1 Background and Review Material

1.1 Algebraic Structures

For a nonempty set $X$, a relation $\sim$ on $X$ is an equivalence relation if the following axioms hold:

\begin{align*}
  x \sim x, \forall x \in X & \quad \text{(Reflexivity)} \quad (1.1) \\
  x \sim y \Rightarrow y \sim x & \quad \text{(Symmetry)} \quad (1.2) \\
  x \sim y \sim z \Rightarrow x \sim z & \quad \text{(Transitivity)} \quad (1.3)
\end{align*}

Given an equivalence relation $\sim$ on a set $X$, the equivalence class of $x \in X$ is the set of all elements equivalent to $x$ and denoted by

$$
[x] = \{ y \in X | x \sim y \}.
$$

Recall that a field is a nonempty set $\mathbb{F}$ together with two operations

\begin{align*}
  (x, y) \mapsto x + y & \quad \text{(Addition)} \quad (1.4) \\
  (x, y) \mapsto xy & \quad \text{(Multiplication)} \quad (1.5)
\end{align*}

from $\mathbb{F} \times \mathbb{F} \to \mathbb{F}$, satisfying

F1) $x + y = y + x, \forall x, y \in \mathbb{F}$,
F2) $x + (y + z) = (x + y) + z, \forall x, y, z \in \mathbb{F}$,
F3) There exists a unique element 0 $\in \mathbb{F}$ such that $x + 0 = x, \forall x \in \mathbb{F}$,
F4) For every $x \in \mathbb{F}$, there exists a unique $y \in \mathbb{F}$ such that $x + y = 0$,
F5) $xy = yx, \forall x, y \in \mathbb{F}$,
F6) $x(yz) = (xy)z, \forall x, y, z \in \mathbb{F}$,
F7) There exists a unique element 1 $\in \mathbb{F}(1 \neq 0)$ such that $x1 = x, \forall x \in \mathbb{F}$,
F8) For every $x \neq 0$ in $\mathbb{F}$, there exists a unique $y \in \mathbb{F}$ such that $xy = 1$,
F9) $x(y + z) = xy + xz, \forall x, y, z \in \mathbb{F}$.

For example, the rational numbers, real numbers and complex numbers, respectively denoted by $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, are all fields. These fields satisfy the correspondence $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

Example 1.1 (Integers modulo $n$). Suppose two integers $a, b \in \mathbb{Z}$ and a positive integer $n \geq 1$ satisfy $a = nq + b$, that is, $q$ is the quotient and $b$ is the remainder in dividing $a$ by $n$. This can be equivalently expressed as $n \mid (a - b)$ (read “$n$ divides $a - b$”). Given $a, b \in \mathbb{Z}$, we say $a$ is congruent to $b$ modulo $n$, and write $a \equiv b \mod n$, if $n \mid (a - b)$. First, note that $\equiv$ is an equivalence relation, which forms the basis for modular arithmetic. In such a case, $n$ is referred to as the modulus.
The set of all equivalence classes of the integers for a modulus \( n \) is called the integers modulo \( n \) and denoted by \( \mathbb{Z}/n\mathbb{Z} = \{[0],[1],\ldots,[n-1]\} \).

Note that the notation \( \mathbb{Z}_n \) is also used for the integers modulo \( n \). For example, \( \mathbb{Z}/2\mathbb{Z} = \{0,1\} \), since any even integer is equivalent to 0 (mod 2), while any odd number is equivalent to 1 (mod 2).

**Fact**: If \( p \) is prime, \( \mathbb{Z}/p\mathbb{Z} \) is a field.

A **group** is a nonempty set \( G \) together with a binary operation (typically denoted \( \star, \cdot \), or +) that satisfies:

- **G1)** \((a \star b) \star c = a \star (b \star c)\),
- **G2)** There exists an element \( e \in G \) such that \( a \star e = a = e \star a \), \( \forall a \in G \),
- **G3)** For each \( a \in G \), there exists an element \( a^{-1} \in G \) such that \( a \star a^{-1} = a^{-1} \star a = e \).

A group is called **abelian** (or commutative) if \( a \star b = b \star a \) for all \( a,b \in G \). If \( G \) is abelian, it is common to write \( \frac{1}{b} \) for \( b^{-1} \) and \( \frac{a}{b} \) for \( ab^{-1} \). Note also that if the group operation is addition (+), then it is common to write 0 for \( e \), \(-a \) for \( a^{-1} \), and \( a - b \) for \( ab^{-1} \).

**Examples**:

1. \((\mathbb{R},+)\) - real numbers with ordinary addition
2. \((\mathbb{C},+)\) - complex numbers with ordinary addition
3. \((\mathbb{Z},+)\) - integers with ordinary addition
4. \((\mathbb{Z},+,+)\) - cyclic group of order \( n \) (addition modulo \( n \))

A subset \( H \) of a group \( G \) is called a **subgroup** of \( G \) if

- **SG1)** \( 1 \in H \),
- **SG2)** \( a,b \in H \Rightarrow ab \in H \),
- **SG3)** \( a \in H \Rightarrow a^{-1} \in H \).

**Example 1.2** (General Linear Groups). The set

\[
\text{GL}_n(\mathbb{R}) = \{A \in M_n(\mathbb{R})| \text{det}(A) \neq 0\}
\]

of (real) \( n \times n \) invertible matrices, along with ordinary matrix multiplication, is a group.

The set

\[
\text{GL}_n(\mathbb{C}) = \{A \in M_n(\mathbb{C})| \text{det}(A) \neq 0\}
\]

of (complex) \( n \times n \) invertible matrices, along with ordinary matrix multiplication, is a group.

These are the **general linear groups** of degree \( n \).
Given a group $G$ and a subset $H$, we may define an equivalence relation as follows:

$$a \equiv b \iff a^{-1}b \in H,$$

or equivalently,

$$a \equiv b \iff b \in aH.$$

Given an element $a \in G$, the equivalence class of $a$ under this equivalence relation is called a left coset of $H$ in $G$, denoted by $aH$, since

$$[a] = \{b \in G | a \equiv b\} = \{b \in G | a^{-1}b \in H\} = \{b \in G | a^{-1}b = c, c \in H\} = \{b \in G | b = ac, c \in H\} = aH.$$

In additive notation, we denote left cosets as $a + H$, since the equivalence relation is expressed as $(-a) + b \in H$.

Likewise, we may define a second equivalence relation

$$a \equiv b \iff ba^{-1} \in H \iff b \in Ha.$$

The equivalence class of $a$ under this equivalence relation is called a right coset of $H$ in $G$, denoted by $Ha$ or $H + a$ in additive notation. In either case, the element $a$ is called a coset representative for $a + H$ or $H + a$.

The following lemma will be useful for our discussion on rings.

**Lemma 1.3.** Let $H$ be a subgroup of a group $G$ and let $a, b \in G$.

i) $aH = bH$ if and only if $b^{-1}a \in H$. In particular, $aH = H$ if and only if $a \in H$.

ii) If $aH \cap bH \neq \emptyset$, then $aH = bH$.

iii) $|aH| = |H|$ for all $a \in G$.

A ring is a nonempty set $R$, together with two binary operations (addition and multiplication), such that:

R1) $R$ is a commutative group with respect to addition,

R2) Multiplication is associative,

R3) Multiplication is distributive with respect to addition.
More concretely, the conditions that $R$ be a ring are:

1) $r + (s + t) = (r + s) + t, \, \forall r, s, t \in R,$
2) $r + s = s + r, \, \forall r, s \in R,$
3) There exists $0 \in R$ such that $r + 0 = 0 + r = r, \, \forall r \in R,$
4) For every $r \in R$ there exists $-r \in R$ such that $r + (-r) = (-r) + r = 0,$
5) $(rs)t = r(st), \, \forall r, s, t \in R,$
6) $(r + s)t = rt + st, \, \forall r, s, t \in R,$
7) $r(s + t) = rs + rt, \, \forall r, s, t \in R.$

A ring is called commutative if $rs = sr$ for all $r, s \in R.$ If there exists an element $e \in R$ such that $re = er = r$ for every $r \in R,$ we call $R$ a ring with identity. The multiplicative identity, $e,$ is most often denoted as 1 (sometimes 1, 1$_R$ or 1$_R$). Note that some authors define a ring as always having a multiplicative identity. In this case, a set satisfying the ring axioms, except the existence of an identity, is referred to as a pseudo-ring. Given the definition of a ring, we may give a more concise definition of a field. A field is a ring that does not only consist of 0 for which every nonzero element has a multiplicative inverse.

**Example 1.4.** The set $\mathbb{Z}_n = \{0, 1, \ldots, n - 1\}$ is a commutative ring under addition and multiplication modulo $n,$ 1 being the identity.

**Example 1.5.** The set $M_n(F)$ is a noncommutative ring under matrix addition and multiplication, where the identity is denoted by $I$ or $I_n.$

**Example 1.6.** Denote the set of all polynomials in a single variable $x,$ with coefficients in the field $F,$ by $F[x].$ With polynomial addition and multiplication, $F[x]$ is a commutative ring. What is the identity?

We denote the set of polynomials in two (or more) variables by $F[x, y]$ or $F[x_1, x_2, \ldots, x_n].$ These are also commutative rings with identity.

Let $R$ be a ring with identity. The smallest positive integer $c$ for which $c \cdot 1 = 0$ is called the characteristic of $R,$ denoted char $R.$ If no such $c$ exists, we say $R$ has characteristic 0. Here, $c \cdot 1$ should be understand as $1 + 1 + \cdots + 1.$

**Fact:** Any finite ring has nonzero characteristic. Any finite field has prime characteristic.

**Example 1.7.**

1) char $\mathbb{R} = 0$ since $a \cdot 1 = 0 \Rightarrow a = 0$ for $a \in \mathbb{R}.$ Likewise, char $\mathbb{Z} = 0.$

2) In the ring $\mathbb{Z}_n,$ we have $n \cdot 1 = 0$ and $k \cdot 1 \neq 0$ for $k = 0, 1, \ldots, n - 1.$ Thus, char $\mathbb{Z}_n = n.$ Recall that $\mathbb{Z}_n$ is a field if and only if $n$ is prime.
Example 1.8 (Fields of characteristic 2). Suppose $\mathbb{F}$ is a field with char $\mathbb{F} = 2$. So, $2 \cdot 1 = 0$ and, furthermore, $2x = 0$ for every $x \in \mathbb{F}$. Equivalently, $x + x = 0 \Rightarrow x = -x$, for every $x \in \mathbb{F}$.

A subring of a ring $R$ is a subset $S$ of $R$ that is a ring using the operations of $R$ and having the same multiplicative identity. Note that if $1_R = 0_R$ then $R$ is called the zero ring.

Theorem 1.9. A nonempty subset $S$ of a ring $R$ is a subring if and only if

1) The multiplicative identity $1_R$ of $R$ is in $S$

2) $S$ is closed under subtraction:
   
   $a, b \in S \Rightarrow a - b \in S$

3) $S$ is closed under multiplication:
   
   $a, b \in S \Rightarrow ab \in S$

Let $R$ be a ring. A nonempty subset $\mathcal{I}$ of $R$ is called an ideal if

1) $a, b \in \mathcal{I} \Rightarrow a - b \in \mathcal{I}$ (closed under subtraction),

2) $a \in \mathcal{I}, r \in R \Rightarrow ar \in \mathcal{I}$ and $ra \in \mathcal{I}$.

Fact: If an ideal $\mathcal{I}$ contains the unit element 1, then $\mathcal{I} = R$.

Example 1.10 (Polynomial ideals).

1) Let $d$ be a polynomial over $\mathbb{F}$ and set $\mathcal{I} = d\mathbb{F}[x]$. That is, $\mathcal{I}$ is the set of all multiples of $d$ by arbitrary $f \in \mathbb{F}[x]$. The set $\mathcal{I}$ is an ideal called the principal ideal generated by $d$.

2) Let $\mathbb{F}$ be a subfield of the complex numbers and define
   
   $\mathcal{I} = (x + 2)\mathbb{F}[x] + (x^2 + 8x + 16)\mathbb{F}[x]$, which is an ideal. Hence,
   
   $(x^2 + 8x + 16) - x(x + 2) = 6x + 16 \in \mathcal{I}$
   
   and
   
   $6x + 16 - 6(x + 2) = 4 \in \mathcal{I}$.
   
   This shows that $1 \in \mathcal{I}$, so $\mathcal{I} = \mathbb{F}[x]$.

3) Let $S$ be a subset of a ring with identity $R$. Consider the set
   
   $\langle S \rangle = \{ \sum_{i=1}^{n} r_is_i \mid r_i \in R, s_i \in S, n \geq 1 \}$

   of all finite linear combinations of elements of $S$, with coefficients in $R$. $S$ is an ideal, called the ideal generated by $S$. It is the smallest ideal of $R$ that contains $S$.  


For rings, we may define cosets under the group structure (with respect to the addition operation). Since addition is commutative, we do not need to distinguish between left and right cosets, rather we simply have cosets. Given a subset $S$ of a commutative ring with identity $R$, we define an equivalence relation

$$a \equiv b \iff a - b \in S.$$ 

If $a \equiv b$ we say that $a$ and $b$ are congruent modulo $S$ (congruent mod $S$) and we write $a \equiv b \mod S$. The cosets of $S$ in $R$ are the sets $a + S = \{a + s | s \in S\}$.

We denote the set of all cosets of $S$ in $R$ by

$$R/S = \{a + S | a \in R\},$$

which is read as “$R$ mod $S$”. We define the operation of coset addition by

$$(a + S) + (b + S) = (a + b) + S.$$ 

If $S$ is a subgroup of the abelian group $R$, then $R/S$ is an abelian group under coset addition. Similarly, we define the operation of coset multiplication by

$$(a + S)(b + S) = ab + S.$$ 

For coset multiplication to be well-defined, we require

$$b + S = b' + S \implies ab + S = ab' + S,$$

or equivalently, by Lemma 1.3(i),

$$b - b' \in S \implies a(b - b') \in S.$$ 

Read carefully, the requirement is

$$a \in R, b \in S \implies ab \in S,$$

that is, $S$ must be an ideal for coset multiplication to be well-defined. On the other hand, if $S$ is an ideal then coset multiplication is well-defined.

**Theorem 1.11.** Let $R$ be a commutative ring with identity. The quotient $R/I$ is a ring under coset addition and multiplication if and only if $I$ is an ideal. In this case, $R/I$ is called the quotient ring of $R$ modulo $I$.

An ideal $I$ in a ring $R$ is a maximal ideal if $I \neq R$ and if whenever $J$ is an ideal satisfying $I \subseteq J \subseteq R$ then either $J = I$ or $J = R$. 
**Theorem 1.12.** \( \mathcal{I} \) is a maximal ideal if and only if \( R/\mathcal{I} \) is a field.

**Proof.**

1) Suppose \( R/\mathcal{I} \) is a field and \( \mathcal{J} \) is an ideal of \( R \) such that \( \mathcal{I} \subset \mathcal{J} \subset R \). Let \( x \in \mathcal{J} \) but \( x \not\in \mathcal{I} \), so \( x + \mathcal{I} \) is a nonzero element of \( R/\mathcal{I} \). By assumption, there is a \( y \in R \) such that \( (x + \mathcal{I})(y + \mathcal{I}) = 1 + \mathcal{I} \). Thus, \( xy + \mathcal{I} = 1 + \mathcal{I} \), which implies that \( 1 - xy \in \mathcal{I} \subset \mathcal{J} \). Since \( \mathcal{J} \) is an ideal, \( xy \in \mathcal{J} \). Finally, \( 1 = (1 - xy) + xy \in \mathcal{J} \), since \( 1 - xy \in \mathcal{J} \) and \( xy \in \mathcal{J} \), so \( \mathcal{J} = R \).

2) Now, assume \( \mathcal{I} \) is a maximal ideal and let \( x \in R, x \not\in \mathcal{I} \). We want to show that \( x \) has a multiplicative inverse. Define

\[
\mathcal{J} = \{xr + a | r \in R, a \in \mathcal{I}\}.
\]

**Claim:** \( \mathcal{J} \) is an ideal and \( \mathcal{I} \subset \mathcal{J} \) (proper). First, let \( xr + a \in \mathcal{J} \) and \( y \in R \). Then,

\[
(xr + a) + y = xr + (a + y) \in \mathcal{J},
\]

since \( a + y \in \mathcal{I} \) (by the fact that \( \mathcal{I} \) is an ideal). Clearly, if \( a \in \mathcal{I} \) then \( a \in \mathcal{J} \), since \( 0 \in R \). By maximality, \( \mathcal{J} = R \), which implies that \( xr + a = 1 \) for some \( r \in R, a \in \mathcal{I} \). Thus,

\[
1 + \mathcal{I} = (xr + a) + \mathcal{I} = xr + \mathcal{I} = (x + \mathcal{I})(r + \mathcal{I}),
\]

since \( a \in \mathcal{I} \). That is, every element of \( R/\mathcal{I} \) has an inverse.

\[\square\]

Let \( R \) be a ring. A nonzero element \( r \in R \) is called a zero divisor if there exists a nonzero \( s \in R \) such that \( rs = 0 \). A commutative ring \( R \) with identity is called an integral domain if it contains no zero divisors.

**Fact:** A finite integral domain is a field.

### 1.2 Some Set Theory Results

A (binary) relation \( R \) on a set \( A \) is a subset of \( A \times A \). A relation \( R \) is usually denoted by \( ~ \) (or \( \leq, \subseteq \)) and defined by \( \alpha \sim \beta \iff (\alpha, \beta) \in R \subset A \times A \). If we wish to specify the relation associated to a specific set, we may specify a pair \( (A, \sim) \). A set \( A \) is said to be partially ordered if there is a binary relation \( \sim \) such that

1) \( \alpha \sim \alpha, \forall \alpha \in A \) (Reflexivity),
2) \( \alpha \sim \beta \text{ and } \beta \sim \alpha \implies \alpha = \beta \) (Antisymmetry),
3) \( \alpha \sim \beta \text{ and } \beta \sim \gamma \implies \alpha \sim \gamma \) (Transitivity).
Note that set inclusion \( \subseteq \) is a relation between sets and most likely the only one we will use.

A set \( B \subseteq A \), of a partially ordered set \((A, \sim)\), is called totally ordered if for any \( x, y \in B \) we have \( x \sim y \) or \( y \sim x \). A partially ordered set \((A, \sim)\) is called inductive if every totally ordered subset of \( A \) (chain in \( A \)) has an upper bound in \( A \). An element \( x \in B \subseteq A \) is called maximal if \( x \sim y \) implies \( x = y \).

**Lemma 1.13** (Zorn’s Lemma). If a partially ordered set \((A, \sim)\) is inductive, then there exists a maximal element of \( A \).

