' Abstract Vector Spaces, Linear Transformations, and Their Coordinate Representations

Contents

1 Vector Spaces

1.1 Definitions

1.1.1 Basics

A **vector space (linear space)** V **over a field** F is a set V on which the operations **addition**, $+ : V \times V \to V$, and **left** F**-action** or **scalar multiplication**, $\cdot : F \times V \to V$, satisfy: for all $x, y, z \in V$ and $a, b, 1 \in F$

VS1 $x + y = y + x$
VS2 $(x + y) + z = x + (y + z)$
VS3 $\exists 0 \in V$ such that $x + 0 = x$
VS4 $\forall x \in V,\ \exists y \in V$ such that $x + y = 0$
(VS1–VS4: $(V, +)$ *is an abelian group*)
VS5 $1x = x$
VS6 $(ab)x = a(bx)$
VS7 $a(x + y) = ax + ay$
VS8 $(a + b)x = ax + ay$

If V is a vector space over F, then a subset $W \subseteq V$ is called a **subspace** of V if W is a vector space over the same field F and with addition and scalar multiplication $+|_{W\times W}$ and $\cdot|_{F\times W}$.

1.1.2 Linear Combinations, Spans, Bases, and Dimensions

If $\varnothing \neq S \subseteq V$, $v_1, \ldots, v_n \in S$ and $a_1, \ldots, a_n \in F$, then a **linear combination** of $v_1, \ldots, v_n$ is the *finite* sum

$$a_1 v_1 + \cdots + a_n v_n \tag{1.1}$$

which is a vector in V. The $a_i \in F$ are called the **coefficients** of the linear combination. If $a_1 = \cdots = a_n = 0$, then the linear combination is said to be **trivial**. In particular, considering the special case of 0 in V, the zero vector, we note that 0 may always be represented as a linear combination of any vectors $u_1, \ldots, u_n \in V$,

$$\underbrace{0u_1 + \cdots + 0u_n}_{0 \in F} = \underbrace{0}_{0 \in V}$$

This representation is called the **trivial representation of** 0 by $u_1, \ldots, u_n$. If, on the other hand, there are vectors $u_1, \ldots, u_n \in V$ and scalars $a_1, \ldots, a_n \in F$ such that

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

where at least one $a_i \neq 0 \in F$, then that linear combination is called a **nontrivial representation of** 0. Using linear combinations we can generate subspaces, as follows. If $S \neq \varnothing$ is a subset of V, then the **span** of S is given by

$$\operatorname{span}(S) := \{v \in V \mid v \text{ is a linear combination of vectors in } S\} \tag{1.2}$$

The span of $\varnothing$ is by definition

$$\operatorname{span}(\varnothing) := \{0\} \tag{1.3}$$

In this case, $S \subseteq V$ is said to **generate** or **span** V, and to be a **generating** or **spanning set**. If $\operatorname{span}(S) = V$, then S said to be a generating set or a spanning set for V.

A nonempty subset S of a vector space V is said to be **linearly independent** if, taking any finite number of distinct vectors $u_1, \ldots, u_n \in S$, we have for all $a_1, \ldots, a_n \in F$ that

$$a_1 u_1 + a_2 u_2 + \cdots + a_n u_n = 0 \quad \Longrightarrow \quad a_1 = \cdots = a_n = 0$$

That is S is linearly independent if the *only* representation of $0 \in V$ by vectors in S is the trivial one. In this case, the vectors $u_1, \ldots, u_n$ themselves are also said to be linearly independent. Otherwise, if there is at least one nontrivial representation of 0 by vectors in S, then S is said to be **linearly dependent**.

