I) \subseteq U.$$

Note that E_λ consists of the eigenvectors for λ together with the zero vector.

8.5 Note: Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. For $\lambda \in F$

$$\begin{aligned} \lambda \text{ is an eigenvalue of } L &\iff \text{there exists } 0 \neq u \in \text{Ker}(L - \lambda I) \\ &\iff (L - \lambda I) \text{ is not invertible} \\ &\iff \det(L - \lambda I) = 0 \\ &\iff \lambda \text{ is a root of } f(x) = \det(L - xI). \end{aligned}$$

Note that when $\mathcal{A}$ is any ordered basis for U, we have

$$f(x) = \det(L - xI) = \det\big(\big[L - xI\big]_{\mathcal{A}}\big) = \det\big([L]_{\mathcal{A}} - xI\big) \in P_n(F).$$

8.6 Definition: Let $L \in \text{End}(U)$ where U is an n-dimensional vector space over a field F. The **characteristic polynomial** of L is the polynomial

$$f_L(x) = \det(L - xI) \in P_n(F).$$

Note that $\text{Spec}(L)$ is the set of roots of $f_L(x)$.

8.7 Note: Let $L \in \mathrm{End}(U)$ where U is an n-dimensional vector space over a field F. Recall that L is diagonalizable if and only if there exists an ordered basis $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ for U such that each u_k is an eigenvector for some eigenvalue λ_k. The eigenvalues of L are the roots of $f_L(x)$, so there are at most n possible distinct eigenvalues. For each eigenvalue, the largest number of linearly independent eigenvectors for λ is equal to the dimension of E_λ. We can try to diagonalize L be finding all the eigenvalues λ for L, then finding a basis for each eigenspace E_λ, then selecting an ordered basis $\mathcal{A}$ from the union of the bases of the eigenspaces. In particular, note that if $\sum_{\lambda \in \mathrm{Spec}(L)} \dim(E_\lambda) < n$ then L cannot be diagonalizable.

8.8 Definition: Let F be a field and let $A \in M_n(F)$. By identifying A with the linear map $L = L_A \in \mathrm{End}(F^n)$ given by $L(x) = Ax$, all of the above definitions and remarks may be applied to the matrix A. The matrix A is **diagonalizable** when there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$ (or equivalently $P^{-1}AP = D$). An **eigenvalue** for A is an element $\lambda \in F$ for which there exists $0 \neq x \in F^n$ such that $Ax = \lambda x$, and then such a vector x is called an **eigenvector** of A for λ. The set of eigenvalues of A, denoted by $\mathrm{Spec}(A)$, is called the **spectrum** of A. For $\lambda \in F$, the **eigenspace** for λ is the vector space $E_\lambda = \mathrm{Null}(A - \lambda I)$. The **characteristic polynomial** of A is the polynomial $f_A(x) = \det(A - xI) \in P_n(F)$,

8.9 Example: Let $A = \begin{pmatrix} 3 & -1 \\ 4 & -1 \end{pmatrix} \in M_2(F)$ where $F = \mathbf{R}$ or $\mathbf{C}$. The characteristic polynomial of A is

$$f_A(x) = \det(A - xI) = \begin{vmatrix} 3-x & -1 \\ 4 & -1-x \end{vmatrix} = (x-3)(x+1) + 4 = x^2 - 2x + 1 = (x-1)^2$$

so the only eigenvalue of A is $\lambda = 1$. When $\lambda = 1$ we have

$$(A - \lambda I) = (A - I) = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix}$$

so the eigenspace $E_1 = \mathrm{Null}(A - I)$ has basis $\{u\}$ where $u = (1, 2)^T$. Since

$$\sum_{\lambda \in \mathrm{Spec}(A)} \dim(E_\lambda) = \dim(E_1) = 1 < 2,$$

we see that A is not diagonalizable.