### 1.3 Vector Spaces

A nonempty set \( V \) with the following two operations on elements of \( V \)

- \((x, y) \mapsto x + y\) maps \( V \times V \to V \) (addition) \((1.6)\)
- \((\alpha, x) \mapsto \alpha x\) maps \( \mathbb{F} \times V \to V \) (scalar multiplication) \((1.7)\)

is called a vector space (sometimes linear space) over the field \( \mathbb{F} \) if the following conditions are satisfied for all \( c_1, c_2 \in \mathbb{F} \) and \( \alpha, \beta, \gamma \in V \)

- \( V1) \quad \alpha + \beta = \beta + \alpha \)
- \( V2) \quad (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma) \)
- \( V3) \quad \exists! \ 0 \in V \text{ such that } \alpha + 0 = \alpha \)
- \( V4) \quad \forall \alpha \ \exists! -\alpha \in V \text{ such that } \alpha + (-\alpha) = 0 \)
- \( V5) \quad 1 \alpha = \alpha \)
- \( V6) \quad c_1(c_2\alpha) = (c_1c_2)\alpha \)
- \( V7) \quad (c_1 + c_2)\alpha = c_1\alpha + c_2\alpha \)
- \( V8) \quad c_1(\alpha + \beta) = c_1\alpha + c_1\beta \).

In particular, if \( \mathbb{F} = \mathbb{R} \) we call \( V \) a real vector space, while if \( \mathbb{F} = \mathbb{C} \) we say \( V \) is a complex vector space. The elements \( \alpha \in V \) are called vectors.

**Examples**

You should first verify that \([0], \mathbb{R}, \mathbb{C}\) are each vector spaces, then verify that the following examples are vector spaces.

1. \( \mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R} \text{ for } i = 1, \ldots, n\} \)
2. \( \mathbb{C}^n = \{(z_1, z_2, \ldots, z_n) \mid z_j \in \mathbb{C}, j = 1, \ldots, n\} \)
3. \( \{(z_1, z_2, z_1 + z_2) \mid z_1, z_2 \in \mathbb{C}\} \)
4. \( \mathbb{F}[x]_n \) which is the set of polynomials in the variable \( x \) up to order \( n \) (less than or equal to \( n \)).
1.3.1 Function Spaces

Let $X$ be a nonempty set and $V$ a vector space. Consider the set of functions mapping $X$ to $V$,
$$F = \{f : X \rightarrow V \}.$$

$F$ is a vector space with operations
$$\begin{align*}
(f + g)(x) &= f(x) + g(x) \\
(\alpha f)(x) &= \alpha f(x).
\end{align*}$$

Note that the 0-vector in $F$ is the function identically defined to be zero: $\mu(x) \equiv 0$ for all $x \in X$. It is also worth noting that we can regard the vector spaces $\mathbb{R}^n, \mathbb{C}^n$ as function spaces by the following identifications:
$$\begin{align*}
\mathbb{R}^n &= \{f : (1, 2, \ldots, n) \rightarrow \mathbb{R}\} \\
\mathbb{C}^n &= \{g : (1, 2, \ldots, n) \rightarrow \mathbb{C}\}.
\end{align*}$$

1.3.2 Subspaces

An important concept when dealing with vector spaces is subspaces. Let $V$ be a vector space. We call a nonempty subset $M \subset V$ a subspace of $V$ if $M$ is a vector space with the operations of $V$ restricted to $M$.

**Theorem 1.14 (Subspace Test).** A nonempty subset $M$ of a vector space $V$ is a subspace if and only if for all $c_1, c_2 \in \mathbb{F}$ and $\alpha, \beta \in M$ the linear combination $c_1 \alpha + c_2 \beta \in M$. That is, $M$ is a nonempty subset of $V$ which is closed with respect to addition and scalar multiplication.

Note that a vector space is a subspace of itself, however, this is usually not of much use.

**Examples**

Let $\Omega$ be an open subset of $\mathbb{R}^n$. The following are subspaces of the space of all functions from $\Omega$ to $\mathbb{C}$:
$$\begin{align*}
\mathcal{C}(\Omega) &= \text{the space of all continuous functions from } \Omega \text{ to } \mathbb{C} \\
\mathcal{C}^k(\Omega) &= \text{the space of all complex-valued functions which are continuous and have continuous partial derivatives up to order } k \\
\mathcal{C}^\infty(\Omega) &= \text{the space of infinitely differentiable functions defined on } \Omega \\
\mathcal{P}(\Omega) &= \text{the space of all polynomials in } n\text{-variables } (x_1, x_2, \ldots, x_n)
\end{align*}$$

**Remarks:**

Given two subspaces $M, N$ of a vector space $V$: 
i) If $M \subset N$, then $M$ is a subspace of $N$

ii) $M \cap N$ will always be a subspace

iii) $M \cup N$ not necessarily a subspace

Exercise:
Show that
$$W = \{x = (x_1, x_2, 0)^T \mid x_1, x_2 \in \mathbb{R}\}$$
is a subspace of $\mathbb{R}^3$.

Exercise:
If $M, N$ are two subspaces of a vector space $V$, then $M \cup N$ is a subspace if and only if $M \subseteq N$ or $N \subseteq M$.

Theorem 1.15. If $V$ is a vector space over an infinite field $F$, then $V$ cannot be the union of a finite collection of proper subspaces.

If $M, N$ are subspaces of a vector space $V$, we define the sum
$$M + N = \{u + v \mid u \in M, v \in N\}.$$
Likewise, given a collection of subspaces $\{S_k\}_{k \in K}$, we define the space
$$\sum_{k \in K} S_k = \left\{s_1 + \cdots + s_n \mid s_j \in \bigcup_{k \in K} S_k\right\}$$
of finite sums of elements in the union. The sum of any collection of subspaces of $V$ is a subspace of $V$.

1.3.3 Linear Independence and Bases

Given a vector space $V$, denote the set of all subsets of $V$ by $\mathcal{P}(V)$ and the set of all subspaces of $V$ by $\mathcal{S}(V)$ (sometimes called the lattice of subspaces). Thus, $\mathcal{S}(V) \subseteq \mathcal{P}(V)$ and there is a natural (canonical) map $L : \mathcal{P}(V) \to \mathcal{S}(V)$ sending a subset $S \in \mathcal{P}(V)$ to its linear span $L(S) \in \mathcal{S}(V)$, defined by
$$L(S) = \left\{\sum_{i=1}^n c_i \alpha_i \mid c_i \in F, \alpha_i \in S, n \geq 0\right\}.$$ Note that $L(S)$ is the subspace of finite linear combinations of elements of $S$ and it is the smallest subspace of $V$ containing $S$. The map $L$ mapping a subset $S$ to its linear span is surjective, since any subspace is a subset. The following theorem contains other properties of the map $L$. 

10
Theorem 1.16. The map \( L : \mathcal{P}(V) \to \mathcal{S}(V) \) satisfies

i) if \( S_1 \subseteq S_2 \) then \( L(S_1) \subseteq L(S_2) \),

ii) if \( \alpha \in L(S) \), then there exists a finite subset \( S' \subseteq S \) such that \( \alpha \in L(S') \),

iii) \( S \subseteq L(S) \) for all \( S \in \mathcal{P}(V) \),

iv) for every \( S \in \mathcal{P}(V) \), \( L(L(S)) = L(S) \),

v) if \( \beta \in L(S \cup \{ \alpha \}) \) and \( \beta \not\in L(S) \), then \( \alpha \in L(S \cup \{ \beta \}) \), where \( \alpha, \beta \in V \) and \( S \in \mathcal{P}(V) \).

A subset \( S \) is \textit{linearly dependent} (over \( F \)) if there exists a finite subset \( \{ \alpha_1, \ldots, \alpha_n \} \subseteq S \) and nonzero scalars \( c_1, \ldots, c_n \in F \) such that
\[
c_1 \alpha_1 + \cdots + c_n \alpha_n = 0.
\]
Otherwise, we say \( S \) is \textit{linearly independent}. That is, if
\[
\sum_{i=1}^{n} c_i \alpha_i = 0 \Rightarrow c_i = 0 \text{ for each } i = 1, \ldots, n.
\]

Example 1.17. Set \( V = \mathbb{R} \), \( F_1 = \mathbb{R} \), \( F_2 = \mathbb{R} \). Note that \( V \) is a vector space over both \( F_1 \) and \( F_2 \). Let
\[
S = \{ \alpha_1 = 1, \alpha_2 = \sqrt{2} \},
\]
which is linearly independent over \( F_1 \). However,
\[
\sqrt{2} \alpha_1 + (-1) \alpha_2 = 0,
\]
so \( S \) is linearly dependent over \( F_2 \).

A \textit{basis} of a vector space \( V \) is a subset \( S \) that is linearly independent such that \( L(S) = V \). If \( L(S) = V \), we say the set \( S \) \textit{spans} \( V \). If \( S \) is a basis for \( V \), then any vector \( \alpha \in V \) can be written uniquely as a linear combination of elements from \( S \), that is, there exist unique \( c_1, \ldots, c_n \in F \) such that \( \alpha = \sum_{i=1}^{n} c_i \alpha_i \) where \( \alpha_i \in S \). Note that every vector space has a basis.

Example 1.18.

a) Consider the vector space \( \mathbb{R}^n \). Define the vectors \( \delta_i = (0, \ldots, 0, 1, 0 \ldots, 0)^t \) of all 0’s, with a 1 in the \( i \)th position. The standard basis for \( \mathbb{R}^n \) consists of the vectors
\[
\{ \delta_1, \delta_2, \ldots, \delta_n \}.
\]
More generally, this is the \textit{standard basis} for the vector space \( \mathbb{F}^n \) of \( n \)-tuples of elements from the field \( \mathbb{F} \). Note that the standard basis elements are sometimes denoted by \( e_i \) or \( \epsilon_i \).
b) Consider the vector space \( \mathbb{F}[x]_n \) of polynomials in the variable \( x \), up to degree \( n \), with coefficients in the field \( \mathbb{F} \). The set of monomials
\[
\{1, x, x^2, \ldots, x^n\}
\]
is a basis for \( \mathbb{F}[x]_n \).

c) Consider the vector space \( M_{m,n}(\mathbb{F}) \) of \( m \times n \) matrices with entries in the field \( \mathbb{F} \). Define the elementary matrices \( \delta_{ij} \) as having 1 in the \((i, j)\) position and 0 everywhere else. The standard basis for \( M_{m,n}(\mathbb{F}) \) is the set
\[
\{\delta_{ij} | i = 1, \ldots, m; j = 1, \ldots, n\}.
\]
As with the standard basis for \( \mathbb{F}^n \), these elementary matrices are often denoted by \( e_{ij} \) or \( \epsilon_{ij} \). Sometimes this is necessary to avoid confusion with Kronecker’s delta or the \( \delta \)-distribution.

Recall the following fact regarding solutions of homogeneous systems of linear equations.

**Theorem 1.19.** If \( A \in M_{m,n}(\mathbb{F}) \) with \( m < n \) then \( Ax = 0 \) has a nontrivial solution.

With this in mind, we can prove the following theorem for finite-dimensional vector spaces.

**Theorem 1.20.** Let \( V \) be a vector space which is spanned by a finite set of vectors \( \{\alpha_1, \ldots, \alpha_n\} \). Then any independent set of vectors in \( V \) is finite and contains no more than \( n \) elements.

**Proof.** Denote the set of vectors that span \( V \) by \( \mathcal{S} = \{\alpha_1, \ldots, \alpha_n\} \), so for any \( v \in V \) there exist scalars \( c_1, \ldots, c_n \in \mathbb{F} \) such that \( v = \sum_{i=1}^{n} c_i \alpha_i \). Suppose \( S \) is a subset of \( V \) that is spanned by a set of \( m \) vectors \( \beta = \{\beta_1, \ldots, \beta_m\} \), where \( m > n \). Since each \( \beta_j \in L(\mathcal{S}), j = 1, \ldots, m \), there exists scalars \( a_{ij} \in \mathbb{F} \) such that
\[
\beta_1 = a_{11} \alpha_1 + \ldots + a_{n1} \alpha_n
\]
\[
\vdots
\]
\[
\beta_j = a_{1j} \alpha_1 + \ldots + a_{nj} \alpha_n
\]
\[
\vdots
\]
\[
\beta_m = a_{1m} \alpha_1 + \ldots + a_{nm} \alpha_n.
\]
Now, suppose \(d_1 \beta_1 + \ldots + d_m \beta_m = 0\). We want to show that not all \(d_i \in \mathbb{F}\) equal 0. Since each \(\beta_j\) can be expressed as a linear combination of the \(\alpha_i \in \alpha\), we have

\[
\sum_{j=1}^{m} d_j \beta_j = \sum_{j=1}^{m} d_j \left( \sum_{i=1}^{n} a_{ij} \alpha_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ij} d_j) \alpha_i.
\]

This defines a homogeneous system of linear equations

\[
\sum_{j=1}^{m} a_{ij} d_j = 0, \quad 1 \leq i \leq n,
\]

for which there exist \(d_1, \ldots, d_m\) not all 0, since \(m > n\). Thus,

\[
d_1 \beta_1 + \ldots + d_m \beta_m = 0,
\]

which shows that \(\beta\) is linearly dependent.

For the finite case, this enables us to prove the following theorem.

**Theorem 1.21.** If \(B_1, B_2\) are two bases of a vector space \(V\), then \(|B_1| = |B_2|\).

**Proof.**

**Finite Case:** The simplest way to prove the finite case is to use the previous theorem twice. Suppose \(V\) is a vector space with a finite basis \(B_1 = \{\alpha_1, \ldots, \alpha_n\}\). We must show that any other basis must have \(n\) elements. Suppose \(B_2 = \{\beta_1, \ldots, \beta_m\}\) is another basis of \(V\). If \(m > n\), the previous theorem shows that \(B_2\) is linearly dependent, thus not a basis. But if \(m < n\), \(B_1\) is linearly dependent by the same argument. In either case, we have a contradiction, so we must have \(m = n\).

**Infinite Case:** Suppose \(V\) does not have a finite basis. Let \(B_1, B_2\) be two infinite bases of \(V\) and let \(a \in B_1\). Since \(B_2\) is a basis, there exists a finite subset \(B_a \subset B_2\) such that \(a \in L(B_a)\) and \(a \notin L(B')\) for any proper subset \(B' \subset B_a\). This yields a well-defined function \(\varphi : B_1 \to P(B_a)\) defined by \(\varphi(a) = B_a\). We may now use the following result from set theory.

**Fact:** Let \(A\) and \(B\) be sets and suppose \(|A| = \infty\). If for each \(a \in A\) we have some finite set \(B_a \subset B\), then

\[
|A| \geq \left| \bigcup_{a \in A} B_a \right|.
\]
Since for each $\alpha \in B_1$ there is a finite subset $B_\alpha \subset B_2$, we must have

$$|B_1| \geq \left| \bigcup_{\alpha \in B_1} B_\alpha \right|.$$ 

Furthermore, since $\alpha \in L(B_\alpha)$ for all $\alpha \in B_1$, $V = L\left( \bigcup_{\alpha \in B_1} B_\alpha \right)$, which shows that $\bigcup_{\alpha \in B_1} B_\alpha$ is a subset of $B_2$ that spans all of $V$. Thus, $\bigcup_{\alpha \in B_1} B_\alpha = B_2$, so $|B_1| \geq |B_2|$. Reversing the roles of $B_1$ and $B_2$ shows that $|B_2| \geq |B_1|$.

The following theorem shows that any subset that spans $V$ contains a basis.

**Theorem 1.22.** Let $V$ be a vector space and suppose $V = L(S)$, where $S$ is a subset of $V$. Then $S$ contains a basis; that is, there is a basis $B$ of $V$ such that $B \subseteq S$.

**Proof.** If $S$ does not contain a nonzero vector, then the results is trivial, since $V$ must be the zero vector space. Suppose $S$ contains a nonzero vector $\alpha$ and define

$$S = \{ A \subseteq S \mid A \text{ is linearly independent over } \mathbb{F} \}.$$ 

Since $\{\alpha\} \in S$, $S$ is nonempty. We may place a partial ordering on $S$ by set inclusion, $\subseteq$. If $A = \{A_k \mid k \in K\}$ is a totally ordered subset of $S$, then $\bigcup_{k \in K} A_k$ is an upper bound for $A$ in $S$, so $(S, \subseteq)$ is inductive. By Zorn’s lemma 1.13, $S$ must have a maximal element, which we denote $B$.

Claim: $B$ is a basis of $V$.

Suppose, for contradiction, that $B$ is not a basis of $V$; in particular, suppose that $L(B) \neq V$, so $S \not\subseteq L(B)$. In this case, there must be an $s \in S$ that is not a linear combination of elements of $B$ (i.e. $s \in S \setminus L(B)$), since otherwise $V = L(S) \subseteq L(L(B)) = L(B)$. Thus, $B \cup \{s\}$ is linearly independent over $\mathbb{F}$, that is, $B \cup \{s\} \in S$. Since $s \not\in L(B)$, $s \not\in B$, hence $B \cup \{s\}$ is strictly larger than $B$, which contradicts the maximality of $B$. Thus, $L(B) = V$. \hfill \Box

On the other hand, the following theorem shows that any linearly independent subset of $V$ can be extended to a basis of $V$.

**Theorem 1.23.** If $S$ is a linearly independent subset of $V$ over $\mathbb{F}$, there exists a basis $B$ of $V$ such that $S \subseteq B$. 

\[ 14 \]
Proof Sketch: Suppose \( L(S) \neq V \), since otherwise \( S \) is a basis of \( V \). So, there is a vector \( \beta \in V \setminus L(S) \) Again, we may appeal to Zorn’s lemma. Consider the collection

\[
\mathcal{A} = \{ A \in \mathcal{P}(V) \mid S \subseteq A, A \text{ is linearly independent} \}.
\]

Note that, since \( S \in \mathcal{A} \), \( \mathcal{A} \) is nonempty. As in the previous theorem, we place a partial ordering on \( \mathcal{A} \) by set inclusion and a similar argument shows that \((\mathcal{A}, \subseteq)\) is inductive. Hence, there is a maximal element \( B \), which by maximality, must be a basis of \( V \). \( \square \)

If for any positive integer \( n \) there exist \( n \) linearly independent vectors in \( V \), we say that \( V \) is infinite dimensional. If \( V \) is not infinite dimensional, then there is a positive integer \( n \) such that:

a) there is a set of \( n \) vectors that are linearly independent;

b) if \( m > n \), then any \( m \) elements are linearly dependent.

In this case, we call \( n \) the dimension of \( V \), which is the largest possible number of linearly independent vectors and the number of vectors in any basis of \( V \).

**Theorem 1.24.** Let \( V \) be a vector space over \( \mathbb{F} \).

a) if \( M \) is a subspace of \( V \), then \( \dim M \leq \dim V \),

b) if \( V \) is finite-dimensional and \( M \) is a subspace of \( V \) such that \( \dim M = \dim V \), then \( M = V \),

c) if \( M \) is a subspace of \( V \), then there exists a subspace \( M' \) of \( V \) such that \( M + M' = V \) and \( M \cap M' = \{0\} \),

d) if \( V \) is finite-dimensional and \( M_1, M_2 \) are subspaces of \( V \), then

\[
\dim(M_1 + M_2) + \dim(M_1 \cap M_2) = \dim M_1 + \dim M_2.
\]

Note, in particular, that b) does not necessarily hold if \( V \) is infinite-dimensional.

**part d) only.** Suppose \( B \) is a basis of \( V \). Since \( S, T \) are subspaces of \( V \), there are subsets \( B_S \subseteq B \) and \( B_T \subseteq B \) that are bases for \( S \) and \( T \), respectively.