Given $\varnothing \neq S \subseteq V$, a nonzero vector $v \in S$ is said to be an **essentially unique linear combination** of the vectors in S if, up to order of terms, there is one and only one way to express v as a linear combination of $u_1, \ldots, u_n \in S$. That is, if there are $a_1, \ldots, a_n, b_1, \ldots, b_m \in F \backslash \{0\}$ and distinct $u_1, \ldots, u_n \in S$ and distinct $v_1, \ldots, v_m \in S$ distinct, then, re-indexing the b_is if necessary,

$$\left.\begin{aligned} v &= a_1 u_1 + \cdots + a_n u_n \\ &= b_1 v_1 + \cdots + b_m v_m \end{aligned}\right\} \quad \Longrightarrow \quad m = n \text{ and } \left\{\begin{aligned} a_i &= b_i \\ u_i &= v_i \end{aligned}\right\} \quad \forall i = 1, \ldots, n$$

A subset $\beta \subseteq V$ is called a (**Hamel**) **basis** if it is linearly independent and $\operatorname{span}(\beta) = V$. We also say that the vectors of β form a basis for V. Equivalently, as explained in Theorem 1.13 below, β is a basis if every nonzero vector $v \in V$ is an essentially unique linear combination of vectors in β. In the context of inner product spaces of inifinite dimension, there is a difference between a vector space basis, the Hamel basis of V, and an orthonormal basis for V, the Hilbert basis for V, because though the two always exist, they are not always equal unless $\dim(V) < \infty$.

The **dimension** of a vector space V is the cardinality of any basis for V, and is denoted $\dim(V)$. V **finite-dimensional** if it is the zero vector space $\{0\}$ or if it has a basis of finite cardinality. Otherwise, if it's basis has infinite cardinality, it is called **infinite-dimensional**. In the former case, $\dim(V) = |\beta| = n < \infty$ for some $n \in \mathbb{N}$, and V is said to be n-**dimensional**, while in the latter case, $\dim(V) = |\beta| = \kappa$, where κ is a cardinal number, and V is said to be κ-**dimensional**.

If V is finite-dimensional, say of dimension n, then an **ordered basis** for V a finite sequence or n-tuple $(v_1, \ldots, v_n)$ of linearly independent vectors $v_1, \ldots, v_n \in V$ such that $\{v_1, \ldots, v_n\}$ is a basis for V. If V is infinite-dimensional but with a countable basis, then an ordered basis is a sequence $(v_n)_{n\in\mathbb{N}}$ such that the set $\{v_n \mid n \in \mathbb{N}\}$ is a basis for V. The term **standard ordered basis** is applied only to the ordered bases for F^n, $(F^\omega)_0$, $(F^B)_0$ and $F[x]$ and its subsets $F_n[x]$. For F^n, the standard ordered basis is given by

$$(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots,$ $\mathbf{e}_n = (0, \ldots, 0, 1)$. Since F is *any* field, 1 represents the multiplicative identity and 0 the additive identity. For $F = \mathbb{R}$, the standard ordered basis is just as above with the usual 0 and 1. If $F = \mathbb{C}$, then $1 = (1, 0)$ and $0 = (0, 0)$, so the ith standard basis vector for $\mathbb{C}^n$ looks like this:

$$\mathbf{e}_i = \Big((0,0), \ldots, \overset{i\text{th term}}{(1,0)}, \ldots, (0,0)\Big)$$

For $(F^\omega)_0$ the standard ordered basis is given by

$$(\mathbf{e}_n)_{n\in\mathbb{N}}, \qquad \text{where} \ \ \mathbf{e}_n(k) = \delta_{nk}$$

and generally for $(F^B)_0$ the standard ordered basis is given by:

$$(\mathbf{e}_b)_{b\in B}, \qquad \text{where} \ \ \mathbf{e}_b(x) = \begin{cases} 1, & \text{if } x = b \\ 0, & \text{if } x \neq b \end{cases}$$

When the term standard ordered basis is applied to the vector space $F_n[x]$, the subspace of $F[x]$ of polynomials of degree less than or equal to n, it refers the basis

$$(1, x, x^2, \ldots, x^n)$$

For $F[x]$, the standard ordered basis is given by

$$(1, x, x^2, \ldots)$$

1.1.3 Sums and Products of Vector Spaces and Subspaces

If $S_1, \ldots, S_k \in P(V)$ are subsets or subspaces of a vector space V, then their **(internal) sum** is given by

$$S_1 + \cdots + S_k = \sum_{i=1}^{k} S_i := \{v_1 + \cdots + v_k \mid v_i \in S_i \text{ for } 1 \leq i \leq k\} \tag{1.4}$$