8.10 Example: Let $A = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \in M_2(F)$ where $F = \mathbf{R}$ or $\mathbf{C}$. The characteristic polynomial of A is

$$f_A(x) = \begin{vmatrix} 1-x & -2 \\ 2 & 1-x \end{vmatrix} = (x-1)^2 + 4 = x^2 - 2x + 5.$$

For $x \in \mathbf{C}$, we have $f_A(x) = 0 \iff x = \frac{2 \pm \sqrt{4-20}}{2} = 1 \pm 2i$. When $F = \mathbf{R}$, A has no eigenvalues (in $\mathbf{R}$) and so A is not diagonalizable. When $F = \mathbf{C}$, the eigenvalues of A are $\lambda_1 = 1 + 2i$ and $\lambda_2 = 1 - 2i$. As an exercise, show that when $\lambda = \lambda_1$ the eigenspace E_{λ_1} has basis $\{u_1\}$ where $u_1 = (i, 1)^T$, and when $\lambda = \lambda_2$ the eigenspace E_{λ_2} has basis $u_2 = (-i, 1)^T$, then verify that the matrix $P = (u_1, u_2) \in M_2(\mathbf{C})$ is invertible and that $P^{-1}AP = \mathrm{diag}(\lambda_1, \lambda_2)$ thus showing that A is diagonalizable.

8.11 Theorem: *Let $L \in \mathrm{End}(U)$ where U is a vector space over a field F. Let $\lambda_1, \lambda_2, \cdots, \lambda_\ell \in F$ be distinct eigenvalues of L. For each index k, let $0 \neq u_k \in U$ be an eigenvector of L for λ_k. Then $\{u_1, u_2, \cdots, u_\ell\}$ is linearly independent.*

Proof: Since $u_1 \neq 0$ the set $\{u_1\}$ is linearly independent. Suppose, inductively, that the set $\{u_1, u_2, \cdots, u_{\ell-1}\}$ is linearly independent. Suppose that $\sum_{i=1}^{\ell} t_i u_i = 0$ with $t_i \in F$. Note that

$$0 = (L - \lambda_\ell I)\big(\sum_{i=1}^{\ell} t_i u_i\big) = \sum_{i=1}^{\ell} t_i(L(u_i) - \lambda_\ell u_i) = \sum_{i=1}^{\ell} t_i(\lambda_i - \lambda_\ell)u_i = \sum_{i=1}^{\ell-1} t_i(\lambda_i - \lambda_\ell)u_i$$

and so $t_i = 0$ for $1 \leq i < \ell$ since $\{u_1, u_2, \cdots, u_{\ell-1}\}$ is linearly independent. Since $t_i = 0$ for $1 \leq i \leq \ell - 1$ and $\sum_{i=1}^{\ell} t_i u_i = 0$, we also have $t_\ell = 0$. Thus $\{u_1, u_2, \cdots, u_\ell\}$ is linearly independent.

8.12 Corollary: *Let $L \in \mathrm{End}(U)$ where U is a vector space over a field F. Let $\lambda_1, \lambda_2, \cdots, \lambda_\ell$ be distinct eigenvalues of L. For each index k, let $\mathcal{A}_k$ be a linearly independent set of eigenvectors for λ_k. Then $\bigcup_{k=1}^{\ell} \mathcal{A}_k$ is linearly independent.*

Proof: Suppose that $\sum_{k=1}^{\ell} \sum_{i=1}^{m_k} t_{k,i} u_{k,i} = 0$ where each $t_{k,i} \in F$ and for each k, the vectors $u_{k,i}$ are distinct vectors in $\mathcal{A}_k$. Then we have $\sum_{i=1}^{\ell} u_k = 0$ where $u_k = \sum_{i=1}^{m_k} t_{k,i} u_{k,i} \in E_{\lambda_k}$. From the above theorem, it follows that $u_k = 0$ for all k, because if we had $u_k \neq 0$ for some values of k, say the values $k_1, k_2, \cdots, k_r$, then $\{u_{k_1}, u_{k_2}, \cdots, u_{k_r}\}$ would be linearly independent but $\sum_{i=1}^{r} u_{k_r} = 0$, which is impossible. Since for each index k we have $0 = u_k = \sum_{i=1}^{m_k} t_{k,i} u_{k,i}$ it follows that each $t_{k,i} = 0$ because $\mathcal{A}_k$ is linearly independent.