If \( S \cap T = \emptyset \), then \( B_S \cap B_T = \emptyset \), in which case, \( B_S \cup B_T \) is a basis of \( S + T \). Hence, \( \dim(S + T) = \dim(S) + \dim(T) \).

Now, suppose \( S \cap T \neq \emptyset \), so \( B_S \cap B_T \neq \emptyset \). Let \( B_{ST} \) be a basis for \( S \cap T \). Note that \( S \cap T \) is a subspace of both \( S \) and \( T \) (and \( V \) as well). So, we may extend \( B_{ST} \) to a basis \( B_{ST} \cup B_S \) of \( S \), where \( B_{ST} \cap B_S = \emptyset \). Likewise, we may extend \( B_{ST} \) to a basis \( B_{ST} \cup B_T \) of \( T \), where
\[ B_{ST} \cap T = \emptyset. \] We claim that \( B_S \cup B_{ST} \cup B_T \) is a basis of \( S + T \). To show linear independence, suppose there are subsets \( B'_S \subseteq B_S, B'_{ST} \subseteq B_{ST}, B'_T \subseteq B_T \) such that

\[
(a_1 \alpha_1 + \cdots + a_n \alpha_n) + (b_1 \gamma_1 + \cdots + b_m \gamma_m) + (c_1 \beta_1 + \cdots + c_p \beta_p) = 0,
\]

where \( B'_S = \{\alpha_1, \ldots, \alpha_n\}, B'_{ST} = \{\gamma_1, \ldots, \gamma_m\}, B'_T = \{\beta_1, \ldots, \beta_p\} \). By (1.8), we have

\[
\sum_{k=1}^{p} c_k \beta_k \in S \cap T, \quad \text{so} \quad \sum_{k=1}^{p} c_k \beta_k = \sum_{j=1}^{m} x_j \gamma_j, \quad \text{for some} \quad x_j \in \mathbb{F}. \]

Since \( B'_{ST} \) is a basis of \( S \cap T \) we must have \( d_k = 0 \), for each \( k \) (also, \( x_j = 0 \) for each \( j \)). Then, we are left with

\[
\sum_{i=1}^{n} a_i \alpha_i + \sum_{j=1}^{m} b_j \gamma_j = 0. \]

Since \( B_S \cup B_{ST} \) is a basis of \( S \), we have \( a_i = 0, i = 1, \ldots, n \) and \( b_j = 0, j = 1, \ldots, m \). Thus, \( B_S \cup B_{ST} \cup B_T \) is a basis and

\[
\dim(S + T) + \dim(S \cap T) = (n + m + p) + m
\]

\[
\dim S + \dim T = (n + m) + (p + m).
\]

\[ \square \]

**Example 1.25.** Consider the infinite dimensional vector space \( \mathbb{F}[x] \) and the subspace

\[ M = \left\{ \sum c_i x^{2i} \mid c_i \in \mathbb{F} \right\} \]

of even polynomials. A basis of \( M \) is all even powers of \( x \) (of which there are infinitely many), so \( \dim M = \dim \mathbb{F}[x] \), but \( M \neq \mathbb{F}[x] \).

**Example 1.26.** Consider the vector space \( M_{n_1}(\mathbb{F}) \) of \( n \times n \) matrices over the field \( \mathbb{F} \). If \( \text{char}(\mathbb{F}) = 2 \), then it is possible to have matrices which are both symmetric and skew-symmetric. Recall from Example 1.8 that if \( \text{char}(\mathbb{F}) = 2 \), then \( x = -x \) for every \( x \in \mathbb{F} \). For example, consider \( M_2(\mathbb{F}) \), where \( \text{char}(\mathbb{F}) = 2 \), and let \( a, b, c, d \in \mathbb{F} \). Then, a matrix would be symmetric if

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.
\]

Additionally, the matrix is skew-symmetric if

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -c \\ -b & -d \end{pmatrix}.
\]

To further highlight this example, consider the field \( \mathbb{F} = \mathbb{Z}_2 \) and let

\[
A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.
\]

Then, \( A^t = A \) and

\[
-A^t = \begin{pmatrix} -1 & -1 \\ -1 & 0 \end{pmatrix} = A,
\]

since \(-1 \equiv 1 \pmod{2}\). Thus, \( A \) is both symmetric and skew-symmetric.
Suppose $V$ is a finite-dimensional vector space over $\mathbb{F}$ and $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ is a basis of $V$. There is a canonical map, called the coordinate map, $\phi_{\mathcal{B}} : V \to \mathbb{F}^n$ defined by

$$
\phi_{\mathcal{B}}(v) = [v]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}
$$

where $v = c_1 \alpha_1 + \cdots + c_n \alpha_n$. That is, the coordinate map sends a vector to the $n$-tuple of coefficients in $\mathbb{F}^n$.

1.3.4 Internal vs. External Direct Sum

Let $\{V_k\}_{k \in K}$ be a family of vector spaces over the (same) field $\mathbb{F}$. We define the external direct sum of this family as

$$
\bigoplus_{k \in K} V_k = \{f : K \to \bigcup_{k \in K} V_k \mid f(k) = 0 \text{ except for finitely many } k \in K\}.
$$

In other words, the function $f$ has finite support. Equivalently,

$$
\bigoplus_{k \in K} V_k = \{(v_1, \ldots, v_n) \mid v_i \in V_i, i = 1, \ldots, n\}.
$$

In either case, we make this into a vector space by defining pointwise addition and scalar multiplication.

On the other hand, we may define a vector space from a family of subspaces of a vector space $V$ over $\mathbb{F}$. Let $\{S_k\}_{k \in K}$ be a family of subspaces of $V$ over the field $\mathbb{F}$. We say $V$ is the internal direct sum of $\{S_k\}_{k \in K}$ if for any $v \in V$ there exist unique $v_k \in S_k$ such that $v = v_1 + \cdots + v_n$. In this case, we write

$$
V = \bigoplus_{k=1}^n S_k,
$$

where $S_k$ is referred to as a direct summand of $V$. If $V = S \oplus T$, then $T$ is called the complement of $S$ in $V$.

**Theorem 1.27.** For any subspace $S$ of a vector space $V$, there exists a subspace $T$ of $V$ such that $V = S \oplus T$. That is, any subspace of $V$ has a complement in $V$.

**Proof.** First, assume $V$ is finite-dimensional, with basis $\mathcal{A} = \{\alpha_1, \ldots, \alpha_n\}$. Suppose $\mathcal{B} = \{\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_m}\} \subseteq \mathcal{A}$ is a basis for $S$. Here, $\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$, with $m \leq n$. Define $\mathcal{B}' = \{\alpha_{i_{m+1}}, \ldots, \alpha_{i_n}\}$, where $\{i_{m+1}, \ldots, i_n\} = \{1, \ldots, n\}\setminus\{i_1, \ldots, i_m\}$. Let $T = L(\mathcal{B}')$. If $v \in V$ has expression $v = c_1 \alpha_1 + \cdots + c_n \alpha_n$ with respect to $\mathcal{A}$, then after reordering (if necessary)
\[ v = (c_1 \alpha_1 + \cdots + c_m \alpha_m) + (c_{i_{m+1}} \alpha_{i_{m+1}} + \cdots + c_n \alpha_i) = v_s + v_t, \text{ with } v_s \in S \text{ and } v_t \in T. \]

Clearly, \( S \cap T = \{0\} \) and \( T \) is a subspace, by definition.

There is nothing dramatically different about the infinite-dimensional case. If \( \alpha \) is an infinite basis of \( V \), then there exists \( B_\alpha \subseteq \alpha \) such that \( L(B_\alpha) = S \). Then, define \( B' = \alpha \setminus B_\alpha \) and \( T = L(B') \).

### 1.3.5 Linear Transformations

Given vector spaces \( V, W \) over a (common) field \( \mathbb{F} \), a map \( T : V \to W \) satisfying

\[ T(c_1 \alpha + c_2 \beta) = c_1 T\alpha + c_2 T\beta, \quad (L1) \]

for every \( c_1, c_2 \in \mathbb{F} \) and \( \alpha, \beta \in V \), is called a linear transformation or homomorphism. In this context, the terms linear transformation and homomorphism are used interchangeably.

#### Example 1.28.

1) The coordinate map \( \phi_B : V \to \mathbb{F}^n \), where \( V \) is finite-dimensional, is a homomorphism from \( V \) to \( n \)-tuples of coordinates.

2) In the vector space \( V = M_{m,n}(\mathbb{F}) \), left matrix multiplication by a matrix \( A \in M_{m}(\mathbb{F}) \),

\[ TB = AB, \quad B \in V, \]

is a homomorphism from \( V \) to \( V \).

3) Let \( V = C^k(I) \), where \( k \geq 2 \) and \( I \) is a subinterval of \( \mathbb{R} \). Ordinary differentiation \( f \mapsto f' \) is a linear transformation from \( C^k(I) \) to \( C^{k-1}(I) \).

4) Let \( A = [a_1, b_2] \times \cdots \times [a_m, b_n] \subset \mathbb{R}^n \) and consider the vector space \( V = \mathcal{R}(A) \) of Riemann-integrable functions. Integration defines a linear transformation \( Tf = \int_A f \).

We denote the set of all linear transformations from \( V \) to \( W \), over \( \mathbb{F} \), by \( \text{Hom}_\mathbb{F}(V, W) \) or, alternately, \( \mathcal{L}(V, W) \). If it is not necessary to specify the field, we often drop the subscript, denoting this set simply as \( \text{Hom}(V, W) \). Note that we may define the vector space \( W^V \), consisting of all maps from \( V \) to \( W \), with pointwise operations.

**Fact:** \( \text{Hom}(V, W) \) is a subspace of \( W^V \).

In the following, we define several special kinds of homomorphisms.

#### Definition 1.29.

1) A homomorphism \( T \in \text{Hom}(V, V) \) is called an endomorphism. The space of all endomorphisms is often simplified as \( \text{Hom}(V) \).
2) An injective (1-1) homomorphism is called a *monomorphism*.

3) A surjective (onto) homomorphism is called an *epimorphism*.

4) A bijective homomorphism is called an *isomorphism*.

5) A bijective endomorphism is called an *automorphism*.

Additionally, we define the following sets associated with a homomorphism $T$.

**Definition 1.30.**

1) The *kernel* of $T$ is defined by

$$\ker(T) = \{ v \in V \mid Tv = 0 \}.$$  

Alternately, the kernel of $T$ is called the *nullspace*, which is denoted by $N(T)$.

2) The *image* of $T$ is defined by

$$\im(T) = \{ w \in W \mid Tv = w, \ v \in V \}.$$  

Alternately, the image of $T$ is called the *range* (space), which is denoted by $R(T)$.

The following theorem is a good exercise.

**Theorem 1.31.** Let $T$ be a homomorphism from $V$ to $W$.

a) $T$ is a monomorphism if and only if $\ker(T) = \{0\}$.

b) $T$ is an epimorphism if and only if $\im(T) = W$.

If there is an isomorphism $T : V \rightarrow W$, then we say $V$ and $W$ are *isomorphic* and write $V \cong W$. By Exercises 10 and 11, $\ker(T)$ is a subspace of $V$, while $\im(T)$ is a subspace of $W$. The dimension of $\ker(T)$ is often called the *nullity* of $T$, denoted by $(T)$. The dimension of $\im(T)$ is called the *rank* of $T$, denoted by $\text{rank}(T)$.

**Theorem 1.32.** Let $V, W$ be vector spaces over $\mathbb{F}$ and let $T \in \text{Hom}(V, W)$.

a) If $T$ is an epimorphism, then $\dim V \geq \dim W$.

b) If $\dim V = \dim W < \infty$, then $T$ is an isomorphism if and only if $T$ is injective or surjective. That is, $T$ is injective if and only if $T$ is surjective.

c) $\dim(\im T) + \dim(\ker T) = \dim V$.

In other notation, that is

$$\text{rank } T + \text{null } T = \dim V.$$
Theorem 1.33. Let $V$ be a finite-dimensional vector space with basis $\alpha = \{\alpha_1, \ldots, \alpha_n\}$. Identify $W^n$ with functions from $\{1, \ldots, n\}$ to $W$. For every collection of vectors $\beta = (\beta_1, \ldots, \beta_n) \in W^n$ there exists a unique homomorphism $T \in \text{Hom}(V, W)$ such that $T\alpha_i = \beta_i$, for each $i = 1, \ldots, n$.

Proof. Recall that the coordinate map $\phi_\alpha : V \to \mathbb{F}^n$ is an isomorphism. Define a homomorphism from $\mathbb{F}^n$ to $W$ by

$$L_\beta(x_1, \ldots, x_n) = \sum_{i=1}^n x_i \beta_i.$$

It is left as an exercise to show that $L_\beta$ is a homomorphism. Now, consider the map $L_\beta \circ \phi_\alpha$. Since $\phi_\alpha(\alpha_i) = \delta_i$, we have

$$L_\beta(\delta_i) = \beta_i,$$

which holds for each $i = 1, \ldots, n$. Now, suppose there is another $S \in \text{Hom}(V, W)$ such that $S\alpha_i = \beta_i$, for each $i$. Let $v \in V$ be an arbitrary vector with expression $v = a_1 \alpha_1 + \cdots + a_n \alpha_n$. Then

$$Tv = T\left(\sum_{i=1}^n a_i\alpha_i\right) = \sum_{i=1}^n a_i T\alpha_i$$

$$= \sum_{i=1}^n a_i \beta_i$$

$$= \sum_{i=1}^n a_i S\alpha_i$$

$$= S\left(\sum_{i=1}^n a_i \alpha_i\right)$$

$$= Sv.$$

Since $Tv = Sv$ for any $v \in V$, we have $T = S$. \qed

Corollary 1.34. Let $V$ be as in the previous theorem. Every basis $\alpha$ of $V$ determines an isomorphism $\Psi(\alpha) : \text{Hom}(V, W) \to W^n$.

Proof. Define $\Psi(\alpha) : \text{Hom}(V, W) \to W^n$ by $\Psi(\alpha) T = (T\alpha_1, \ldots, T\alpha_n)^t$ and inverse $\xi : W^n \to \text{Hom}(V, W)$ by $\xi((\beta_1, \ldots, \beta_n)^t) = \phi_\alpha(L_\beta(\cdot))^t$. \qed

Let $V$ and $W$ be finite-dimensional vector spaces over $\mathbb{F}$, with $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ a basis for $V$ and $\beta = \{\beta_1, \ldots, \beta_m\}$ a basis for $W$. Consider any homomorphism $T \in \text{Hom}(V, W)$. If
$T \alpha_i \neq 0$, then $\alpha_i \in \text{im}(T)$. Thus, there are $\{a_{ij}\} \in \mathbb{F}$ such that $T \alpha_i = \sum_{j=1}^{m} a_{ij} \beta_j$ for each $i = 1, \ldots, n$. Now, recall the coordinate map $\phi_{\beta} : W \rightarrow \mathbb{F}^m$. By the definition, we have

$$\phi_{\beta}(T \alpha_i) = \phi_{\beta} \left( \sum_{j=1}^{m} a_{ij} \beta_j \right) = (a_{i1}, a_{i2}, \ldots, a_{im})^t \in \mathbb{F}^m.$$ 

Then, we define the homomorphism $\Gamma(\alpha, \beta) : \text{Hom}(V, W) \rightarrow M_{m,n}(\mathbb{F})$ by

$$\Gamma(\alpha, \beta)T = [\phi_{\beta} \circ T \alpha_1, \phi_{\beta} \circ T \alpha_2, \ldots, \phi_{\beta} \circ T \alpha_n].$$

(1.9)

This construction yields the matrix $\Gamma(\alpha, \beta) \in M_{m,n}(\mathbb{F})$ called the matrix of $T$ relative to the bases $\alpha$ and $\beta$ or, simply, matrix representation of $T$. This suggests the following; if $V$ is $n$-dimensional and $W$ is $m$-dimensional, then $\dim(\text{Hom}(V, W))$ is $nm$-dimensional.

Given a homomorphism $T \in \text{Hom}(V)$, we say a subspace $S$ of $V$ is $T$-invariant if $T(S) \subseteq S$. Furthermore, given a subset $\mathcal{L}$ of $\text{Hom}(V)$, a subspace $S$ of $V$ is called $\mathcal{L}$-invariant if $S$ is $T$-invariant for every $T \in \mathcal{L}$. The vector space $V$ is called $\mathcal{L}$-irreducible if the only $\mathcal{L}$-invariant subspaces of $V$ are $\{0\}$ and $V$.

### 1.4 Quotient Spaces

Let $S$ be a subspace of a vector space $V$. Recall that we may define an equivalence relation by

$$u \equiv v \iff u - v \in S,$$

in which case, we say $u$ is equivalent (congruent) to $v$ mod $S$. Furthermore, the equivalence classes are

$$[v] = v + S,$$

cosets of $S$ in $V$, where $v$ is the coset representative. Additionally, recall that we denote the set of all cosets of $S$ in $V$ by

$$V/S = \{v + S \mid v \in V\},$$

which is the quotient space of $V$ mod $S$. We make $V/S$ into a vector space by defining

$$a(v + S) = av + S$$

$$(u + S) + (v + S) = (u + v) + S,$$

for $a \in \mathbb{F}$ and $u, v \in V$. In this vector space, recall that the equivalence class of 0 is the subspace $S$. When looking at quotient spaces, an important mapping is the canonical projection (natural projection) $\pi_S : V \rightarrow V/S$ defined by

$$\pi_S(v) = v + S,$$

which sends each vector to its corresponding coset of $S$ in $V$.  

-21
Lemma 1.35. The canonical projection $\pi_S : V \to V/S$ is surjective and $\ker(\pi_S) = S$.

Proof. First, suppose $\pi_S(v) = 0$, so $v + S = S$, since the 0-coset is the space $S$. But, $v + S = S \iff v \in S$. If $w + S$ is a coset in $V/S$, then $\pi_S(w) = w + S$, so $\pi_S$ is surjective. \hfill $\Box$

By the rank-nullity theorem 1.32(c), we have the following result, since $\text{im}\, \pi_S = V/S$ and $\ker \pi_S = S$.

Corollary 1.36. If $V$ is finite-dimensional and $S$ a subspace of $V$, then

$$\dim V = \dim S + \dim V/S.$$ 

Here, $\dim V/S$ is called the codimension of $S$ in $V$.

Theorem 1.37 (First Isomorphism Theorem). Let $T \in \text{Hom}(V, W)$ and suppose $S$ is a subspace of $V$ such that $T(S) = 0$; that is, $S \subseteq \ker T$. There exists a unique $\xi \in \text{Hom}(V/S, W)$ such that the following diagram commutes:

$$
\begin{array}{ccc}
V & \xrightarrow{T} & W \\
\downarrow{\pi_S} & & \downarrow{\xi} \\
V/S & & 
\end{array}
$$

That is, $\xi \circ \pi_S = T$.