More generally, if $\{S_i \mid i \in I\} \subseteq P(V)$ is any collection of subsets or subspaces of V, then their **sum** is defined to be

$$\sum_{i\in I} S_i = \Big\{v_1 + \cdots + v_n \mid s_i \in \bigcup_{i\in I} S_i \text{ and } n \in \mathbb{N}\Big\} \tag{1.5}$$

If S is a subspace of V and $v \in V$, then the sum $\{v\} + S = \{v + s \mid s \in S\}$ is called a **coset** of S and is usually denoted

$$v + S \tag{1.6}$$

We are of course interested in the **(internal) sum of *subspaces*** of a vector space. In this case, we may have the special condition

$$V=\sum_{i=1}^{k}S_i \quad \text{and} \quad S_j\cap\sum_{i\neq j}S_i=\{0\},\ \forall j=1,\dots,k \tag{1.7}$$

When this happens, V is said to be the **(internal) *direct* sum** of $S_1,\dots,S_k$, and this is symbolically denoted

$$V=S_1\oplus S_2\oplus\cdots\oplus S_k=\bigoplus_{i=1}^{k}S_i=\{s_1+\cdots+s_k\mid s_i\in S_i\} \tag{1.8}$$

In the infinite-dimensional case we proceed similarly. If $\mathcal{F}=\{S_i\mid i\in I\}$ is a family of subsets of V such that

$$V=\sum_{i\in I}S_i \quad \text{and} \quad S_j\cap\sum_{i\neq j}S_i=\{0\},\ \forall j\in I \tag{1.9}$$

then V is a **direct sum** of the $S_i\in\mathcal{F}$, denoted

$$V \;=\; \bigoplus\mathcal{F} \;=\; \bigoplus_{i\in I}S_i \;=\; \Big\{s_1+\cdots+s_n\;\Big|\; s_j\in\bigcup_{i\in I}S_i,\ \text{and } s_{j_i}\in S_i \text{ if } s_{j_i}\neq 0\Big\} \tag{1.10}$$

We can also define the **(external) sum of distinct vector spaces** which do not lie inside a larger vector space: if $V_1,\dots,V_n$ are vector spaces over the same field F, then their external direct sum is the cartesian product $V_1\times\cdots\times V_n$, with addition and scalar multiplication defined componentwise. And we denote the sum, confusingly, by the same notation:

$$V_1\oplus V_2\oplus\cdots\oplus V_n:=\{(v_1,v_2,\dots,v_n)\mid v_i\in V_i,\ i=1,\dots,n\} \tag{1.11}$$

In the infinite-dimensional case, we have two types of external direct sum, one where there is no restriction on the sequences, the other where we only allow sequences with finite support. To distinguish these two cases we use the terms *direct product* and *external direct product*, respectively: let $\mathcal{F}=\{V_i\mid i\in I\}$ be a family of vector spaces over the same field F. Then their **direct product** is given by

$$\prod_{i\in I}V_i \;:=\; \Big\{\mathbf{v}=(v_i)_{i\in I}\;\Big|\; v_i\in V_i\Big\} \tag{1.12}$$

$$\;=\; \Big\{f:I\to\bigcup_{i\in I}V_i\;\Big|\; f(i)\in V_i\Big\} \tag{1.13}$$

The infinite-dimensional **external direct sum** is then defined:

$$\bigoplus_{i\in I}V_i \;:=\; \Big\{\mathbf{v}=(v_i)_{i\in I}\;\Big|\; v_i\in V_i \text{ and } \mathbf{v} \text{ has finite support}\Big\} \tag{1.14}$$

$$\;=\; \Big\{f:I\to\bigcup_{i\in I}V_i\;\Big|\; f(i)\in V_i \text{ and } f \text{ has finite support}\Big\} \tag{1.15}$$