8.13 Corollary: *Let $L \in \mathrm{End}(U)$ where U is a finite dimensional vector space over a field F. Then*

$$L \text{ is diagonalizable if and only if } \sum_{\lambda \in \mathrm{Spec}(L)} \dim(E_\lambda) = \dim U.$$

In this case, if $\mathrm{Spec}(L) = \{\lambda_1, \lambda_2, \cdots, \lambda_\ell\}$ and, for each k, $\mathcal{A}_k = \{u_{k,1}, u_{k,2}, \cdots, u_{k,m_k}\}$ is an ordered basis for E_{λ_k}, and then

$$\mathcal{A} = \bigcup_{k=1}^{\ell} \mathcal{A}_k = \big\{u_{1,1}, u_{1,2}, \cdots, u_{1,m_1}, u_{2,1}, u_{2,2}, \cdots, u_{2,m_2}, \cdots, u_{\ell,1}, u_{\ell,2}, \cdots, u_{\ell,m_\ell}\big\}$$

is an ordered basis for U such that $[L]_{\mathcal{A}}$ is diagonal.

8.14 Definition: Let F be a field. For $f \in F[x]$ and $a \in F$, the **multiplicity** of a as a root of f, denoted by $\text{mult}(a, f(x))$, is the smallest $m \in \mathbf{N}$ such that $(x-a)^m$ is a factor of $f(x)$. Note that a is a root of f if and only if $\text{mult}(a, f) > 0$. For a non-constant polynomial $f \in F[x]$, we say that f **splits** (over F) when f factors into a product of linear factors in $F[x]$, that is when f is of the form $f(x) = c \prod_{i=1}^{n} (x - a_i)$ for some $a_i \in F$.

8.15 Theorem: *Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. Let $\lambda \in \text{Spec}(L)$ and let $m_\lambda = \text{mult}\big(\lambda, f_L(x)\big)$. Then*

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

Proof: Since λ is an eigenvalue of L we have $E_\lambda \neq \{0\}$ so $\dim(E_\lambda) \geq 1$. Let $m = \dim(E_\lambda)$ and let $\mathcal{A} = (u_1, u_2, \cdots, u_m)$ be an ordered basis for E_λ. Extend $\mathcal{A}$ to an ordered basis $\mathcal{B} = (u_1, \cdots, u_m, \cdots, u_n)$ for U. Since $L(u_i) = \lambda u_i$ for $1 \leq i \leq m$, the matrix $[L]_\mathcal{B}$ is of the form

$$[L]_\mathcal{B} = \begin{pmatrix} \lambda I & A \\ 0 & B \end{pmatrix} \in M_n(F)$$

where $I \in M_m(F)$. The characteristic polynomial of L is

$$f_L(x) = \begin{vmatrix} (\lambda - x)I & A \\ 0 & B - xI \end{vmatrix} = (\lambda - x)^m f_B(x).$$

Thus $(x - \lambda)^m$ is a factor of $f_L(x)$ and so $m_\lambda = \text{mult}\big(\lambda, f_L(x)\big) \geq m$.

8.16 Corollary: *Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. Then L is diagonalizable if and only if $f_L(x)$ splits and $\dim(E_\lambda) = \text{mult}\big(\lambda, f_L(x)\big)$ for every $\lambda \in \text{Spec}(L)$.*

Proof: Suppose that L is diagonalizable. Choose an ordered basis $\mathcal{A}$ so that $[L]_\mathcal{A}$ is diagonal, say $[L]_\mathcal{A} = D = \text{diag}(\lambda_1, \cdots, \lambda_n)$. Note that $f_L(x) = f_D(x) = \prod_{k=1}^{n} (\lambda_k - x)$, and so $f_L(x)$ splits. For each $\lambda \in \text{Spec}(L)$, let $m_\lambda = \text{mult}(\lambda, f_L(x))$. Then, by the above theorem together with Corollary 8.13, we have

$$n = \dim(U) = \sum_{\lambda \in \text{Spec}(L)} \dim(E_\lambda) \leq \sum_{\lambda \in \text{Spec}(L)} m_\lambda = \deg(f_L) = n$$

which implies that $\dim(E_\lambda) = m_\lambda$ for all λ. Conversely, if f_L splits and $\dim(E_\lambda) = m_\lambda$ for all λ then

$$\sum_{\lambda \in \text{Spec}(L)} \dim(E_\lambda) = \sum_{\lambda \in \text{Spec}(L)} m_\lambda = \deg(f_L) = n = \dim(U)$$

and so L is diagonalizable.