Proof. By the condition we are trying to satisfy, we define $\xi \in \text{Hom}(V/S, W)$ by

$$\xi(v + S) = Tv, \quad v \in V$$

and show that $\xi$ is, indeed, a linear transformation, followed by a uniqueness argument. We must also show that this gives a well-defined mapping. To show this, suppose $u \in v + S$. We must show that $Tu = Tv$. Equivalently, we suppose $u - v \in S$, in which case, $T(u - v) = 0$, since $S \subseteq \ker T$. By linearity, it follows that $Tu = Tv$. This shows that $\xi$ is well-defined, in the sense that this mapping only depends on the coset, rather than a particular coset representative.

Now, let $a, b \in \mathbb{F}, u, v \in V$. By definition, we have

$$\xi(a[u] + b[v]) = \xi([au + bv]) = T(au + bv) = aT(u) + bT(v) = a\xi([u]) + b\xi([v]),$$

so $\xi$ is indeed a linear transformation. Also, for $v \in V$, we have

$$\xi(\pi_S(v)) = \xi([v]) = Tv$$

by definition, so the diagram commutes.

Lastly, suppose there is another $\tau \in \text{Hom}(V/S, W)$ such that $\tau \circ \pi_S = T$. So, $\tau([v]) = \xi([v])$ for any $[v] \in V/S$. Since $\pi_S$ is surjective, we have $\tau = \xi$ (equality must hold for all $v \in V$).

\textbf{Corollary 1.38.} For $T \in \text{Hom}(V, W)$ we have $\text{im} \, T \cong V/\ker T$.

Now, we consider the scenario in which $V = M \oplus N$, so $N$ is a complement of $M$ in $V$. In this case, any element $v \in V$ can be expressed as $v = v_m + v_n$, where $v_m \in M$ and $v_n \in N$.

As in Exercise 12, we may define a projection operator $P : V \to N$ by $P(v_m + v_n) = v_n$. The operator $P$ is called the \textit{projection onto} $N$ \textit{along} $M$. Clearly, we have

$$\text{im}(P) = N$$

and since

$$P(v_m + v_n) = 0 \Rightarrow v_n = 0,$$

we also have $\ker(P) = M$. If we apply the First Isomorphism Theorem, in particular Corollary 1.38, we have

$$\text{im}(P) \cong V/\ker(P) \iff N \cong V/M.$$ 

This proves the following result.

\textbf{Corollary 1.39.} Any complement of the subspace $S$ in $V$ is isomorphic to $V/S$.

\textbf{Theorem 1.40 (Second Isomorphism Theorem).} Suppose $M, N$ are subspaces of a vector space $V$ such that $M \subseteq N$. Then

$$(V/M)/(N/M) \cong V/N.$$ 

\textit{Proof.} First, note that $N/M$ is a subspace of $V/M$. Define a homomorphism $T \in \text{Hom}(V/M, V/N)$ by $T(v + M) = v + N$. Since $M \subseteq N$, if $u + M = v + M$, then $u - v \in M \subseteq N$, hence $u + N = v + N$, showing that $T$ is well-defined. The 0-coset in $V/N$ is $N$, thus,

$$T(u + M) = N \iff u \in N \iff u + M \in N/M,$$

which shows that $\ker T = N/M$. Also, $T$ is surjective, $\text{im} \, T = V/N$. By Corollary 1.38,

$$\text{im} \, T \cong V/\ker T,$$

which, in this case, reads

$$V/N \cong (V/M)/(N/M).$$
The next isomorphism theorem is a generalization of Corollary 1.39. Note that, in some of the literature, the second and third isomorphism theorems are in reverse order.

**Theorem 1.41** (Third Isomorphism Theorem). Let $M,N$ be subspaces of a vector space $V$. Then

$$(M + N)/M \cong N/(M \cap N).$$

**Proof.** Define $T: M + N \rightarrow N/M \cap N$ by $T(m + n) = n + (M \cap N)$. Since $\ker T = M$, Corollary 1.38 gives

$$(M + N)/M \cong N/(M \cap N).$$

Likewise, a similar map may be defined to show

$$(M + N)/N \cong M/(M \cap N).$$

A vector space $A$ over $\mathbb{F}$ with a third operation, vector multiplication, from $A \times A \rightarrow A$, satisfying

A1) $(ab)c = a(bc)$, for all $a,b,c \in A$

A2) $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$, for all $a,b,c \in A$

A3) $x(ab) = (xa)b = a(xb)$, for all $x \in \mathbb{F}, a,b \in A$

is called an associative algebra (or linear algebra). If there is a multiplicative identity, $1 \in \mathbb{F}$ such that $a1 = 1a = a$ for every $a \in A$, then $A$ is an (associative) algebra with identity. An algebra is commutative if $ab = ba$ for every $a,b \in A$.

If $A,B$ are algebras, then a homomorphism $\varphi \in \text{Hom}(A,B)$ is called an algebra homomorphism if it additionally satisfies $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a,b \in A$. That is, an algebra homomorphism preserves multiplication.

**Exercises - Section 1.**

1. (a) Define $S = \{A \in M_n(\mathbb{F}) \mid A^t = A\}$ the subset of all symmetric matrices in $M_n(\mathbb{F})$. Show that $S$ is a subspace of $M_n(\mathbb{F})$.

   (b) Define $T = \{A \in M_n(\mathbb{F}) \mid A^t = -A\}$ the subset of all skew-symmetric matrices in $M_n(\mathbb{F})$. Show that $T$ is a subspace of $M_n(\mathbb{F})$.

2. Let $S$ be a subspace of a vector space $V$ and let $\alpha, \beta \in V$. Set $A = \alpha + S$ and $B = \beta + S$. Show that $A = B$ or $A \cap B = \emptyset$.

3. If $S_1,S_2$ are subsets of a vector space $V$, show that $L(S_1 \cup S_2) = L(S_1) + L(S_2)$.

4. Determine two subspaces $S_1,S_2$ of $\mathbb{R}^2$ such that $S_1 \cup S_2$ is not a subspace of $\mathbb{R}^2$. 

24
5. (Dedekind’s Modular Law) Let $R, S, T$ be subspaces of a vector space $V$ with $S \subseteq R$. Show $R \cap (S + T) = S + (R \cap T)$.

6. Let $V$ be a finite-dimensional vector space with basis $B$. Show that the coordinate map $\phi_B$ is bijective. Show $\phi_B$ preserves addition and scalar multiplication, that is

$$\phi_B(av + bw) = a\phi_B(v) + b\phi_B(w),$$

where $a, b \in \mathbb{F}$ and $v, w \in V$. This shows that $\phi_B$ is an isomorphism from $V$ onto $\mathbb{F}^n$, where $\dim V = n$.

7. Let $V$ be the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$ and define

$$V_e = \{f \in V \mid f(-x) = f(x)\}$$
$$V_o = \{f \in V \mid f(-x) = -f(x)\}$$

the subsets of even and odd functions, respectively.

(a) Show $V_e$ and $V_o$ are subspaces of $V$
(b) Prove $V_e + V_o = V$
(c) Prove $V_e \cap V_o = \{0\}$

This shows that $V = V_e \oplus V_o$.

8. Let $S_1, S_2$ be subspaces of a vector space $V$ such that $S_1 + S_2 = V$ and $S_1 \cap S_2 = \{0\}$. Prove that for each $\alpha \in V$ there are unique vectors $\alpha_1 \in S_1, \alpha_2 \in S_2$ such that $\alpha = \alpha_1 + \alpha_2$.

9. Define $V = \mathbb{R}$ as a vector space over the field $\mathbb{F} = \mathbb{Q}$. Show that this vector space is infinite-dimensional.

10. For a homomorphism $T \in \text{Hom}(V, W)$, show that $\ker(T)$ is a subspace of $V$.

11. For a homomorphism $T \in \text{Hom}(V, W)$, show that $\text{im}(T)$ is a subspace of $W$.

12. Suppose $V = M \oplus N$ and define linear operators $P_M, P_N \in \text{Hom}(V)$ by

$$P_M(v_m + v_n) = v_m$$
$$P_N(v_m + v_n) = v_n,$$

where $v_m \in M, v_n \in N$. The operators $P_M : V \rightarrow M$ and $P_N : V \rightarrow N$ are called projection operators. Show that

(a) $P_M^2 = P_M$ and $P_N^2 = P_N$.
(b) $P_M + P_N = I_V$, where $I_V$ is the identity on $V$. 

25
(c) $P_M P_N = 0 = P_N P_M$
(d) $V = \text{im}(P_M) \oplus \text{im}(P_N)$

13. Let $T \in \text{Hom}(V)$. Show that $T^2 = 0$ if and only if there exists two subspaces $M, N$ of $V$ such that
   (a) $M + N = V$
   (b) $M \cap N = \{0\}$
   (c) $T(N) = \{0\}$
   (d) $T(M) \subseteq N$.

14. An operator $T \in \text{Hom}(V)$ is called an involution if $T^2 = I_V$. If $T$ is an involution, show there exist subspaces $M, N$ of $V$ such that
   (a) $M + N = V$
   (b) $M \cap N = \{0\}$
   (c) $T v = v$ for every $v \in M$
   (d) $T w = -w$ for every $w \in N$.

15. Let $T \in \text{Hom}(\mathbb{R}^3, \mathbb{R}^3)$ be defined by
   $$T \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 3u \\ u - v \\ 2u + v + w \end{pmatrix}.$$ 
   Determine if $T$ is a monomorphism. Why is the answer to this question enough to determine whether $T$ is an isomorphism (automorphism) or not?

16. Suppose $T \in \text{Hom}(\mathbb{F}^3, \mathbb{F}^3)$ is an endomorphism such that
   $$T \delta_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}; \quad T \delta_2 = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}; \quad T \delta_3 = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix};$$
   where $\{\delta_1, \delta_2, \delta_3\}$ is the standard basis of $\mathbb{F}^3$. Describe $T$ explicitly; that is, give a formula for
   $$T \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$
   as in Problem 15.
17. Consider the vector space \( V = C([-1, 1]) \) of continuous, real-valued functions defined on \([-1, 1]\). Let \( S_e = \{ f \in V \mid f(-x) = f(x) \} \) be the subspace of even functions and \( S_o = \{ f \in V \mid f(-x) = -f(x) \} \) the subspace of odd functions. Define the operator \( T \) as the indefinite integral

\[
(T f)(x) = \int_0^x f(s) \, ds.
\]

Are the subspaces \( S_e \) and \( S_o \) \( T \)-invariant?

18. Show that \( \mathcal{E}(V) = \text{Hom}(V, V) \) with composition of endomorphisms,

\[
(T_1, T_2) \mapsto T_1 \circ T_2,
\]

defines an associative algebra with identity. Is this algebra commutative?

19. Let \( M \) be a subspace of a vector space \( V \) and let \( \{ m_1, \ldots, m_n \} \) be a basis of \( M \). Explain how to determine a basis for \( V/M \).

## 2 Modules; Linear and Multilinear functionals

### 2.1 Linear functionals

In this section, we discuss a special class of homomorphisms defined on a vector space \( V \), namely those that map \( V \) to the field \( \mathbb{F} \).

A map \( f \in \text{Hom}(V, \mathbb{F}) \) is called a *linear functional*. That is, \( f \) is a map from \( V \) to \( \mathbb{F} \) such that

\[
f(av + bw) = af(v) + bf(w),
\]

for all \( a, b \in \mathbb{F} \) and \( v, w \in V \).

The set of all linear functionals defined on \( V \) is called the *dual space* of \( V \), denoted by \( V^* \). So, \( V^* \) is simply a shorthand notation for \( \text{Hom}(V, \mathbb{F}) \). We have already shown, in section 1.3.5, that \( \text{Hom}(V, W) \) is a vector space in general, so \( V^* \) is a vector space (with pointwise operations).

**Example 2.1.**

1) Consider the vector space \( C[0, 1] \) of all continuous, real-valued functions defined on \([0, 1]\). Definite integration defines a linear functional; for any \( f \in C[0, 1] \) define

\[
\varphi(f) = \int_0^1 f(x) \, dx.
\]

Since integration is linear, \( \varphi \) is an element of the dual space \( C[0, 1]^* \).
2) On any function space, \( C[0,1] \) or \( \mathbb{F}[x] \) for example, point evaluation defines a linear functional. For any \( p \in \mathbb{F}[x] \), define

\[
g(p) = p(0),
\]

point evaluation at 0. So, \( g \) is an element of the dual space \( \mathbb{F}[x]^* \).

**Remark 2.2.** Recall the matrix representation of a homomorphism \( T \in \text{Hom}(V, W) \), given by (1.9). If \( f \) is a linear functional on \( V \), we may determine its matrix representation, since linear functionals are special types of homomorphisms. Since linear functionals map to the field \( \mathbb{F} \), which has basis \( \beta = \{1\} \), the calculation of \( \Gamma(\alpha, \beta)f \) is greatly simplified. As an example, let \( V = \mathbb{F}[x]_2 \) and define

\[
f(p) = \int_0^2 p(x) \, dx.
\]

Taking the standard basis \( \alpha = \{\alpha_1, \alpha_2, \alpha_3\} \) of \( V \), where \( \alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2 \), we have

\[
\begin{align*}
\Gamma(\alpha, \beta)f &= \left( \phi_\beta \circ f(\alpha_1), \phi_\beta \circ f(\alpha_2), \phi_\beta \circ f(\alpha_3) \right) = (2, 2, \frac{8}{3}).
\end{align*}
\]

Thus, the matrix of \( f \) is given by

\[
\begin{pmatrix}
2 & 2 & \frac{8}{3}
\end{pmatrix}
\]

As a verification, let \( p(x) = ax^2 + bx + c \). Evaluating \( f(p) \) directly yields

\[
\int_0^2 ax^2 + bx + c \, dx = \frac{8}{3}a + 2b + 2c.
\]

We may also compute \( f(p) \) by multiplying the coefficients of \( p \) in \( \alpha \) by the matrix of \( f \), giving

\[
\begin{pmatrix}
2 & 2 & \frac{8}{3}
\end{pmatrix}
\begin{pmatrix}
c \\ b \\ a
\end{pmatrix} = 2c + 2b + \frac{8}{3}a.
\]

Similarly, if \( V = \mathbb{F}[x]_2 \) and \( f(p) = p(t) \), for some fixed \( t \in \mathbb{F} \), it is straightforward to verify that the matrix of \( f \) is given by

\[
\begin{pmatrix}
1 & t & t^2
\end{pmatrix}.
Our next few results are concerned with the structure of the dual space $V^*$ and its relation with the vector space $V$. First, note that if $v \in V$ is nonzero, then there exists a linear functional $f \in V^*$ such that $f(v) \neq 0$. In other words, if $f(v) = 0$ for every $f \in V^*$, then $v = 0$. Since $\dim \mathbb{F} = 1$, any linear functional is either identically 0 or surjective (Exercise). Furthermore, by the rank-nullity theorem, if $f \neq 0$ and $\dim V < \infty$ we have

$$\dim(\ker f) = \dim V - 1.$$  

In fact, the following theorem shows how the image and kernel of a nonzero functional relate to $V$.

**Theorem 2.3.** If $f \in V^*$ and $f(v) \neq 0$ for some $v \in V$, then $V = L(\{v\}) \oplus \ker(f)$.

**Proof.** We know that $V = \text{im}(f) \oplus \ker(f)$, since $\text{im}(f) \cong V/\ker(f)$, by the first isomorphism theorem. First, we may show that $L(\{v\}) \cap \ker(f) = \{0\}$. Suppose $u \in L(\{v\}) \cap \ker(f)$, so $f(u) = u$ and $u = av$ for some $a \in \mathbb{F}$. Thus, $f(av) = 0$, which implies that $af(v) = 0$, but $f(v) \neq 0$; hence, $a = 0$, so $u = 0$.

Now, let $u \in V$ be an arbitrary vector with $f(u) \neq 0$, say $f(u) = b \in \mathbb{F}$. We will show that $u \in L(\{v\}) + \ker(f)$. Since $f(v) \neq 0$, we may set $f(v) = c$ for some $c \in \mathbb{F}$. We have $f(b/c v) = b/c f(v) = b = f(u)$ and clearly $b/c v \in L(\{v\})$. Also,

$$f\left(u - \frac{b}{c} v\right) = f(u) - f(u) = 0,$$

so $u - \frac{b}{c} v \in \ker(f)$. Since

$$u = \frac{b}{c} v + (u - \frac{b}{c} v),$$

we have $u \in L(\{v\}) + \ker(f)$. Finally, $V = L(\{v\}) \oplus \ker(f)$. \(\square\)

Now, suppose $V$ has a basis $\alpha$. We may define a set of linear functionals $\alpha_i^* \in V^*$ by

$$\alpha_i^*(\alpha_j) = \delta_{i,j}.$$  

The set $\alpha^* = \{\alpha_i^* \mid i \in \Delta\}$ is linearly independent, as shown in the following theorem.

**Theorem 2.4.** If $V$ is finite-dimensional, then $\dim V = \dim V^*$. Furthermore, the set $\alpha^*$ is a basis of $V^*$.

**Proof.** It suffices to show that $\alpha^*$ is a basis of $V^*$, from which it follows that $\dim V = \dim V^*$. To show linear independence, suppose $\sum_{i=1}^n c_i \alpha_i^* = 0$ for some $c_i \in \mathbb{F}$.  

29
Thus, $\sum_{i=1}^{n} c_i \alpha_i^*(v) = 0$ for all $v \in V$. In particular, this must be true on each basis element $\alpha_i \in \mathfrak{a}$ of $V$. So,

$$\sum_{i=1}^{n} c_i \alpha_i^*(\alpha_j) = 0, \quad \text{for each } j = 1, \ldots, n,$$

which implies that $c_j = 0$, for each $j = 1, \ldots, n$ by the definition $\alpha_i^*(\alpha_j) = \delta_{i,j}$. Hence, the set $\mathfrak{a}^*$ is linearly independent.

Now, we must show that $\mathfrak{a}^*$ spans $V^*$, that is, $L(\mathfrak{a}^*) = V^*$, which follows from linearity.

Let $f \in V^*$ be an arbitrary nonzero linear functional. If $v \in V$ has expression $v = \sum_{j=1}^{n} a_j \alpha_j$, then

$$f(v) = f \left( \sum_{j=1}^{n} a_j \alpha_j \right) = \sum_{j=1}^{n} a_j f(\alpha_j),$$

which we will show is equivalent to the expression $\sum_{i=1}^{n} f(\alpha_i) \alpha_i^*(v)$. To wit,

$$\sum_{i=1}^{n} f(\alpha_i) \alpha_i^*(v) = \sum_{i=1}^{n} f(\alpha_i) \alpha_i^* \left( \sum_{j=1}^{n} a_j \alpha_j \right) = \sum_{i=1}^{n} f(\alpha_i) \sum_{j=1}^{n} a_j \alpha_i^*(\alpha_j) = \sum_{i=1}^{n} f(\alpha_i) a_i,$$

which again follows from the definition $\alpha_i^*(\alpha_j) = \delta_{i,j}$. In conclusion, any $f \in V^*$ can be expressed as $f = \sum_{i=1}^{n} f(\alpha_i) \alpha_i^*$, hence $L(\mathfrak{a}^*) = V^*$. \qed

The basis $\mathfrak{a}^*$ is called the dual basis of $\mathfrak{a}$.