When all the V_i are equal, there is some special notation:

$$V^n \quad (\textit{finite-dimensional}) \tag{1.16}$$

$$V^I:=\prod_{i\in I}V \quad (\textit{infinite-dimensional}) \tag{1.17}$$

$$(V^I)_0:=\bigoplus_{i\in I}V \tag{1.18}$$

Example 1.1 *Common examples of vector spaces are the* **sequence space** F^n, F^ω, F^B, $(F^\omega)_0$ *and* $(F^B)_0$ *in a field* F *over that field, i.e. the various types of cartesian products of* F *equipped with addition and scalar multiplication operations defined componentwise (*$\omega = \mathbb{N}$ *and* B *is any set, and where* $(F^\omega)_0$ *and* $(F^B)_0$ *denote all functions* $f : \omega \to F$*, respectively* $f : B \to F$*, with finite support). Some typical ones are*

$$\begin{array}{lll} \mathbb{Q}^n & \mathbb{R}^n & \mathbb{C}^n \\ \mathbb{Q}^\omega & \mathbb{R}^\omega & \mathbb{C}^\omega \\ \mathbb{Q}^B & \mathbb{R}^B & \mathbb{C}^B \\ (\mathbb{Q}^\omega)_0 & (\mathbb{R}^\omega)_0 & (\mathbb{C}^\omega)_0 \\ (\mathbb{Q}^B)_0 & (\mathbb{R}^B)_0 & (\mathbb{C}^B)_0 \end{array}$$

each over $\mathbb{Q}$ *or* $\mathbb{R}$ *or* $\mathbb{C}$*, respectively.* ■

1.2 Basic Vector Space Theory

Theorem 1.2 *If V is a vector space over a field F, then for all $x, y, z \in V$ and $a, b \in F$ the following hold:*

(1) $x + y = z + y \implies x = z$ (*Cancellation Law*)

(2) $0 \in V$ *is unique.*

(3) $x + y = x + z = 0 \implies y = z$ (*Uniqueness of the Additive Inverse*)

(4) $0x = 0$

(5) $(-a)x = -(ax) = a(-x)$

(6) *The finite external direct sum $V_1 \oplus \cdots \oplus V_n$ is a vector space over F if each V_i is a vector space over F and if we define $+$ and $\cdot$ componentwise by*

$$\begin{aligned}(v_1, \ldots, v_n) + (u_1, \ldots, u_n) &:= (v_1 + u_1, \ldots, v_n + u_n)\\ c(v_1, \ldots, v_n) &:= (cv_1, \ldots, cv_n)\end{aligned}$$

More generally, if $\mathcal{F} = \{V_i \mid i \in I\}$ is a family of vector spaces over the same field F, then the direct product $\prod_{i \in I} V_i$ and external direct sum $\bigoplus_{i \in I} V_i = (V^I)_0$ are vector spaces over F, with $+$ and $\cdot$ given by

$$\begin{aligned}v + w &= \Big(f : I \to \bigcup_{i \in I} V_i\Big) + \Big(g : I \to \bigcup_{i \in I} V_i\Big) := \Big(f + g : I \to \bigcup_{i \in I} V_i\Big)\\ cv &= c\Big(f : I \to \bigcup_{i \in I} V_i\Big) := \Big(cf : I \to \bigcup_{i \in I} V_i\Big)\end{aligned}$$

Proof: (1) $x = x+0 = x+(y+(-y)) = (x+y)+(-y) = (z+y)+(-y) = z+(y+(-y)) = z+0 = z$. (2) $\exists 0_1,\ 0_2 \implies 0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2$. (3)(a) $x + y_1 = x + y_2 = 0 \implies y_1 = y_2$ by the Cancellation Law. (b) $y_1 = y_1 + 0 = y_1 + (x + y_2) = (y_1 + x) + y_2 = 0 + y_2 = y_2 + 0 = y_2$. (4) $0x + 0x = (0 + 0)x = 0x = 0x + 0_V \implies 0x = 0_V$ by the Cancellation Law. (5) $ax + (-(ax)) = 0$, and by 3 $-(ax)$ is unique. But note $ax + (-a)x = (a + (-a))x = 0x = 0$, hence $(-a)x = -(ax)$, and also $a(-x) = a((-1)x) = (a(-1))x = (-a)x$, which we already know equals $-(ax)$. (6) All 8 vector space properties are verified componentwise, since that is how addition and scalar multiplication are defined. But the components are vectors in a vector space, so all 8 properties will be satisfied. ■