8.17 Corollary: *Let $A \in M_n(F)$ where F is a field. Then A is diagonalizable if and only if $f_A(x)$ splits and $\dim(E_\lambda) = \text{mult}(\lambda, f_A(x))$ for all $\lambda \in \text{Spec}(A)$.*

8.18 Note: To summarize the above results, given a matrix $A \in M_n(F)$, where F is a field, we can determine whether A is diagonalizable as follows. We find the characteristicc polynomial $f_A(x) = \det(A - xI)$. We factor $f_A(x)$ to find the eigenvalues of A and the multiplicity of each eigenvalue. If $f_A(x)$ does not split then A is not diagonalizable. If $f_A(x)$ does split, then for each eigenvalue λ with multiplicity $m_\lambda \geq 2$, we calculate $\dim(E_\lambda)$. If we find one eigenvalue λ for which $\dim(E_\lambda) < m_\lambda$ then A is not diagonalizable. Otherwise A is diagonalizable. In particular we remark that if $f_A(x)$ splits and has n distinct roots (so the eigenvalues all have multiplicity 1) then A is diagonalizable.

In the case that A is diagonalizable and $f_A(x) = (-1)^n \prod_{k=1}^{\ell} (x - \lambda_k)^{m_k}$, if we find an ordered basis $\mathcal{A}_k = \{u_{k,1}, u_{k_2}, \cdots, u_{k,m_k}\}$ for each eigenspace, then we have $P^{-1}AP = D$ with

$$P = \big(u_{1,1}, u_{1,2}, \cdots, u_{1,m_1}, u_{2,1}, u_{2,2}, \cdots, u_{2,m_2}, \cdots, u_{\ell,1}, u_{\ell,2}, \cdots, u_{\ell,m_\ell}\big)$$
$$D = \operatorname{diag}(\lambda_1, \lambda_1, \cdots, \lambda_1, \lambda_2, \lambda_2, \cdots, \lambda_2, \cdots, \lambda_\ell, \lambda_\ell, \cdots, \lambda_\ell)$$

where each λ_k is repeated m_k times.

8.19 Example: Let $A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{pmatrix} \in M_3(\mathbf{Q})$. Determine whether A is diagonalizable and, if so, find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

Solution: The characteristic polynomial of A is

$$\begin{aligned} f_A(x) = \big|A - xI\big| &= \begin{vmatrix} 3-x & 1 & 1 \\ 2 & 4-x & 2 \\ -1 & -1 & 1-x \end{vmatrix} \\ &= -(x-3)(x-4)(x-1) - 2 - 2 - 2(x-3) + 2(x-1) - (x-4) \\ &= -(x^3 - 8x^2 + 19x - 12) - x + 4 = -(x^3 - 8x^2 + 20x - 16) \\ &= -(x-2)(x^2 - 6x + 8) = -(x-2)^2(x-4) \end{aligned}$$

so the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 4$ of multiplicities $m_1 = 2$ and $m_2 = 1$. When $\lambda = \lambda_1 = 2$ we have

$$A - \lambda I = A - 2I = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so the eigenspace E_2 has basis $\{u_1, u_2\}$ with $u_1 = (-1, 0, 1)^T$ and $u_2 = (-1, 1, 0)^T$. When $\lambda = \lambda_2 = 4$ we have

$$A - \lambda I = A - 4I = \begin{pmatrix} -1 & 1 & 1 \\ 2 & 0 & 2 \\ -1 & -1 & -3 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 & -1 \\ 0 & 2 & 4 \\ 0 & 2 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

so the eigenspace E_4 has basis $\{u_3\}$ where $u_3 = (-1, -2, 1)^T$. Thus we have $P^{-1}AP = D$ where

$$P = \big(u_1, u_2, u_3\big) = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & -2 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad D = \operatorname{diag}\big(\lambda_1, \lambda_1, \lambda_2\big) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

8.20 Definition: For a square matrix $T \in M_n(R)$ with entries in a ring R, we say that T is **upper triangular** when $T_{k,l} = 0$ whenever $k > l$.