**Example 2.5.** Let $V = \mathbb{R}[x]_2 = \{ ax^2 + bx + c \mid a, b, c \in \mathbb{R} \}$ be the vector space of real
polynomials of degree 2 or less. For any \( p \in V \), define three linear functionals by

\[
\begin{align*}
  f_1(p) &= \int_0^1 p(x) \, dx \\
  f_2(p) &= \int_0^2 p(x) \, dx \\
  f_3(p) &= \int_0^{-1} p(x) \, dx.
\end{align*}
\]

We will determine a basis of \( V \) for which \( \{ f_1, f_2, f_3 \} \) is the dual basis. Let

\[
\begin{align*}
  p_1(x) &= a_1 x^2 + b_1 x + c_1 \\
  p_2(x) &= a_2 x^2 + b_2 x + c_2 \\
  p_3(x) &= a_3 x^2 + b_3 x + c_3
\end{align*}
\]

be polynomials in \( V \). For \( f \) to be the dual basis of \( \{ p_1, p_2, p_3 \} \) we must have \( f_i(p_j) = \delta_{i,j} \) for \( 1 \leq i, j \leq 3 \). To determine \( p_1 \), the conditions

\[
\begin{align*}
  \int_0^1 (a_1 x^2 + b_1 x + c) \, dx &= 1 & \Rightarrow & \frac{1}{3} a_1 + \frac{1}{2} b_1 + c_1 &= 1 \\
  \int_0^2 (a_1 x^2 + b_1 x + c) \, dx &= 0 & \Rightarrow & \frac{8}{3} a_1 + 2 b_1 + 2 c_1 &= 0 \\
  \int_0^{-1} (a_1 x^2 + b_1 x + c) \, dx &= 0 & \Rightarrow & -\frac{1}{3} a_1 + \frac{1}{2} b_1 - c_1 &= 0
\end{align*}
\]

This yields a linear system of equations

\[
\begin{pmatrix}
  \frac{1}{3} & \frac{1}{2} & 1 \\
  \frac{8}{3} & 2 & 2 \\
  -\frac{1}{3} & \frac{1}{2} & -1
\end{pmatrix}
\begin{pmatrix}
  a_1 \\
  b_1 \\
  c_1
\end{pmatrix}
= 
\begin{pmatrix}
  1 \\
  0 \\
  0
\end{pmatrix}.
\]

Likewise, the conditions for \( p_2 \) and \( p_3 \) yield the linear systems

\[
\begin{pmatrix}
  \frac{1}{3} & \frac{1}{2} & 1 \\
  \frac{8}{3} & 2 & 2 \\
  -\frac{1}{3} & \frac{1}{2} & -1
\end{pmatrix}
\begin{pmatrix}
  a_2 \\
  b_2 \\
  c_2
\end{pmatrix}
= 
\begin{pmatrix}
  0 \\
  1 \\
  0
\end{pmatrix}.
\]
and
\[
\begin{pmatrix}
\frac{1}{3} & \frac{1}{2} & 1 \\
\frac{8}{3} & 2 & 2 \\
-\frac{1}{3} & \frac{1}{2} & -1
\end{pmatrix}
\begin{pmatrix}
a_3 \\
b_3 \\
c_3
\end{pmatrix}
= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.
\]

Thus, if we define
\[
G = \begin{pmatrix}
\frac{1}{3} & \frac{1}{2} & 1 \\
\frac{8}{3} & 2 & 2 \\
-\frac{1}{3} & \frac{1}{2} & -1
\end{pmatrix},
\]
then
\[
G^{-1} = \begin{pmatrix}
a_1 & a_2 & a_3 \\
b_1 & b_2 & b_3 \\
c_1 & c_2 & c_3
\end{pmatrix}.
\]

One can check that this yields
\[
p_1(x) = -\frac{3}{2}x^2 + x + 1,
\]
\[
p_2(x) = \frac{1}{2}x^2 - \frac{1}{6},
\]
\[
p_3(x) = -\frac{1}{2}x^2 + x - \frac{1}{2}.
\]

Given our previous remarks, the dual space \( V^* \) is, itself, a vector space, so we may define its dual space, \((V^*)^*\), which we typically denote by \( V^{**} \). The space \( V^{**} \) is called the double dual and it consists of linear functionals in \( \text{Hom}(V^*, \mathbb{F}) \). A vector space \( V \) is called reflexive if \( V \cong V^{**} \).

We define a map \( \varphi : V \times V^* \to \mathbb{F} \), called the dual pairing, by \( \varphi(v, f) = f(v) \), for all \( v \in V, f \in V^* \). The dual pairing simply evaluates the given linear functional at the given vector. This is an example of a bilinear functional, meaning that the functional is linear over \( V \) and over \( V^* \) (linear in both components). In general, a map \( \omega : U \times V \to W \) is called bilinear if
\[
\omega(au_1 + bu_2, v) = a\omega(u_1, v) + b\omega(u_2, v)
\]
\[
\omega(u, av_1 + bv_2) = a\omega(u, v_1) + b\omega(u, v_2),
\]
for all \( a, b \in \mathbb{F}, u, u_1, u_2 \in U \) and \( v, v_1, v_2 \in V \). You should verify that the dual pairing satisfies the above conditions. Sometimes, the dual pairing is referred to as the natural...
bilinear functional. The dual pairing \( \varphi \) yields a natural homomorphism \( \psi : V \to V^{**} \) defined by \( \psi(v) = \varphi(v, \cdot) \), where the right hand side takes any \( f \in V^* \) and evaluates \( f(v) \). Thus, \( \psi(v) \in V^{**} \) for each \( v \in V \). The linearity of \( \psi \) follows directly from the linearity of \( \varphi \).

**Theorem 2.6.** The map \( \psi : V \to V^{**} \) is injective. If \( \dim V < \infty \), then \( \psi \) is an isomorphism.

**Proof.** To show \( \psi \) is injective, suppose \( \psi(v) = 0 \). So, \( \varphi(v, \cdot) = 0 \), which means that \( f(v) = 0 \) for all \( f \in V^* \), which is true if and only if \( v = 0 \). Hence, \( \ker \psi = \{0\} \).

Now, suppose \( \dim V < \infty \), so \( \dim V^* < \infty \), from which it follows that \( \dim V = \dim V^{**} < \infty \). By Theorem 1.32b, \( \psi \) must also be surjective, hence \( \psi \) is an isomorphism.

**Corollary 2.7.** If \( \dim V < \infty \), \( V \) is reflexive.

### 2.1.1 Operator adjoints and perp spaces

Now, if \( S \) is a subspace of \( V \), we define the set
\[
S^\perp = \{ f \in V^* \mid \varphi(s, f) = 0, \forall s \in S \},
\]
which is read as \( S \) “perp.” Note that this is exactly the same as what some texts refer to as the annihilator of \( S \), denoted by \( S^0 \), since any linear functional in this set sends any element of \( S \) to 0. Analogously, if \( M \) is a subspace of \( V^* \), we define the set
\[
M^\perp = \{ v \in V \mid \varphi(v, f) = 0, \forall f \in M \}.
\]

Note that \( S^\perp \) is a subspace of \( V^* \), while \( M^\perp \) is a subspace of \( V \).

**Example 2.8.**

a) Let \( V = \mathbb{R}^4 \) and consider the following three linear functionals
\[
\begin{align*}
f_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_2 + 2x_3 + x_4 \\
f_2(x_1, x_2, x_3, x_4) &= 2x_2 + x_4 \\
f_3(x_1, x_2, x_3, x_4) &= -2x_1 - 4x_3 + 3x_4.
\end{align*}
\]

We can determine the subspace \( S \) of \( V \) that is perpendicular to \( \{f_1, f_2, f_3\} \); that is, the subspace that \( \{f_1, f_2, f_3\} \) annihilates. We must determine vectors \((x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \) such that
\[
\begin{align*}
x_1 + 2x_2 + 2x_3 + x_4 &= 0 \\
2x_2 + x_4 &= 0 \\
-2x_1 - 4x_3 + 3x_4 &= 0,
\end{align*}
\]
which can be expressed in matrix form

\[
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
-2 & 0 & -4 & 3
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix}
= \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}.
\]

The row-reduced echelon form of the above matrix is

\[
\begin{pmatrix}
1 & 0 & 2 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix},
\]

hence \(S\) is spanned by the vector \((-2, 0, 1, 0)^t\).

b) Now, let \(V = \mathbb{F}^4\) where \(\mathbb{F} = \mathbb{Z}_5\) and consider the same linear functionals \(\{f_1, f_2, f_3\}\). Now, \(f_3\) can be rewritten as

\[
\begin{align*}
f_3(x_1, x_2, x_3, x_4) &= -2x_1 - 4x_3 + 3x_4 \\ &\equiv 3x_1 + x_3 + 3x_4 \mod 5.
\end{align*}
\]

In this case, determining the subspace perpendicular to \(\{f_1, f_2, f_3\}\) is equivalent to solving the linear system

\[
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
3 & 0 & 1 & 3
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix}
= \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}.
\]

Again, we need only row reduce to determine \(S\), keeping in mind that all operations are modulo 5. By the following calculations,

\[
\begin{align*}
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
3 & 0 & 1 & 3
\end{pmatrix} &\to
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
0 & 4 & 0 & 0
\end{pmatrix} \to
\begin{pmatrix}
1 & 0 & 2 & 0 \\
0 & 4 & 0 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix} \\
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
3 & 0 & 1 & 3
\end{pmatrix} &\to
\begin{pmatrix}
1 & 2 & 2 & 1 \\
0 & 2 & 0 & 1 \\
0 & 1 & 0 & 0
\end{pmatrix} \to
\begin{pmatrix}
1 & 0 & 2 & 0 \\
0 & 1 & 0 & 3 \\
0 & 0 & 1 & 0
\end{pmatrix}
\end{align*}
\]

we find that \(S\) is spanned by the vector \((-2, 0, 1, 0)^t\), which is equivalent to \((3, 0, 1, 0)^t \mod 5\).

The following results follow by careful application of the definitions and are left as an exercise.

**Theorem 2.9.** Let \(N, S\) be subspaces of a vector space \(V\).
a) If $N \subseteq S$, then $S^\perp \subseteq N^\perp$.

b) $S \subseteq (S^\perp)^\perp$

c) $(N \cup S)^\perp = N^\perp \cap S^\perp$

d) $L(S)^\perp = S^\perp$.

**Theorem 2.10.** Let $S$ be a subspace of a finite-dimensional vector space $V$. Then

$$\dim V = \dim S + \dim S^\perp.$$  

**Proof.** Let $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of $V$ and let $\alpha^* = \{\alpha_1^*, \ldots, \alpha_n^*\}$ be the corresponding dual basis. We already know that $\dim V = \dim S + \dim(V/S)$, so we will show that $\dim S^\perp = \dim(V/S)$. After possibly reindexing, we may assume that $B_S = \{\alpha_1, \ldots, \alpha_m\}$ is a basis of $S$, in which case, $B'_S = \{\alpha_{m+1}^*, \ldots, \alpha_n^*\}$ is a basis of $V/S$. Since $S^\perp$ is a subspace of $V^*$, we may find a basis $B^* \subset \alpha^*$ for $S^\perp$. Assume $B^* = \{\alpha_1^*, \ldots, \alpha_k^*\}$. By definition, we must have $\alpha_j^*(s) = 0$ for $j = 1, \ldots, k$ and all $s \in S$. In particular, $\alpha_j^*(\alpha_i) = 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, k$. However, since $\alpha^*$ is the dual basis of $\alpha$, for each $j = 1, \ldots, k$ there must be a corresponding $m + 1 \leq i \leq n$ such that $\alpha_i^*(\alpha_j) = 1$. Thus, there is a one-to-one correspondence between elements of $B'$ and $B^*$, so $\dim S^\perp = \dim(V/S)$ and $\dim V = \dim S + \dim S^\perp$. \qed 

Given any homomorphism $T \in \text{Hom}(V, W)$, the duality between a vector space and its dual space allow us to define the (operator) adjoint of $T$, denoted by $T^*$. The adjoint $T^* \in \text{Hom}(W^*, V^*)$ is defined by the condition $T^*f = fT$ for every $f \in W^*$. In particular, if $\varphi_W : W \times W^* \to \mathbb{F}$ and $\varphi_V : V \times V^* \to \mathbb{F}$ denote the dual pairings for $W$ and $V$, respectively, then the adjoint satisfies

$$\varphi_W(Tv, f) = \varphi_V(v, T^*f),$$

for all $v \in V, f \in W^*$.

As a consequence of the first isomorphism theorem, we have $(V/S)^* \cong S^\perp$ (Exercise 2). The following theorem details one connection between adjoints and perp spaces.

**Theorem 2.11.** Let $T \in \text{Hom}(V, W)$. Then

a) $(\text{im } T^*)^\perp = \ker T$

b) $\ker T^* = (\text{im } T)^\perp$.

**Proof.** a) Let $\varphi_V$ and $\varphi_W$ denote the dual pairings for $V$ and $W$, respectively. Note that

$$(\text{im } T^*)^\perp = \{a \in V \mid \varphi_V(a, T^*g) = 0, \forall T^*g \in \text{im } T^*\}.$$
Let $\alpha \in (\text{im } T^*)^\perp$. By definition of the adjoint, we have

$$
\varphi_V(\alpha, T^*g) = 0, \forall g \in W^* \\
\iff \varphi_W(T\alpha, g) = 0 \quad \forall g \in W^* \\
\iff g(T\alpha) = 0 \quad \forall g \in W^*.
$$

Since this is true for every $g \in W^*$, we must have $T\alpha = 0$, so $\alpha \in \ker T$. Thus, $(\text{im } T^*)^\perp \subseteq \ker T$. On the other hand, since our logic holds in the reverse order, we also have $\ker T \subseteq (\text{im } T^*)^\perp$, hence $\ker T = (\text{im } T^*)^\perp$.

b) Note that $\ker T^* = \{g \in W^* \mid T^*g = 0\}$, where $T^*g \in V^*$ equals 0 if and only if $(T^*g)v = 0$ for every $v \in V$. So, if $g \in \ker T^*$, we have

$$
(T^*g)v = 0, \forall v \in V \\
\iff \varphi_V(v, T^*g) = 0, \forall v \in V \\
\iff \varphi_W(Tv, g) = 0, \forall v \in V,
$$

$g \in (\text{im } T)^\perp \iff g \in \ker T^*$.

An important property of adjoints is detailed in the following theorem.

**Theorem 2.12.** If $T \in \text{Hom}(U, V)$ and $L \in \text{Hom}(V, W)$, then

$$(LT)^* = T^*L^*.$$ 

**Proof.** Note that $(LT)^* \in \text{Hom}(W^*, U^*)$, since $LT \in \text{Hom}(U, W)$. Let $u \in U, g \in W^*$ be arbitrary. Then,

$$
\varphi_W(LTu, g) = \varphi_V(Tu, L^*g) = \varphi_U(u, T^*L^*g),
$$

by the corresponding definitions for $L^*$ and $T^*$. Since $u, g$ are arbitrary, the equality follows. $\Box$

The following corollary is left as an exercise (Exercise 3).

**Corollary 2.13.** If $T \in \text{Hom}(V)$ is invertible, then

$$(T^*)^{-1} = (T^{-1})^*.$$ 

We record two important results about the adjoint of a homomorphism.

**Theorem 2.14.** Let $T \in \text{Hom}(V, W)$, where $V$ and $W$ are finite-dimensional, then

$\text{rank}(T) = \text{rank}(T^*)$. 

36
Proof. Since $V$ is finite-dimensional, we have

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

$$\Rightarrow \text{rank } T = \dim V - \text{nullity } T.$$  

By Theorem 2.11a, ker $T = (\text{im } T^*)^\perp$, hence

$$\text{rank } T = \dim V^* - \dim (\text{im } T^*)^\perp,$$

where we also used the fact that $\dim V^* = \dim V$. Note that, since $T^* \in \text{Hom}(W^*, V^*)$, $\text{im}(T^*)$ is a subspace of $V^*$. Finally, by Theorem 2.10, $\dim V^* = \dim (\text{im } T^*) + \dim (\text{im } T^*)^\perp$, so $\text{rank } T^* = \dim V^* - \dim (\text{im } T^*)^\perp$. \hfill $\square$

**Theorem 2.15.** Let $T \in \text{Hom}(V, W)$, where $V$ and $W$ are finite-dimensional with bases $\alpha$ and $\beta$, respectively. Denote the corresponding dual bases by $\alpha^*$ and $\beta^*$, respectively. Then,

$$\Gamma(\alpha^*, \beta^*)T^* = (\Gamma(\alpha, \beta)T)^t.$$  

That is, the matrix representation of the adjoint $T^*$ is the transpose of the matrix representation of $T$.

### 2.2 Multilinear Functionals and modules

Let $V_1, \ldots, V_n$ and $V$ be vector spaces over the field $\mathbb{F}$, where $n \geq 1$ is finite. A mapping $\phi : V_1 \times \cdots \times V_n \to V$ is called multilinear (of degree $n$) if it is linear in each component. That is, $\phi$ is multilinear if for each $i = 1, \ldots, n$

$$\phi(v_1, \ldots, av_i + bw_i, \ldots, v_n) = a\phi(v_1, \ldots, v_i, \ldots, v_n) + b\phi(v_1, \ldots, w_i, \ldots, v_n),$$

for all $a, b \in \mathbb{F}, v_i, w_i \in V_i$ and $v_j \in V_j, j \neq i$. Note that multilinear mappings of degree 1 and 2 are what we usually call linear and bilinear mappings, respectively. So, homomorphisms are a special case of multilinear mappings.

One of the most important multilinear mappings in linear algebra is the determinant of a square matrix. We will spend some time developing the theory of special types of multilinear mappings and modules as we develop the notion of determinants.

Another special type of multilinear mapping are those for which $V = \mathbb{F}$ in the definition, which we call multilinear functionals.

#### 2.2.1 Modules

Let $R$ be a commutative ring with identity. A set $M$ is called an $R$-module (or module over $R$) if
M1) there is an addition operation \((\alpha, \beta) \mapsto \alpha + \beta\) from \(M \times M \to M\), under which \(M\) is a commutative group,

M2) there is a multiplication operation \((c, \alpha) \mapsto c\alpha\) from \(R \times M \to M\) such that
   
   i) \((c_1 + c_2)\alpha = c_1\alpha + c_2\alpha\),  
   
   ii) \(c(\alpha + \beta) = c\alpha + c\beta\),  
   
   iii) \((c_1c_2)\alpha = c_1(c_2\alpha)\),  
   
   iv) \(1\alpha = \alpha\),

   for all \(c, c_1, c_2 \in R\) and \(\alpha, \beta \in M\).

If \(R\) is not commutative, we may still define the structure of a module, but we must distinguish between left \(R\)-modules and right \(R\)-modules. Note that modules play the role of vector spaces, however, since every element of \(R\) does not necessarily have an inverse, we must be cautious with scalar multiplication.

**Example 2.16.**

1) The set \(R^n\) consisting of \(n\)-tuples of ring elements is an \(R\)-module. This is one of the most basic, and common, examples; one that we will likely refer to often. As in the case of \(F^n\), the operations on \(R^n\) are componentwise addition and scalar multiplication.

2) Likewise, the set \(M_{m,n}(R)\) of \(m \times n\) matrices with elements in \(R\), is an \(R\)-module.