Theorem 1.3 *Let V be a vector space over F. A subset S of V is a subspace iff for all $a, b \in F$ and $x, y \in S$ we have one of the following equivalent conditions:*

(1) $ax + by \in S$

(2) $x + y \in S$ *and* $ax \in S$

(3) $x + ay \in S$

Proof: If S satisfies any of the above implications, then it is a subspace because vector space properties VS1-2 and VS5-8 follow from the fact that $S \subseteq V$, while VS3 follows because we can choose $a = b = 0$ and $0 \in V$ is unique (Theorem 1.2), and VS4 follows from VS3, in the first case by choosing $b = -a$ and $x = y$, in the second case by choosing $y = (-1)x$, and in the third case by choosing $a = -1$ and $y = x$. Conversely, if S is a vector space, then (1)-(3) follow by definition. ■

Theorem 1.4 *Let V be a vector space. If $\mathcal{C} = \{S_i \mid i \in K\}$ is a collection of subspaces of V, then*

$$\sum_{i\in K} S_i, \qquad \bigoplus_{i\in K} S_i \qquad \text{and} \qquad \bigcap_{i\in K} S_i \tag{1.19}$$

are subspaces.

Proof: (1) If $\mathcal{C} = \{S_i \mid i \in K\}$ is a collection of subspaces, then none of the S_i are empty by the previous theorem, so if $a, b \in F$ and $v, w \in \sum_{i\in K} S_i$, then $v = s_{i_1} + \cdots + s_{i_n} \in \bigcup_{i\in K} S_i$ and $w = s_{j_1} + \cdots + s_{j_m} \in \bigcup_{i\in K} S_i$ such that $s_{i_k} \in S_{i_k}$ for all $i_k \in K$. Consequently

$$\begin{aligned} av + bw &= a(s_{i_1} + \cdots + s_{i_n}) + b(s_{j_1} + \cdots + s_{j_n}) \\ &= as_{i_1} + \cdots + as_{i_n} + bs_{j_1} + \cdots + bs_{j_n} \\ &\in \sum_{i\in K} S_i \end{aligned}$$

because $as_{i_k} \in S_{i_k}$ for all $i_k \in K$ because the S_{i_k} are subspaces. Hence by the previous theorem $\sum_{i\in K} S_i$ is a subspace of V. The direct sum case is a subcase of this general case.

(2) If $a, b \in F$ and $u, v \in \bigcap_{i\in K} S_i$, then $u, v \in S_i$ for all $i \in K$, whence $au + bv \in S_i$ for all $i \in K$, or $au + bv \in \bigcap_{i\in K} S_i$, and $\bigcap_{i\in K} S_i$ is a subspace by the previous theorem. ■

Theorem 1.5 *If S and T are subspaces of a vector space V, then $S \cup T$ is a subspace iff $S \subseteq T$ or $T \subseteq S$.*

Proof: If $S \cup T$ is a subspace, then for all $a, b \in F$ and $x, y \in S \cup T$ we have $ax + by \in S \cup T$. Suppose neither $S \subseteq T$ nor $S \subseteq T$ holds, and take $x \in S \backslash T$ and $y \in T \backslash S$. Then, $x + y$ is neither in S nor in T (for if it were in S, say, then $y = (x + y) + (-x) \in S$, a contradiction), contradicting the fact that $S \cup T$ is a subspace. Conversely, suppose, say, $S \subseteq T$, and choose any $a, b \in F$ and $x, y \in S \cup T$. If $x, y \in S$, then $ax + by \in S \subseteq S \cup T$, while if $x, y \in T$, then $ax + by \in T \subseteq S \cup T$. If $x \in S$ and $y \in T$, then $x \in T$ as well and $ax + by \in T \subseteq S \cup T$. Thus in all cases $ax + by \in S \cup T$, and $S \cup T$ is a subspace. ■

Theorem 1.6 *If V is a vector space, then for all $S \subseteq V$, $\operatorname{span}(S)$ is a subspace of V. Moreover, if T is a subspace of V and $S \subseteq T$, then $\operatorname{span}(S) \subseteq T$.*