8.21 Definition: Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over a field F. We say that L is (upper) **triangularizable** when there exists an ordered basis $\mathcal{A}$ for U such that $[L]_{\mathcal{A}}$ is upper triangular.

8.22 Definition: For a square matrix $A \in M_n(F)$, where F is a field, we say that A is (upper) **triangularizable** when there exists an invertible matrix $P \in GL_n(F)$ such that $P^{-1}AP$ is upper triangular.

8.23 Theorem: *(Schur's Theorem) Let F be a field.*

(1) Let $L \in \text{End}(U)$ where U is a finite dimensional vector space over F. Then L is triangularizable if and only if $f_L(x)$ splits, and
(2) Let $A \in M_n(F)$. Then A is triangularizable if and only if $f_A(x)$ splits.

Proof: We shall prove Part (2), and we leave it as an exercise to show that Part (1) holds if and only if Part (2) holds. Suppose first that A is triangularizable. Choose an invertible matrix P and an upper triangular matrix T with $P^{-1}AP = T$. Then

$$f_A(x) = f_T(x) = \prod_{k=1}^{n} (T_{k,k} - x)$$

and so $f_A(x)$ splits.

Conversely, suppose that $f_A(x)$ splits. Choose a root λ_1 of $f_A(x)$ and note that λ_1 is an eigenvalue of A. Choose an eigenvector u_1 for λ_1, so we have $Au_1 = \lambda_1 u_1$. Since $u_1 \neq 0$ the set $\{u_1\}$ is linearly independent. Extend the set $\{u_1\}$ to a basis $\mathcal{A} = \{u_1, u_2, \cdots, u_n\}$ for F^n. Let $Q = (u_1, u_2, \cdots, u_n) \in M_n(F)$, and note that Q is invertible because $\mathcal{A}$ is a basis for F^n. Since $Q^{-1}Q = I$, the first column of $Q^{-1}Q$ is equal to e_1, so we have

$$\begin{aligned} Q^{-1}AQ &= Q^{-1}A(u_1, u_2, \cdots, u_n) = Q^{-1}(Au_1, Au_2, \cdots, Au_n) \\ &= Q^{-1}\big(\lambda_1 u_1, A(u_2, \cdots, u_n)\big) = \big(\lambda_1 Q^{-1}u_1, Q^{-1}A(u_2, \cdots, u_n)\big) \\ &= \big(\lambda_1 e_1, Q^{-1}A(u_2, \cdots, u_n)\big) = \begin{pmatrix} \lambda_1 & x^T \\ 0 & B \end{pmatrix} \end{aligned}$$

with $x \in F^{n-1}$ and $B \in M_{n-1}(F)$.. Note that $f_A(x) = (x - \lambda_1) f_B(x)$ and so $f_B(x)$ splits. We suppose, inductively, that B is triangularizable. Choose $R \in GL_{n-1}(F)$ so that $R^{-1}BR = S$ with S upper-triangular. Let $P = Q \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \in M_n(F)$. Then P is invertible with $P^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & R^{-1} \end{pmatrix} Q^{-1}$ and

$$\begin{aligned} P^{-1}AP &= \begin{pmatrix} 1 & 0 \\ 0 & R^{-1} \end{pmatrix} Q^{-1}AQ \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & R^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 & x^T \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & R^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 & x^T R \\ 0 & BR \end{pmatrix} = \begin{pmatrix} \lambda_1 & x^T R \\ 0 & R^{-1}BR \end{pmatrix} = \begin{pmatrix} \lambda_1 & x^T R \\ 0 & S \end{pmatrix} \end{aligned}$$

which is upper triangular.