3) \(R\), itself, can be regarded as an \(R\)-module.

Most of the concepts defined for vector spaces can be analogously defined for \(R\)-modules, although extra care must be taken. For instance, if \(a, b \in R\) are nonzero and \(u, v \in M\), suppose \(au + bv = 0\), that is to say \(\{u, v\}\) is linearly dependent. As opposed to vector spaces, we cannot necessarily conclude that \(u\) is a scalar multiple of \(v\), or vice versa, since \(a\) and \(b\) may not have a multiplicative inverse. Another important difference between vector spaces and modules, is that a module does not necessarily admit a basis. A **basis** for an \(R\)-module \(M\) is a linearly independent subset of \(M\) which spans (generates) \(M\). An \(R\)-module \(M\) is called **free** if either \(M = \{0\}\) or \(M\) has a basis. If \(M\) has a finite basis with \(n\) elements, \(M\) is called a **free \(R\)-module with \(n\) generators**. Furthermore, we say a module \(M\) is **finitely generated** if it contains a finite subset which spans \(M\). The **rank** of a finitely generated module is the smallest integer \(k\) such that some \(k\) elements span \(M\). Note that a module may be finitely generated without having a finite basis!

A **submodule** of an \(R\)-module \(M\) is a nonempty subset \(S\) of \(M\) that is an \(R\)-module under the same operations.
Theorem 2.17. A nonempty subset $S$ of an $R$-module $M$ is a submodule if and only if 
$$c_1, c_2 \in R, \alpha, \beta \in M \Rightarrow c_1 \alpha + c_2 \beta \in M.$$

Theorem 2.18. Let $R$ be a commutative ring with identity. If $M$ is a free $R$-module with $n$ generators, then the rank of $M$ is $n$.

Analogous to vector spaces, if $M, N$ are $R$-modules and $f : M \rightarrow N$, we call $f$ an $R$-homomorphism if
$$f(au + bv) = af(u) + bf(v),$$
for all $a, b \in R, u, v \in M$. Note, however, that the term “linear transformation” is not typically used in this context. The set of all $R$-homomorphisms from $M$ to $N$ is denoted by $\text{Hom}_R(M, N)$, which is an $R$-module under addition of functions and scalar multiplication. We will also refer to $R$-monomorphisms, $R$-epimorphisms, $R$-endomorphisms and $R$-endomorphisms, which are defined analogously. In particular, we define the dual module, $M^*$, which consists of all linear functions from $M$ to $R$; that is $M^* = \text{Hom}_R(M, R)$.

2.2.2 Multilinear forms

As in the previous section, $R$ denotes a commutative ring with identity. Let $M_1, \ldots, M_k$ be $R$-modules. A mapping $L : M_1 \times \cdots \times M_k \rightarrow R$ which satisfies
$$L(v_1, \ldots, av_i + bw_i, \ldots, v_k) = aL(v_1, \ldots, v_i, \ldots, v_k) + bL(v_1, \ldots, w_i, \ldots, v_k),$$
for each $i = 1, \ldots, k$, all $a, b \in R, v_i, w_i \in M_i$ and $v_j \in M_j, j \neq i$ is called a multilinear form of degree $k$ or a $k$-linear form. Note that multilinear forms are analogous to multilinear functionals, the difference being whether we are working over a ring or a field (modules or vector spaces).

Let $M$ be an $R$-module. For a positive integer $k$ let

$$M^k = \underbrace{M \times \cdots \times M}_k \{ (v_1, \ldots, v_k) | v_i \in V \},$$

that is, the Cartesian product of $k$ copies of $M$. Of particular interest are multilinear forms mapping $M^k$ to $R$. We will denote the collection of all multilinear forms on $M^k$ by $T^k(M)$. Under pointwise addition and scalar multiplication,

$$(aL)(v_1, \ldots, v_k) = a(L(v_1, \ldots, v_k))$$

$$(L + L')(v_1, \ldots, v_k) = L(v_1, \ldots, v_k) + L'(v_1, \ldots, v_k),$$

$T^k(M)$ becomes an $R$-module. Note that, for $k = 1$, $T^1(M)$ is the dual space $V^*$.

Let $L$ be an element of $T^k(M)$. We say $L$ is alternating if $L(v_1, \ldots, v_k) = 0$ whenever $v_i = v_j$ for some $i \neq j$. As a consequence of the definition, we have the following useful fact, which says that if we swap the order of any two vectors $L$ changes sign.
Lemma 2.19. Let $L \in T^k(M)$ be alternating. Then
\[ L(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -L(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) \]
for any $i \neq j$.

Proof. If $L$ is alternating, then
\[ L(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_k) = 0. \]
However, since $L$ is linear in each component, the left hand side can be expanded to yield
\[
L(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_k) + L(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_k) + L(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) + L(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k) = 0.
\]
Again, by the definition of alternating, the first and last terms on the left hand side are zero, leaving
\[
L(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) + L(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = 0
\implies L(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -L(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)
\]
\]

Example 2.20.

1) Consider the $R$-module $M_n(R)$ of $n \times n$ matrices with entries in $R$. We may form many examples of $n$-linear forms by constructing functions which are linear with respect to the rows of an $n \times n$ matrix or the columns of the matrix. We will define $n$-linear forms on $M_n(R)$ as being linear with respect to the rows. By this, we mean a mapping which associates a scalar to each matrix and is linear in each row, when all other rows are held fixed. More specifically, we may construct an $n$-linear form as follows. Let $r_1, \ldots, r_n$ denote the rows of a matrix $A \in M_n(R)$. When constructing $n$-linear forms of this type, we may write $L(A) = L(r_1, \ldots, r_n)$, so that the $n$-linear form may be regarded as a function of the rows of $A$. Define the $n$-linear form $L : R^n \times \cdots \times R^n \to R$ by $L(r_1, \ldots, r_n) = \det(A)$. On $M_n(R)$, an $n$-linear form $L$ is alternating if $L(A) = 0$ whenever any two rows of $A$ are identical. By Lemma 2.19, we have $L(A') = L(A)$, where $A'$ is the same matrix as $A$ with two rows swapped.

2) Let $j_1, \ldots, j_n$ be positive integers, $1 \leq j_k \leq n$, and let $s$ be a fixed element of $R$. For each $A \in M_n(R)$, denote the $(i, j)$ entry by $A(i, j)$ and define
\[
L(A) = sA(1, j_1) \cdots A(n, j_n).
\]
We want to show that $L$ is $n$-linear with respect to the rows of $A$. Fix all rows of $A$ except the $i$th row. Then, we may write
\[
L(r_i) = sA(1, j_1) \cdots A(n, j_n) = bA(i, j_i),
\]
where \( b \) is the product of all terms except the \( i \)th. Then, if \( r_i' = (A'(i,1), \ldots, A'(i,n)) \) is any row vector in \( R^n \), we have

\[
L(cr_i + r_i') = (cA(i,j_i) + A'(i,j_i))b = cL(r_i) + L(r_i').
\]

Clearly, this holds for each \( i \), so \( L \) is \( n \)-linear. For example, let \( \phi(A) = a_{11}a_{22} \cdots a_{nn} \) be the product of the diagonal entries of \( A \). Then, \( \phi \) is an \( n \)-linear function of the form (2.2).

### 2.2.3 Formal definition of determinant

Suppose \( L \) is an alternating \( n \)-linear form on \( M_n(R) \). Let \( A \in M_n(R) \) be a matrix with rows \( r_1, \ldots, r_n \). If we denote the rows of the \( n \times n \) identity matrix on \( R \) by \( e_1, \ldots, e_n \), then

\[
r_i = \sum_{j=1}^{n} A(i,j)e_j, \quad 1 \leq j \leq n.
\]

Since \( L \) is \( n \)-linear, we have

\[
L(A) = L(r_1, \ldots, r_n) = L \left( \sum_{j=1}^{n} A(1,j)e_j, r_2, \ldots, r_n \right)
\]

\[
= \sum_{j=1}^{n} A(1,j)L(e_j, r_2, \ldots, r_n).
\]

Similarly, since \( L \) is linear in each component, we may write

\[
L(A) = L(r_1, \ldots, r_n) = L \left( \sum_{j_1=1}^{n} A(1,j_1)e_{j_1}, r_{j_2}, \ldots, r_{j_n} \right)
\]

\[
= \sum_{j_1=1}^{n} A(1,j_1)L(e_{j_1}, \sum_{j_2=1}^{n} A(2,j_2)e_{j_2}, \ldots, r_{j_n})
\]

\[
= \sum_{j_1, j_2} A(1,j_1)A(2,j_2)L \left( e_{j_1}, e_{j_2}, \sum_{j_3=1}^{n} A(3,j_3)e_{j_3}, \ldots, r_{j_n} \right)
\]

\[
= \ldots
\]

\[
= \sum_{j_1, j_2, \ldots, j_n} A(1,j_1) \cdots A(n,j_n)L(e_{j_1}, \ldots, e_{j_n}), \quad (2.3)
\]

where the sum in (2.3) is over all sequences of positive integers \( (j_1, \ldots, j_n) \) less than or equal to \( n \). Thus, \( L \) is a finite sum of functions of the form (2.2), since \( L \) is \( n \)-linear. By
assumption, \( L \) is also alternating, thus \( L(e_{j_1}, \ldots, e_{j_n}) = 0 \) whenever any two of the indices are equal. To simplify our analysis of (2.3), we may assume no two indices are equal, since otherwise, the alternating form is 0.

In order to study (2.3) in more detail, let us digress and take a brief look at permutations. A sequence of integers \((j_1, \ldots, j_n)\), with \(1 \leq j_k \leq n\), such that no two of the \(j_k\)'s are equal, is called a permutation of degree \(n\). In general, a permutation of a set \(A\) is a function from \(A\) to \(A\) that is 1-1 and onto (bijective). For our purposes, we will primarily be interested in permutations of the set \(\Delta = \{1, \ldots, n\}\). A bijective map permuting the elements of a set is often denoted by \(\sigma\). For example, suppose \(\sigma(1) = j_1, \sigma(2) = j_2, \ldots, \sigma(n) = j_n\). In this case, the action of \(\sigma\) may be represented by a \(2 \times n\) array

\[
\sigma = \begin{pmatrix}
1 & 2 & \ldots & n \\
 j_1 & j_2 & \ldots & j_n
\end{pmatrix}.
\]

**Example 2.21.** Suppose \(n = 6\) and let

\[
\sigma = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 \\
3 & 2 & 4 & 6 & 1 & 5
\end{pmatrix}.
\]

Then, \(\sigma\) is a bijection given by \(\sigma(1) = 3, \sigma(2) = 2, \sigma(3) = 4, \sigma(4) = 6, \sigma(5) = 1, \sigma(6) = 5\).

A permutation group of a set \(A\) is a set of permutations of \(A\) that forms a group under function composition. In particular, we denote the group of permutations of \(\Delta = \{1, \ldots, n\}\) by \(S_n\), which is called the symmetric group of degree \(n\). More precisely, the law \((\sigma, \tau) \mapsto \sigma \tau\) mapping \(S_n \times S_n \rightarrow S_n\) satisfies

1) \(\sigma(\tau \gamma) = (\sigma \tau) \gamma,\) for all \(\sigma, \tau, \gamma \in S_n\)
2) There exists an element \(1 \in S_n\) such that \(\sigma 1 = 1 = 1 \sigma\) for all \(\sigma \in S_n\)
3) For every \(\sigma \in S_n\) there exists a \(\tau \in S_n\) such that \(\sigma \tau = 1 = \tau \sigma\),

making \(S_n\) into a group. The order of a permutation group of a set \(A\) is the number of distinct permutations of \(A\). Note that the order of \(S_n\) is \(n!\), which is denoted by \(|S_n| = n!\).

A permutation \(\sigma \in S_n\) is called an \(m\)-cycle if \(\sigma\) cyclically permutes a sequence of \(m\) elements, leaving all other elements fixed. That is, for \(m > 1\) and elements \(i_1, \ldots, i_m \in \Delta, \sigma(i_1) = i_2, \sigma(i_2) = i_3, \ldots, \sigma(i_{m-1}) = i_m, \sigma(i_m) = i_1\) and \(\sigma(k) = k\) for all \(k \in \Delta - \{i_1, \ldots, i_m\}\). If \(\sigma\) is such an \(m\)-cycle, we often use cycle notation to express the action of \(\sigma\). In array notation we have

\[
\sigma = \begin{pmatrix}
i_1 & i_2 & \cdots & i_{m-1} & i_m & k \\
i_2 & i_3 & \cdots & i_m & i_1 & k
\end{pmatrix},
\]

whereas, in cycle notation we would write

\[
\sigma = (i_1 i_2 \cdots i_m)(k).
\]

Note that any number not appearing in a cycle is fixed by that cycle.
Example 2.22.

1) The permutation

\[ \sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 4 & 1 & 6 \end{pmatrix} \]

is a 3-cycle, expressed in cycle notation as

\[ \sigma_1 = (135)(2)(4)(6). \]

However, it is more common to omit the cycles containing a single element, in which case, the above permutation is written as

\[ \sigma_1 = (135). \]

2) The permutation

\[ \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 6 & 1 & 4 & 5 \end{pmatrix} \]

is a 6-cycle, expressed in cycle notation as

\[ \sigma_2 = (123654). \]

3) Now, consider the permutation

\[ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 2 & 3 & 6 & 5 \end{pmatrix}. \]

By definition, \( \tau \) is not a cycle. However, \( \tau \) can be expressed as the product of cycles, specifically the product of a 2-cycle and a 4-cycle,

\[ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 2 & 3 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 6 & 5 \end{pmatrix}. \]

Each of the permutations in the product can be expressed in cycle notation as

\[ (1432)(5)(6) \quad \text{and} \quad (1)(2)(3)(4)(56), \]

but again, it is more common to omit the 1-cycles, expressing the permutation as the product

\[ \tau = (1432)(56). \]

We say two cycles are *disjoint* if they have no numbers in common. For example, the cycles (135) and (46) are disjoint, while the cycles (135) and (654) are not disjoint. A 2-cycle in \( S_n \) is called a *transposition*.

We now list several standard results about permutation groups, omitting proofs.
Theorem 2.23. Any permutation $\sigma \in S_n$ can be expressed as a product of disjoint cycles.

Theorem 2.24. Every permutation $\sigma \in S_n$, $n > 1$, can be expressed as a product of transpositions.

Note that a factorization of a permutation into a product of transpositions is not unique. However, if $\sigma$ can be factored into an even number of transpositions, then any factorization of $\sigma$ into transpositions will have an even number of transpositions. Likewise, if $\sigma$ can be factored into an odd number of transpositions, any such factorization will have an odd number of transpositions. This allows us to define $\sigma \in S_n$ to be even if it can be factored into an even number of transpositions, and odd if $\sigma$ can be factored into an odd number of transpositions. If $\sigma$ is even, we define the sign of $\sigma$ to be 1 and write $\text{sgn}(\sigma) = 1$, while if $\sigma$ is odd we define $\text{sgn}(\sigma) = -1$.

With the tools of basic permutation theory, we may return to our discussion of determinants. We may now express the formula (2.3) as

$$L(A) = \sum_{\sigma \in S_n} A(1, \sigma(1)) \cdots A(n, \sigma(n)) L(e_{\sigma(1)}, \ldots, e_{\sigma(n)}).$$

We are also in position to restate Lemma 2.19 in terms of permutations.

Lemma 2.25. If $L$ is an alternating $n$-linear form on $M^n$ and $\sigma \in S_n$, then for all $(v_1, \ldots, v_n) \in M^n$

$$L(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}) = \text{sgn}(\sigma)L(v_1, v_2, \ldots, v_n).$$

Proof. Since every $\sigma \in S_n$ is a product of transpositions, we may repeatedly apply Lemma 2.19 for each transposition. So, if $\sigma = \sigma_1 \cdots \sigma_k$, where each $\sigma_j$ is a transposition, note that $\text{sgn}(\sigma_j) = -1$ and $\text{sgn}(\sigma) = (-1)^k$. For each $\sigma_j$, we apply Lemma 2.19, which results in a sign change each time, hence

$$L(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}) = (-1)^k L(v_1, v_2, \ldots, v_n).$$

Finally, we define the determinant of a matrix $A \in M_n(R)$ as the $n$-linear form

$$\text{det}(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma)A(1, \sigma(1)) \cdots A(n, \sigma(n)).$$

This definition is arrived at by seeking a determinant function on $M_n(R)$. A determinant function on $M_n(R)$ is an alternating, $n$-linear form, $D$, such that $D(I) = 1$. The above definition of $\text{det}$ is clearly a determinant function, in fact, it is the only one.
**Theorem 2.26.** Let $R$ be a commutative ring with identity and let $n$ be a positive integer. The function $\det$ defined by (2.4) is the unique alternating $n$-linear form on $M_n(R)$ such that $\det(I) = 1$. Furthermore, if $L$ is any alternating $n$-linear form on $M_n(R)$, then for each $A \in M_n(R)$

$$L(A) = \det(A)L(I).$$

**Example 2.27.** Let $A \in M_2(R)$ be a $2 \times 2$ matrix. We want to compute the determinant of $A$ using (2.4). Note that $|S_2| = 2!$, hence there are only 2 unique permutations of $\{1, 2\}$, namely $\sigma_1 = (12)$ and $\sigma_2 = (21)$. Furthermore, $\text{sgn}(\sigma_1) = 1, \text{sgn}(\sigma_2) = -1$, since $\sigma_1$ is the product of 0 transpositions and $\sigma_2$ is a single transposition. Using (2.4), we have

$$\det(A) = \sum_{\sigma \in S_2} \text{sgn}(\sigma)A(1, \sigma(1))A(2, \sigma(2))$$

$$= \text{sgn}(\sigma_1)A(1, \sigma_1(1))A(2, \sigma_1(2)) + \text{sgn}(\sigma_2)A(1, \sigma_2(1))A(2, \sigma_2(2))$$

$$= (1A(1, 1)A(2, 2)) + (-1A(1, 2)A(2, 1)).$$

In more familiar form, if

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

this yields $\det(A) = ad - bc$.

**Theorem 2.28.** Let $A, B \in M_n(R)$. Then

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

**Proof.** Let $B$ be a fixed $n \times n$ matrix over $R$ and define

$$D(A) = \det(AB),$$

for any $A \in M_n(R)$. Denoting the rows of $A$ by $r_1, \ldots, r_n$, we may write

$$D(A) = D(r_1, \ldots, r_n) = \det(r_1B, \ldots, r_nB).$$

It is straightforward to verify that $D$ is alternating and $n$-linear, since $\det$ is alternating and $n$-linear. By the previous theorem, we have

$$D(A) = \det(A)D(I).$$

But, $D(I) = \det(e_1B, \ldots, e_nB) = \det(B)$, hence

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

\[\square\]

**Exercises - Section 2.**
1. Let $f \in \text{Hom}(V, \mathbb{F})$ be a linear functional with $f \neq 0$. Show that $V/\ker(f) \cong \mathbb{F}$.