Proof: If $S = \varnothing$, $\operatorname{span}(\varnothing) = \{0\}$, which is a subspace of V; if T is a subspace, then $\varnothing \subseteq T$ and $\{0\} \subseteq T$. If $S \neq \varnothing$, the any $x, y \in S$ have the form $x = a_1v_1 + \cdots + a_nv_n$ and $y = b_1w_1 + \cdots + b_mw_m$ for some $a_1, \ldots a_n, b_1, \ldots, b_n \in F$ and $v_1, \ldots, v_n, w_1, \ldots, w_n \in S$. Then, if $a, b \in F$ we have

$$\begin{aligned} ax + by &= a(a_1v_1 + \cdots + a_nv_n) + b(b_1w_1 + \cdots + b_mw_m) \\ &= (aa_1)v_1 + \cdots + (aa_n)v_n + (bb_1)w_1 + \cdots + (bb_m)w_m \\ &\in \operatorname{span}(S) \end{aligned}$$

so $\operatorname{span}(S)$ is a subspace of V. Now, suppose T is a subspace of V and $S \subseteq T$. If $w \in \operatorname{span}(S)$, then $w = a_1w_1 + \cdots + a_kw_n$ for some $w_1, \ldots, w_n \in S$ and $a_1, \ldots, a_n \in F$. Since $S \subseteq T$, $w_1, \ldots, w_k \in T$, whence $w = c_1w_1 + \cdots + c_kw_k \in T$. ■

Corollary 1.7 *If V is a vector space, then $S \subseteq V$ is a subspace of V iff $\operatorname{span}(S) = S$.*

Proof: If S is a subspace of V, then by the theorem $S \subseteq S \implies \operatorname{span}(S) \subseteq S$. But if $s \in S$, then $s = 1s \in \operatorname{span}(S)$, so $S \subseteq \operatorname{span}(S)$, and therefore $S = \operatorname{span}(S)$. Conversely, if $S = \operatorname{span}(S)$, then S is a subspace of V by the theorem. ■

Corollary 1.8 *If V is a vector space and $S, T \subseteq V$, then the following hold:*

(1) $S \subseteq T \subseteq V \implies \text{span}(S) \subseteq \text{span}(T)$

(2) $S \subseteq T \subseteq V$ *and* $\text{span}(S) = V \implies \text{span}(T) = V$

(3) $\text{span}(S \cup T) = \text{span}(S) + \text{span}(T)$

(4) $\text{span}(S \cap T) \subseteq \text{span}(S) \cap \text{span}(T)$

Proof: (1) and (2) are immediate, so we only need to prove 3 and 4:

(3) If $v \in \text{span}(S \cup T)$, then $\exists v_1, \ldots, v_m \in S$, $u_1, \ldots, u_n \in T$ and $a_1, \ldots, a_m, b_1, \ldots, b_n \in F$ such that

$$v = v_1 a_1 + \cdots + v_m a_m + b_1 u_1 + \cdots + b_n u_n$$

Note, however, $v_1 a_1 + \cdots + v_m a_m \in \text{span}(S)$ and $b_1 u_1 + \cdots + b_n u_n \in \text{span}(T)$, so that $v \in \text{span}(S) + \text{span}(T)$. Thus $\text{span}(S \cup T) \subseteq \text{span}(S) + \text{span}(T)$. Conversely, if $v = s + t \in \text{span}(S) + \text{span}(T)$, then $s \in \text{span}(S)$ and $t \in \text{span}(T)$, so that by 1, since $S \subseteq S \cup T$ and $T \subseteq S \cup T$, we must have $\text{span}(S) \subseteq \text{span}(S \cup T)$ and $\text{span}(T) \subseteq \text{span}(S \cup T)$. Consequently, $s, t \in \text{span}(S \cup T)$, and since $\text{span}(S \cup T)$ is a subspace, $v = s + t \in \text{span}(S \cup T)$. That is $\text{span}(S) + \text{span}(T) \subseteq \text{span}(S \cup T)$. Thus, $\text{span}(S) + \text{span}(T) = \text{span}(S \cup T)$.