2. If $S$ is a subspace of a vector space $V$, use the first isomorphism theorem to show:
   (a) $(V/S)^* \cong S^\perp$
   (b) $V^* = S^* \oplus S^\perp$

3. Let $T \in \text{Hom}(V)$ be an invertible linear operator. Use Theorem 2.12 to show that
   $$(T^*)^{-1} = (T^{-1})^*.$$ 

4. If $N_1, N_2$ are submodules of an $R$-module $M$, show that $N_1 \cap N_2$ and $N_1 + N_2$ are submodules of $M$.

3 Tensor products

3.1 Free vector spaces and free modules

We begin our pursuit of constructing the tensor product by discussing how to form a vector space (or $R$-module) from a given set. Let $X$ be any set. Formally, we construct a vector space over the field $\mathbb{F}$ by constructing the direct sum of $|X|$ copies of $\mathbb{F}$. Let $\mathbb{F}\langle X \rangle = \{f : X \to \mathbb{F} \mid f(x) = 0 \text{ for all but finitely many } x \in X\}$. We make $\mathbb{F}\langle X \rangle$ into a vector space by defining addition and scalar multiplication component-wise; that is, if $f, g \in \mathbb{F}\langle X \rangle, a \in \mathbb{F}$ then the elements $(f + g)(x) \in \mathbb{F}\langle X \rangle$ and $(af)(x) \in \mathbb{F}\langle X \rangle$ are defined by

$$(f + g)(x) = f(x) + g(x)$$
$$ (af)(x) = af(x).$$

A basis of $\mathbb{F}\langle X \rangle$ is given by the collection $\{\delta_x \mid x \in X\}$ where $\delta_x(u) = \begin{cases} 1, & x = u \\ 0, & x \neq u \end{cases}$

To show that $\{\delta_x \mid x \in X\}$ is linearly independent, suppose

$$\sum_{i=1}^{n} a_i \delta_{x_i} = 0 \implies \sum_{i=1}^{n} a_i \delta_{x_i}(u) = 0 \forall u \in X.$$ 

In particular, for each $x_j, j = 1, \ldots, n$, we have

$$\sum_{i=1}^{n} a_i \delta_{x_i}(x_j) = 0 \implies a_j = 0.$$
Now, to show that \( \{ \delta_x \mid x \in X \} \) spans \( \mathbb{F}\langle X \rangle \), let \( f \in \mathbb{F}\langle X \rangle \) be arbitrary. By definition, we know \( f(x) = 0 \) for all but finitely many \( x \in X \). Suppose \( f(x_i) = a_i \) for \( i = 1, \ldots, n \), where \( x_1, \ldots, x_n \) is the finite collection of elements for which \( f \) is nonzero. We will show that \( f \) may be expressed by the sum \( f = \sum_{i=1}^{n} a_i \delta_{x_i} \). Evaluating this sum at \( x_j \), we have

\[
\sum_{i=1}^{n} a_i \delta_{x_i}(x_j) = a_1 \delta_{x_1}(x_j) + a_2 \delta_{x_2}(x_j) + \cdots + a_n \delta_{x_n}(x_j) = a_j,
\]

since \( \delta_{x_i}(x_j) = 0 \) for \( i \neq j \) and \( \delta_{x_j}(x_j) = 1 \). The vector space \( \mathbb{F}\langle X \rangle \) is called the free vector space on \( X \). Although, formally, the basis is given by \( \{ \delta_x \mid x \in X \} \), there is a natural injection map \( \iota : X \to \mathbb{F}\langle X \rangle \) defined by \( \iota(x) = \delta_x \). It is straightforward to show the injection map is bijective, so we may regard the set \( X \) as a basis of \( \mathbb{F}\langle X \rangle \). That is, we identify an element \( x \in X \) with the function that has value 1 on \( x \) and 0 on every other element of \( X \). Furthermore, the free vector space \( \mathbb{F}\langle X \rangle \) consists of formal, finite linear combinations of elements of \( X \).

Similarly, if \( M \) is a set, we may construct the free \( \mathbb{R}\)-module on \( M \) by the same process. We omit the details of this construction, as they are almost identical to the vector space case. Analogous to free vector spaces, we will denote the free \( \mathbb{R}\)-module on \( M \) by \( \mathbb{R}\langle M \rangle \).

### 3.2 Universal problem

Suppose \( V_1, V_2, V \) are all vector spaces over \( \mathbb{F} \) and let \( \phi : V_1 \times V_2 \to V \) be a bilinear mapping. We may construct other bilinear mappings from \( \phi \) as follows. Let \( h : V \to W \) be a homomorphism from \( V \) to another vector space \( W \) (over \( \mathbb{F} \)). Then \( h \circ \phi \) is a bilinear mapping from \( V_1 \times V_2 \to W \).

**Exercise 3.1.** Verify that \( h \circ \phi \) is a bilinear mapping.

We wish to answer the following question.

**Problem 3.1.** Can we determine a vector space \( V \) and bilinear mapping \( \phi \) (as above) so that given any bilinear mapping

\[
\psi : V_1 \times V_2 \to W
\]

there exists exactly one homomorphism \( h \in \text{Hom}(V, W) \) such that \( h \circ \phi = \psi \).

Problem 3.1 is called the universal problem (for bilinear mappings), a solution of which consists of a vector space and a bilinear mapping, which we denote as a pair \( (V, \phi) \).

What we seek is, in some sense, the most general bilinear mapping on \( V_1 \times V_2 \). We will produce a solution to the universal problem by construction. However, we will first show that a solution \( (V, \phi) \) is unique up to isomorphism.
Lemma 3.1. The solution to the universal problem is unique up to isomorphism.

Proof. Suppose \((V, \phi)\) and \((V', \phi')\) are two solutions to Problem 1. Since \((V, \phi)\) is a solution, for any \(\psi : V_1 \times V_2 \rightarrow W\) there exists \(h : V \rightarrow W\) such that \(h \circ \phi = \psi\). Setting \(W = V'\) and \(\psi = \phi'\), there must exist an \(h : V \rightarrow V'\) such that
\[
    h \circ \phi = \phi'.
\]

Likewise, since \((V', \phi')\) is a solution, for any \(\psi : V_1 \times V_2 \rightarrow W\) there exists \(h' : V' \rightarrow W\) such that \(h' \circ \phi' = \psi\). Setting \(W = V\) and \(\psi = \phi\), there must exist an \(h' : V' \rightarrow V\) such that
\[
    h' \circ \phi' = \phi.
\]

On one hand, we may plug (3.2) into (3.1), yielding
\[
    h \circ (h' \circ \phi') = \phi',
\]
which can only be true if \(h \circ h' = \text{id}_V\). On the other hand, we may plug (3.1) into (3.2), yielding
\[
    h' \circ (h \circ \phi) = \phi,
\]
which can only be true if \(h' \circ h = \text{id}_V\). Thus, \(h : V \rightarrow V'\) and \(h' : V' \rightarrow V\) are inverse isomorphisms. \(\square\)

We have shown that, if a solution exists, the solution to Problem 3.1 is unique. However, it is worth reiterating that the solution depends on \(V_1 \times V_2\). More plainly, we may express this by saying \((V, \phi)\) is a solution to the universal problem for \(V_1 \times V_2\) or \((V, \phi)\) is universal for \(V_1 \times V_2\).

3.2.1 Constructing a solution to the universal problem

Let \(V_1, V_2\) be two vector spaces over \(\mathbb{F}\) and consider the free vector space \(\mathbb{F} \langle V_1 \times V_2 \rangle\) consisting of linear combinations of elements in \(V_1 \times V_2\) with coefficients in \(\mathbb{F}\). Note that \(\mathbb{F} \langle V_1 \times V_2 \rangle\) is a vector space of dimension \(n_1 n_2\), where \(\dim(V_1) = n_1\) and \(\dim(V_2) = n_2\).

Recall, from Section 3.1, that \(\mathbb{F} \langle V_1 \times V_2 \rangle\) consists of functions mapping \(V_1 \times V_2\) into \(\mathbb{F}\) and has a basis consisting of \(\{\delta_{(u,v)} | u \in V_1, v \in V_2\}\). Also, recall that there is a natural injection \(i : V_1 \times V_2 \rightarrow \mathbb{F} \langle V_1 \times V_2 \rangle\) defined by \(i(u, v) = \delta_{(u,v)}\).

Let \(S\) be the subspace of \(\mathbb{F} \langle V_1 \times V_2 \rangle\) generated by elements of one of the following forms
\[
\delta_{(u_1+u_2,v)} - \delta_{(u_1,v)} - \delta_{(u_2,v)} \quad (3.3)
\]
\[
\delta_{(u,v_1+v_2)} - \delta_{(u,v_1)} - \delta_{(u,v_2)} \quad (3.4)
\]
\[
\delta_{(au,v)} - a\delta_{(u,v)} \quad (3.5)
\]
\[
\delta_{(u,av)} - a\delta_{(u,v)} \quad (3.6)
\]
where \( a \in F, u, u_1, u_2 \in V_1 \) and \( v, v_1, v_2 \in V_2 \). By the natural injection, we may identify such elements with elements of the form
\[
(u_1 + u_2, v) - (u_1, v) - (u_2, v) \\
(u, v_1 + v_2) - (u, v_1) - (u, v_2) \\
(au, v) - a(u, v) \\
(u, av) - a(u, v),
\]
where \( a \in F, u, u_1, u_2 \in V_1 \) and \( v, v_1, v_2 \in V_2 \). However, despite the notation being more complicated, we will utilize the former notation.

Set \( V = F \langle V_1 \times V_2 \rangle / S \), elements of which are cosets of the form \( \delta(u,v) + S \). There is a natural mapping \( \phi: V_1 \times V_2 \to V \) defined by
\[
\phi(u,v) = \delta(u,v) + S,
\]
which can be expressed as the composition \( \phi = \pi_S \circ \iota \), where \( \pi_S: F \langle V_1 \times V_2 \rangle \to V \) is the canonical projection given by \( \pi_S(\delta(u,v)) = \delta(u,v) + S \), as defined in Section 1.4.

**Theorem 3.2.** The pair \((V, \phi)\) is a solution to Problem 3.1.

**Proof.** We first show that \( \phi: V_1 \times V_2 \to V \) is bilinear. We need to show that
\[
\phi(u_1 + u_2, v) = \phi(u_1, v) + \phi(u_2, v),
\]
\[
\phi(u, v_1 + v_2) = \phi(u, v_1) + \phi(u, v_2),
\]
and
\[
\phi(au, v) = a\phi(u, v),
\]
\[
\phi(u, av) = a\phi(u, v).
\]

By the definition (3.7) of \( \phi \), (3.8) is equivalent to showing
\[
\delta(u_1 + u_2,v) + S = (\delta(u_1,v) + S) + (\delta(u_2,v) + S).
\]

However, this is true if and only if
\[
\delta(u_1 + u_2,v) - \delta(u_1,v) - \delta(u_2,v) \in S,
\]
which follows from the definition of \( S \). Similarly, (3.9) is true if and only if
\[
\delta(u,v_1 + v_2) - \delta(u,v_1) - \delta(u,v_2) \in S.
\]

Relations (3.10)-(3.11) follow from the definition of \( S \), as well.
Now, we need to show for any bilinear $\psi : V_1 \times V_2 \to W$ there exists a linear $h : V \to W$ such that $h \circ \phi = \psi$. Define a linear transformation $T \in \text{Hom}(\mathbb{F}(V_1 \times V_2), W)$ by $T(\delta_{(u,v)}) = \psi(u, v)$. Since the elements $\delta_{(u,v)}$ are a basis for $\mathbb{F}(V_1 \times V_2)$, this definition of $T$ is well defined. The map $\psi$ is bilinear, so

\[
\begin{align*}
\psi(u_1 + u_2, v) &= \psi(u_1, v) + \psi(u_2, v), \\
\psi(u, v_1 + v_2) &= \psi(u, v_1) + \psi(u, v_2), \\
\psi(au, v) &= a\psi(u, v), \\
\psi(u, av) &= a\psi(u, v),
\end{align*}
\]

or equivalently,

\[
\begin{align*}
\psi(u_1 + u_2, v) - \psi(u_1, v) - \psi(u_2, v) &= 0, \\
\psi(u, v_1 + v_2) - \psi(u, v_1) - \psi(u, v_2) &= 0, \\
\psi(au, v) - a\psi(u, v) &= 0, \\
\psi(u, av) - a\psi(u, v) &= 0.
\end{align*}
\]

By the definition of $T$, these conditions translate to the conditions

\[
\begin{align*}
T(\delta_{(u_1+u_2,v)}) - T(\delta_{(u_1,v)}) - T(\delta_{(u_2,v)}) &= 0, \\
T(\delta_{(u,v_1+v_2)}) - T(\delta_{(u,v_1)}) - T(\delta_{(u,v_2)}) &= 0, \\
T(\delta_{(au,v)}) - aT(\delta_{(u,v)}) &= 0, \\
T(\delta_{(u,av)}) - aT(\delta_{(u,v)}) &= 0.
\end{align*}
\]

Since $T$ is linear, we equivalently have

\[
\begin{align*}
T(\delta_{(u_1+u_2,v)}) - \delta_{(u_1,v)} - \delta_{(u_2,v)} &= 0, \\
T(\delta_{(u,v_1+v_2)}) - \delta_{(u,v_1)} - \delta_{(u,v_2)} &= 0, \\
T(\delta_{(au,v)}) - a\delta_{(u,v)} &= 0, \\
T(\delta_{(u,av)}) - a\delta_{(u,v)} &= 0,
\end{align*}
\]

which shows that $T(S) = \{0\}$, or equivalently, $S \subseteq \ker(T)$, since each of the elements in the above conditions are one of the forms (3.3)-(3.6). Thus, the first isomorphism theorem yields a linear transformation $h \in \text{Hom}(V, W)$ such that $h \circ \pi_S = T$, that is $h(\pi_S(\delta_{(u,v)})) = T(\delta_{(u,v)})$. Since $\phi = \pi_S \circ \iota$ and by the definition of $T$, we have $h(\phi(u, v)) = \psi(u, v)$ for all $(u, v) \in V_1 \times V_2$. Hence, $h \circ \phi = \psi$, so the pair $(V, \phi)$ is a solution to the universal problem.

Now that we know a solution to the universal problem exists and is unique, we are ready to introduce notation. The vector space $V = \mathbb{F}(V_1 \times V_2)/S$ is henceforth denoted as $V_1 \otimes V_2$ and called the tensor product of $V_1$ and $V_2$. Elements of $V_1 \otimes V_2$ are cosets of the
form $\phi(u,v) = \delta_{(u,v)} + S$, which we henceforth denote as $u \otimes v$, so $\phi(u,v) = u \otimes v$.

Furthermore, we call such elements tensors and we call $\phi$ the canonical map. Since $\phi$ is bilinear we know

$$\phi(u_1 + u_2, v) = \phi(u_1, v) + \phi(u_2, v),$$

which implies

$$(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v.$$

Likewise, since

$$\phi(u, v_1 + v_2) = \phi(u, v_1) + \phi(u, v_2),$$
$$\phi(au, v) = a\phi(u,v),$$
$$\phi(u, av) = a\phi(u,v),$$

we have

$$u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2,$$
$$ (au) \otimes v = a(u \otimes v),$$
$$ u \otimes (av) = a(u \otimes v).$$

Tensors of the form $u \otimes v$ are called pure tensors (or elementary, decomposable). It may not be apparent that all tensors are not pure. In general, we can say that any tensor is a finite sum of pure tensors, that is, any element of $V_1 \otimes V_2$ is of the form

$$\sum_{i,j} a_{ij} u_i \otimes v_j,$$

where $u_i, v_j$ are basis elements of $V_1, V_2$, respectively. This is due to the following result.

**Theorem 3.3.** Let $B_1, B_2$ be bases for $V_1, V_2$, respectively. The set $\{\beta_1 \otimes \beta_2 | \beta_1 \in B_1, \beta_2 \in B_2\}$ is a basis of $V_1 \otimes V_2$.

This theorem can be proven by a straightforward argument, however, the next theorem provides another approach and also provides a useful result.

**Theorem 3.4.** If $u_1, \ldots, u_n$ are linearly independent elements of $V_1$ and $v_1, \ldots, v_n \in V_2$ are arbitrary, then

$$\sum_{i} u_i \otimes v_i = 0 \Rightarrow v_i = 0 \text{ for all } i.$$

Moreover, $u \otimes v = 0$ if and only if $u = 0$ or $v = 0$. 

51
Proof. Assume \( u_i, v_i \) are nonzero and suppose \( \sum u_i \otimes v_i = 0 \). The universal problem states that for any bilinear mapping \( \psi : V_1 \times V_2 \to W \) there is a linear map \( h : V_1 \otimes V_2 \to W \) such that \( h \circ \phi = \psi \). Thus,

\[
    h \left( \sum u_i \otimes v_i \right) = \sum \psi(u_i, v_i) = 0
\]

\[
    \Rightarrow \sum_i (h \circ \phi)(u_i, v_i) = \sum_i \psi(u_i, v_i) = 0.
\]

Since this holds for any bilinear mapping, we may define a particular bilinear mapping that does what we need. Let \( f \in V_1^* \), \( g \in V_2^* \) be two linear functionals and define

\[
    \psi(u, v) = f(u)g(v).
\]

The reader may check that \( \psi : V_1 \times V_2 \to \mathbb{F} \) is a bilinear mapping. Having assumed that \( u_1, \ldots, u_n \) are linearly independent, we may consider the corresponding dual functionals \( u_j^* \in V_1^* \) defined by

\[
    u_j^*(u_i) = \delta_{ij}.
\]

Then, setting \( f = u_j^* \), we have

\[
    \psi(u, v) = u_j^*(u)g(v)
\]

and

\[
    \sum_i u_j^*(u_i)g(v_i) = 0,
\]

where \( g \in V_2^* \) is arbitrary. Thus, by the definition of \( u_j^* \), we have \( g(v_j) = 0 \) for any \( g \in V_2^* \), hence \( v_j = 0 \). Repeating the argument for each \( j \) yields the desired result.

\[\square\]

**Corollary 3.5.** If \( \dim(V_1) = n_1 \) and \( \dim(V_2) = n_2 \), then \( \dim(V_1 \otimes V_2) = n_1 n_2 \).

For example, using the standard basis \( \{e_1, e_2\} \) of \( \mathbb{F}^2 \), a basis of \( \mathbb{F}^2 \otimes \mathbb{F}^2 \) is \( \{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\} \).

Now, suppose \( \varphi \in \text{Hom}(V, V') \) and \( \rho \in \text{Hom}(W, W') \) are linear transformations. The universal problem allows us to associate linear maps with bilinear maps. Define a mapping from \( V \times W \) to \( V' \otimes W' \) by \( (v, w) \mapsto \varphi(v) \otimes \rho(w) \). It is straightforward to check that this map is bilinear, since both \( \varphi \) and \( \rho \) are linear. For instance, we have

\[
    (v_1 + v_2, w) \mapsto \varphi(v_1 + v_2) \otimes \rho(w),
\]

but since \( \varphi \) is linear, we have

\[
    \varphi(v_1 + v_2) \otimes \rho(w) = (\varphi(v_1) + \varphi(v_2)) \otimes \rho(w) = \varphi(v_1) \otimes \rho(w) + \varphi(v_2) \otimes \rho(w),
\]

52
by the bilinearity of the canonical map. By the universal property, we may associate to
the bilinear map \((v, w) \mapsto \varphi(v) \otimes \rho(w)\) a linear map \(h \in \text{Hom}(V \otimes W, V' \otimes W')\). Note that
the map \((v, w) \mapsto \varphi(v) \otimes \rho(w)\) from \(V \times W \rightarrow V' \otimes W'\) is taking the place of \(\psi\) in the
universal problem. Thus, we have
\[
h(\varphi(v, w)) = \varphi(v) \otimes \rho(w),
\]
or equivalently,
\[
h(v \otimes w) = \varphi(v) \otimes \rho(w).
\]
We denote this particular linear map \(h\) by \(\varphi \otimes \rho\). Note that \(\varphi \otimes \rho : V \otimes W \rightarrow V' \otimes W'\) is a
linear map. In conclusion, given linear maps \(\varphi : V \rightarrow V'\) and \(\rho : W \rightarrow W'\), we may define
the tensor product of these maps by the condition
\[
\varphi \otimes \rho(v \otimes w) = \varphi(v) \otimes \rho(w).
\]

**Example 3.6.** Suppose \(\varphi : \mathbb{F}^2 \rightarrow \mathbb{F}^2\) and \(\rho : \mathbb{F}^2 \rightarrow \mathbb{F}^2\) are linear transformations. Let
\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\]
and
\[
B = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}
\]
be the matrices of \(\varphi\) and \(\rho\) in the standard basis \(\{e_1, e_2\}\), respectively. By our previous
discussion, we may define a linear map \(\varphi \otimes \rho : \mathbb{F}^2 \otimes \mathbb{F}^2 \rightarrow \mathbb{F}^2 \otimes \mathbb{F}^2\). Let’s compute the
matrix of \(\varphi \otimes \rho\) with respect to the basis \(\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}\). Just as we would
for any other linear transformation, we must compute the action of \(\varphi \otimes \rho\) on each basis
element. What helps us here is that for any \(u \in \mathbb{F}^2\), \(\varphi(u) = Au\). Likewise, \(\rho(u) = Bu\).
Furthermore, by the definition of \(\varphi \otimes \rho\), we know
\[
\varphi \otimes \rho(u \otimes v) = \varphi(u) \otimes \rho(v) = (Au) \otimes (Bv).
\]

Naturally, we will denote the matrix of \(\varphi \otimes \rho\) as \(A \otimes B\). Thus, we have
\[
A \otimes B(e_1 \otimes e_1) = (Ae_1) \otimes (Be_1) = (ae_1 + ce_2) \otimes (a'e_1 + c'e_2)
\]
\[
= ae_1 \otimes (a'e_1 + c'e_2) + ce_2 \otimes (a'e_1 + c'e_2)
\]
\[
= ae_1 \otimes a'e_1 + ae_1 \otimes c'e_2 + ce_2 \otimes a'e_1 + ce_2 \otimes c'e_2
\]
\[
= a'(e_1 \otimes e_1) + ac'(e_1 \otimes e_2) + ca'(e_2 \otimes e_1) + cc'(e_2 \otimes e_2).
\]

This shows that the coordinates of \(A \otimes B(e_1 \otimes e_1)\) in the basis \(\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}\) are
\[
\begin{pmatrix} aa' & ac' & ca' & cc' \end{pmatrix}^t.
\]
Likewise, we have

\[ A \otimes B(e_1 \otimes e_2) = (Ae_1) \otimes (Be_2) = (ae_1 + ce_2) \otimes (b'e_1 + d'e_2) = ab'(e_1 \otimes e_1) + ad'(e_1 \otimes e_2) + cb'(e_2 \otimes e_1) + cd'(e_2 \otimes e_2), \]

\[ A \otimes B(e_2 \otimes e_1) = (Ae_2) \otimes (Be_1) = (be_1 + de_2) \otimes (a'e_1 + c'e_2) = ba'(e_1 \otimes e_1) + bc'(e_1 \otimes e_2) + da'(e_2 \otimes e_1) + dc'(e_2 \otimes e_2), \]

\[ A \otimes B(e_2 \otimes e_2) = (Ae_2) \otimes (Be_2) = (be_1 + de_2) \otimes (b'e_1 + d'e_2) = bb'(e_1 \otimes e_1) + bd'(e_1 \otimes e_2) + db'(e_2 \otimes e_1) + dd'(e_2 \otimes e_2). \]

Hence, the matrix \( A \otimes B \) is given by

\[
\begin{pmatrix}
  aa' & ab' & ba' & bb' \\
  ac' & ad' & bc' & bd' \\
  ca' & cb' & da' & db' \\
  cc' & cd' & dc' & dd'
\end{pmatrix}.
\]

Upon closer inspection, we see that \( A \otimes B \) can be expressed as

\[
\begin{pmatrix}
  a & a' & b' & c' \\
  c' & b' & d' & c' \\
  c' & a' & d' & c' \\
  d' & a' & c' & d'
\end{pmatrix}
\begin{pmatrix}
  a' & b' \\
  c' & d'
\end{pmatrix}
\]

or in block form as

\[
\begin{pmatrix}
  aB & bB \\
  cB & dB
\end{pmatrix}.
\]

The previous example illustrates the Kronecker product. More generally, if \( \varphi \in \text{Hom}(\mathbb{F}^n, \mathbb{F}^m) \) has matrix representation \( A \in M_{m,n}(\mathbb{F}) \) and \( \rho \in \text{Hom}(\mathbb{F}^k, \mathbb{F}^l) \) has matrix representation \( B \in M_{l,k}(\mathbb{F}) \), then the matrix

\[
A \otimes B = \begin{pmatrix}
  a_{11}B & \cdots & a_{1n}B \\
  \vdots & \ddots & \vdots \\
  a_{m1}B & \cdots & a_{mn}B
\end{pmatrix}
\]

is called the Kronecker product of \( A \) and \( B \).
### 3.3 tensor algebras

Let $A$ be a commutative ring, written additively, and $M$ an $A$-module. First, define the tensor product of $n$ copies of $M$, denoted by $M^\otimes n$ or $T^n(M)$ or $T_A^n(M)$, if the underlying ring needs to be explicitly stated. In these notes, the typical notation will be $T^n(M)$. The tensor product $T^n(M)$ consists of elements of the form

$$x_1 \otimes x_2 \otimes \cdots \otimes x_n$$

where each $x_i$ belongs to $M$. Define a family of $A$-linear maps

$$m_{pq} : T^p(M) \otimes T^q(M) \to T^{p+q}(M)$$

by the product

$$(x_1 \otimes \cdots \otimes x_p) \cdot (x_{p+1} \otimes \cdots \otimes x_q) = x_1 \otimes \cdots \otimes x_p \otimes x_{p+1} \otimes \cdots x_q.$$

The family of multiplication maps defined by these products are associative. Define the tensor algebra of the module $M$, denoted by $T(M)$ or $T_A(M)$, as the direct sum $T(M) = \bigoplus_{n \geq 0} T^n(M)$ with the corresponding associative product. Given this definition, the tensor algebra of $M$ is a graded $A$-algebra of type $\mathbb{N}$. For each $n \geq 0$, the module $T^n(M)$ consists of the homogeneous elements of degree $n$. Naturally, we have the following correspondence between the tensor powers, the ring $A$, and the module $M$; $T^0(M) = A, T^1(M) = M$. Note that the additive unit, denoted by 0, belongs to $T^0(M) = A$. There is a canonical injection, often denoted by $\phi, \phi_M$, or $i_M$, mapping the module $M$ to the tensor algebra $T(M)$. As a consequence of the product on $T(M)$, the elements of $T^n(M)$ can be expressed as

$$x_1 \otimes x_2 \otimes \cdots \otimes x_n = \phi(x_1) \cdot \phi(x_2) \cdots \phi(x_n). \quad (3.12)$$

Hence, the elements of $T(M)$ are often expressed as $x_1 x_2 \cdots x_n$, by suppressing the tensor notation (and the dot notation).

The following proposition becomes quite useful when the exterior algebra and the universal enveloping algebra (of a Lie algebra) are introduced.

**Proposition 3.7.** Suppose $E$ is an $A$-algebra and $f$ is an $A$-linear map from $M$ to $E$. There exists a unique $A$-algebra homomorphism $g : T(M) \to E$ such that $f = g \circ \phi$.

![Diagram]

By the canonical injection $\phi$ and the relations (3.12), if $g$ exists, we have

$$g(x_1 \otimes \cdots \otimes x_n) = f(x_1) \cdots f(x_n). \quad (3.13)$$

55
4 Normed Spaces

For vector spaces, and various other spaces, it is quite useful to have some notion of distance. This section is devoted to reviewing the basics about normed spaces. A normed space (or normed linear space) is a vector space \( V \) on which there is defined a function \( \| \cdot \| : V \to \mathbb{F} \) called the norm such that the following hold:

a) \( \| x \| \geq 0 \) and \( \| x \| = 0 \) if and only if \( x = 0 \)

b) \( \| \alpha x \| = |\alpha| \| x \| \)

c) \( \| x + y \| \leq \| x \| + \| y \| \).

In this case, we often write \( (V, \| \cdot \|) \) is a normed space, unless it should be apparent which norm corresponds to \( V \).

Note that a normed space is the same as a metric space with the metric given by \( d(x, y) = \| x - y \| \).

Example 4.1.  1) We define a norm on the real vector space \( \mathbb{R}^n \). Let \( x = (x_1, x_2, \ldots, x_n)^t \in \mathbb{R}^n \), then

\[
\| x \|_2 = \left( \sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}.
\]  \hspace{1cm} (4.1)

2) More generally, for \( 1 \leq p \leq \infty \) we define the \( p \)-norms for \( x \in \mathbb{R}^n \) by

\[
\| x \|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad \text{for } 1 \leq p < \infty
\] \hspace{1cm} (4.2)

\[
\| x \|_\infty = \max_{1 \leq i \leq n} |x_i|
\] \hspace{1cm} (4.3)

As an exercise, you should show that \( \| \cdot \|_p \) is indeed a norm for \( p = 1, 2, \infty \). Note that when \( n = 1 \) all of the \( p \)-norms coincide with the absolute value. It can be shown that

\[
\| x \|_\infty = \lim_{p \to \infty} \| x \|_p.
\]

Example 4.2. For \( p \in [1, \infty] \), we define the sequence spaces \( \ell^p \) (the little l-\( p \) spaces) as

\[
\ell^p = \{ x = (x_n)_{n \geq 1} : \| x \|_{\ell^p} < \infty \}
\]
where the corresponding \( p \)-norms are given by

\[
\|x\|_\ell^p = \left( \sum_{n=1}^{\infty} |x_n|^p \right)^{1/p}
\]

for \( 1 \leq p < \infty \),

\[
\|x\|_\ell^\infty = \sup_{n \geq 1} |x_n|
\]

for \( p = \infty \).

The \( \ell^p \) spaces consist of all infinite sequences of complex numbers which are bounded with respect to the above norms.

We can define similar norms on function spaces. We must be careful, however, as some of these notions require technical details that we do not have time to cover in this course. If any of these details become necessary, we will cover them as they arise.

**Example 4.3.** For the space of continuous functions over the closed interval \([a, b]\), denoted \( C[a, b] \), we define the maximum norm by

\[
\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|, \quad f \in C[a, b].
\]

In addition to being able to discuss the relative size of a vector or the distance between two vectors, norms allow us to discuss the convergence of sequences.

**Definition 4.4.** Let \((V, \| \cdot \|)\) be a normed space. We say that a sequence \((x_n)\) in \( V \) converges to an element \( x \in V \), if for every \( \varepsilon > 0 \) there exists a number \( \delta(\varepsilon) \) such that for every \( n \geq \delta(\varepsilon) \) we have \( \|x_n - x\| < \varepsilon \). In this case, we may write \( \lim_{n \to \infty} x_n = x \) or \( x_n \to x \).

**Definition 4.5.** Let \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) be norms on a vector space \( V \). We say that the norms \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) are equivalent if there exist positive constants \( \alpha, \beta \) such that

\[
\alpha \|x\|_1 \leq \|x\|_2 \leq \beta \|x\|_1
\]

for all \( x \in V \).

An important fact, which we may or may not prove, depending on time is the following.

**Claim.** Let \( V \) be a finite dimensional vector space. Then any two norms \( \| \cdot \|_1, \| \cdot \|_2 \) on \( V \) are equivalent.

This statement is not true for an infinite dimensional space. Norms also allow us to define the concepts of continuity, open/closed sets, extreme values and a host of other notions we have from calculus/real analysis.
Exercise 4.1. Define a normed space as the vector space $V = C[0,1]$ with the norm defined by

$$
\|v\|_p = \left( \int_0^1 |v(x)|^p \, dx \right)^{1/p}, \quad 1 \leq p < \infty
$$

$$
\|v\|_\infty = \sup_{0 \leq x \leq 1} |v(x)|.
$$

Consider the sequence

$$
u_n(x) = \begin{cases} 
1 - nx, & 0 \leq x \leq \frac{1}{n} \\
0, & \frac{1}{n} < x \leq 1
\end{cases}
$$

Show that

$$
\|u_n\|_p = (n(p+1))^{-1/p}, \quad 1 \leq p < \infty,
$$

however

$$
\|u_n\|_\infty = 1.
$$

Thus, the sequence $\{u_n\}$ converges in the normed spaces $(V, \|\cdot\|_p)$ for $1 \leq p < \infty$, but diverges in the normed space $(V, \|\cdot\|_\infty)$.

Exercise 4.2. Show that the following two norms are equivalent on $C[0,1]$: 

$$
\|f\|_a = |f(a)| + \int_0^1 |f'(x)| \, dx
$$

$$
\|f\|_b = \int_0^1 |f(x)| \, dx + \int_0^1 |f'(x)| \, dx.
$$

4.1 Operators on normed spaces

Given two spaces $V, W$, an operator $L$ from $V \rightarrow W$ is a mapping which assigns to each element of $V$ a unique element in $W$. The domain of $L$, $\mathcal{D}(L)$, is the following subset of $V$

$$
\mathcal{D}(L) = \{ v \in V : L(v) \text{ exists} \},
$$

while the range of $L$ is the following subset of $W$

$$
\mathcal{R}(L) = \{ w \in W : w = L(v) \text{ for some } v \in V \}.
$$

Another important set related to an operator, is the nullspace

$$
\mathcal{N}(L) = \{ v \in V : L(v) = 0 \}.
$$

Given two operators $L$ and $M$ from $V \rightarrow W$, we define the addition of operators by

$$(L + M)(v) = L(v) + M(v).$$
This defines a new operator, $L + M$, with $\mathcal{D}(L + M) = \mathcal{D}(L) \cap \mathcal{D}(M)$. Likewise, for $\alpha \in \mathbb{K}$ we define scalar multiplication of an operator by

$$(\alpha L)(v) = \alpha L(v).$$

Again, this defines a new operator $\alpha L : V \to W$ with domain $\mathcal{D}(L)$.

**Definition 4.6.** An operator $L : V \to W$ is one-to-one (or 1-1) if

$$L(v_1) = L(v_2) \to v_1 = v_2, \quad v_1, v_2 \in \mathcal{D}(L).$$

An operator which is 1-1 is also said to be injective. If $\mathcal{R}(L) = W$, we say $L$ is surjective. Furthermore, if $L$ is both injective and surjective, we say $L$ is bijective.

**Example 4.7.** Let $V = \mathcal{C}[0,1] = W$ be the vector space of continuous functions on $[0,1]$. Define the differentiation operator by

$$\frac{d}{dx}f = f', \quad f \in \mathcal{C}[0,1].$$

In this case, we have $\mathcal{D}(\frac{d}{dx}) = \mathcal{C}^1[0,1] \subset \mathcal{C}[0,1]$. Is $\frac{d}{dx}$ an injection and/or surjection?

For operators on normed spaces, properties such as continuity and boundedness are defined in terms of vectors and sequences of vectors in the domain of the operator.

**Definition 4.8.** Let $V,W$ be two normed spaces. An operator $L : V \to W$ is continuous at $v \in \mathcal{D}(L) \subset V$ if for a sequence $\{v_n\} \in \mathcal{D}(L)$ we have

$$v_n \to v \text{ in } V \Rightarrow L(v_n) \to L(v) \text{ in } W.$$

We say that $L$ is continuous if it is continuous for all $v \in \mathcal{D}(L)$. The operator $L$ is said to be bounded if for any $\gamma > 0$, there is a $R > 0$ such that for $v \in \mathcal{D}(L)$ we have

$$\|v\| \leq \gamma \Rightarrow \|L(v)\| \leq R.$$

**Exercise 4.3.** Let $V = \mathcal{C}[0,1]$ be the space of complex-valued continuous functions defined on the closed interval $[0,1]$. Define an operator $L$ by

$$L(v)(t) = v(0) + \int_0^t v(s)ds.$$ 

What is the domain of $L$? How about the range? Show that the function $v(t) = ae^t$ is a fixed point of $L$ for any $a \in \mathbb{C}$. Analogous to functions, a fixed point of an operator $L$ is an element of $\mathcal{D}(L)$ such that $L(v) = v$. 

59
**Definition 4.9.** If $L : V \to W$ satisfies the following properties:

$$L(\alpha v_1 + \beta v_2) = \alpha L(v_1) + \beta L(v_2) \quad \text{for all } v_1, v_2 \in V, \alpha, \beta \in K$$

we say that $L$ is *linear*. We denote the set of all linear operators mapping $V$ to $W$ by $\mathcal{L}(V, W)$ and by $\mathcal{L}(V)$ whenever $V = W$. In this case, we often use the notation $Lv$ in place of $L(v)$.

**Proposition 4.10.** Suppose $L$ is a linear operator from a vector space $V$ to a vector space $W$. Then

a) the continuity of $L$ over the whole space $V$ is equivalent to continuity at any one point (usually taken to be $v = 0$.)

b) $L$ is bounded if and only if there exists a constant $\alpha \geq 0$ such that

$$\|Lv\|_W \leq \alpha \|v\|_V \quad \text{for all } v \in V.$$ 

c) $L$ is continuous on $V$ if and only if $L$ is bounded on $V$.

As a consequence of Proposition 4.10, we define a norm on $\mathcal{L}(V, W)$ by

$$\|L\|_{V,W} = \sup_{v \neq 0} \frac{\|Lv\|_W}{\|v\|_V}. \quad (4.6)$$

This norm is usually referred to as the *operator norm*. The following theorem is obtained as a consequence.

**Theorem 4.11.** The set $\mathcal{L}(V, W)$ with the norm (4.6) is a normed space.

**Proof.** Exercise. ☐

In fact, as we will discover soon, under certain conditions the space $\mathcal{L}(V, W)$ is a Banach space. The following property of the operator norm can sometimes be useful

$$\|Lv\|_W \leq \|L\|_{V,W} \|v\|_V \quad \text{for all } v \in V.$$ 

**Example 4.12.** For a normed space $V$, the identity operator $I : V \to V$, defined by

$$Iv = v,$$

is an element of $\mathcal{L}(V)$ and $\|I\| = 1$. 

---

*CR Humber Intermediate Linear Algebra Lecture Notes Spring 2015*

---

60