(4) By Theorem 1.6 $\text{span}(S \cap T)$, $\text{span}(S)$ and $\text{span}(T)$ are subspaces, and by Theorem 1.4 $\text{span}(S) \cap \text{span}(T)$ is a subspace. Now, consider $x \in \text{span}(S \cap T)$. There exist vectors $v_1, \ldots, v_n \in S \cap T$ and scalars $a_1, \ldots, a_n \in F$ such that $x = a_1 v_1 + \cdots + a_n v_n$. But since $v_1, \ldots, v_n$ belong to both S and T, $x \in \text{span}(S)$ and $x \in \text{span}(T)$, so that $x \in \text{span}(S) \cap \text{span}(T)$. It is not in general true, however, that $\text{span}(S) \cap \text{span}(T) \subseteq \text{span}(S \cap T)$. For example, if $S = \{\mathbf{e}_1, \mathbf{e}_2\} \subseteq \mathbb{R}^2$ and $T = \{\mathbf{e}_1, (1,1)\}$, then $\text{span}(S) \cap \text{span}(T) = \mathbb{R}^2 \cap \mathbb{R}^2 = \mathbb{R}^2$, but $\text{span}(S \cap T) = \text{span}(\{e_1\}) = \mathbb{R}$, and $\mathbb{R}^2 \not\subseteq \mathbb{R}$. ■

Theorem 1.9 *If V is a vector space and $S \subseteq T \subseteq V$, then*

(1) *S linearly dependent $\implies$ T linearly dependent*

(2) *T linearly independent $\implies$ S linearly independent*

Proof: (1) If S is linearly dependent, there are scalars $a_1, \ldots, a_m \in F$, not all zero, and vectors $v_1, \ldots, v_m \in S$ such that $a_1 v_1 + \cdots + a_m v_m = 0$. If $S \subseteq T$, then $v_1, \ldots, v_m$ are also in T, and have the same property, namely $a_1 v_1 + \cdots + a_m v_m = 0$. Thus, T is linearly dependent. (2) If T is linearly independent, then if $v_1, \ldots, v_n \in S$, $a_1, \ldots, a_n \in F$ and $a_1 v_1 + \cdots + a_n v_n = 0$, since $S \subseteq T$ implies $v_1, \ldots, v_n \in T$, we must have $a_1 = a_2 = \cdots = a_n = 0$, and S is linearly independent. ■

Theorem 1.10 *If V is a vector space, then $S \subseteq V$ is linearly dependent iff $S = \{0\}$ or there exist distinct vectors $v, u_1, \ldots, u_n \in S$ such that v is a linear combination of $u_1, \ldots, u_n$.*

Proof: If S is linearly dependent, then S consists of at least one element by definition. Suppose S has only one element, v. Then, either $v = 0$ or $v \neq 0$. If $v \neq 0$, consider $a \in F$. Then, $av = 0 \implies a = 0$, which means S is linearly independent, contrary to assumption. Therefore, $v = 0$, and so $S = \{0\}$. If S contains more than one element and $v \in S$, then since S is linearly dependent, there exist distinct $u_1, \ldots, u_n \in S$ and $a_1, \ldots, a_n \in F$, not all zero such that

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

If $v = 0$, we're done. If $v \neq 0$, then, since $\{u_1, \ldots, u_n\}$ is linearly dependent, so is $\{u_1, \ldots, u_n, v\}$, by Theorem 1.9, so that there are $b_1, \ldots, b_{n+1} \in F$, not all zero, such that

$$b_1 u_1 + \cdots + b_n u_n + b_{n+1} v = 0$$

Choose $b_{n+1} \neq 0$, for if we let $b_{n+1} = 0$ we have gained nothing. Then,

$$v = \left(-\frac{b_1}{b_{n+1}}\right) u_1 + \cdots + \left(-\frac{b_n}{b_{n+1}}\right) u_n$$

and v is a linear combination of distinct vectors in S.

Conversely, if $S = \{0\}$, clearly S is linearly dependent, for we can choose $a, b \in F$ both nonzero and have $a0 + b0 = 0$. Otherwise, if there are distinct $u_1, \ldots, u_n, v \in S$ and $a_1, \ldots, a_n \in F$ such that $v = a_1 u_1 + \cdots + a_n u_n$, then obviously $0 = -1v + a_1 u_1 + \cdots + a_n u_n$, and since $-1 \neq 0$, S is linearly dependent. ■

Corollary 1.11 *If $S = \{u_1, \ldots, u_n\} \subseteq V$, then S is linearly dependent iff $u_1 = 0$ or for some $1 \leq k < n$ we have $u_{k+1} \in \operatorname{span}(u_1, \ldots, u_k)$.* ■

Theorem 1.12 *If V is a vector space and $u, v \in V$ are distinct, then $\{u, v\}$ is linearly dependent iff u or v is a multiple of the other.*

Proof: If $\{u, v\}$ is linearly dependent, there are scalars $a, b \in F$ not both zero such that $au + bv = 0$. If $a = 0$, then $b \neq 0$, and so $au + bv = bv = 0 \implies v = 0 \implies u \neq 0$, and so $v = 0u$. If neither a nor b is equal to zero, then $au + bv = 0 \implies u = -a^{-1}bv$. Conversely, suppose, say, $u = av$. If $a = 0$, then $u = 0$ and v can be any vector other than v, since by hypothesis $u \neq v$, whence $au + 0v = 0$ for any $a \in F$, including nonzero a, so that $\{u, v\}$ is linearly dependent. Otherwise, if $a \neq 0$, we have $u = av \implies 1u + (-a)v = 0$, and both $1 \neq 0$ and $-a \neq 0$, so that $\{u, v\}$ is linearly dependent. ■

Theorem 1.13 (Properties of a Basis) *If V is a vector space, then the following are equivalent, and any $S \subseteq V$ satisfying any of them is a basis:*

(1) *S is linearly dependent and $V = \operatorname{span}(S)$.*

(2) **(Essentially Unique Representation)** *Ever nonzero vector $v \in V$ is an essentially unique linear combination of vectors in S.*

(3) *No vector in S is a linear combination of other vectors in S.*

(4) *S is a* **minimal spanning set***, that is $V = \operatorname{span}(S)$ but no proper subset of S spans V.*

(5) *S is a* **maximal linearly independent set***, that is S is linearly independent but any proper superset is not linearly independent.*

Proof: (1)$\implies$ (2): If S is linearly independent and $v \in S$ is a nonzero vector such that for distinct s_i and t_i in S

$$\begin{aligned} v &= a_1 s_1 + \cdots + a_n s_n \\ &= b_1 t_1 + \cdots + b_m s_m \end{aligned}$$

then grouping any s_{i_j} and t_{i_j} that are equal, we have

$$0 = (a_{i_1} - b_{i_1}) s_{i_1} + \cdots + (a_{i_k} - b_{i_k}) s_{i_k} + a_{i_{k+1}} s_{i_{k+1}} + \cdots + a_{i_n} s_{i_n} \\ + b_{i_{k+1}} t_{i_{k+1}} + \cdots + b_{i_m} t_{i_m}$$

implies $a_{i_{k+1}} = \cdots = a_{i_n} = b_{i_{k+1}} = \cdots = b_{i_m} = 0$, so $n = m = k$, and for $j = 1, \ldots, k$ we have $a_{i_j} = b_{i_j}$, whence v is essentially uniquely represented as a linear combination of vectors in S.

(1)$\Longrightarrow$ (2): The contrapositive of Theorem 1.10.

(3)$\Longrightarrow$ (1): If 3 holds and $a_1 s_1 + \cdots + a_n s_n = 0$, where the s_i are distinct and $a_1 \neq 0$, then $n > 1$ and $s_1 = -\frac{1}{a_1}(a_2 s_2 + \cdots + a_n s_n)$, which violates 3. Thus 3 implies 1.

(1)$\Longleftrightarrow$ (4): If S is a linearly independent spanning set and T is a proper subset of S that also spans V, then any vectors in $S\backslash T$ would have to be linear combinations of the vectors in T, violating 3 for S (3 is equivalent to 1, as we just showed). Thus S is a minimal spanning set. Conversely, if S is a minimal spanning set, then it must be linearly independent, for otherwise there would be some $s \in S$ that is a linear combination of other vectors in S, which would mean $S\backslash\{s\}$ also spans V, contradicting minimality. Thus 4 implies 1.

(1)$\Longleftrightarrow$ (5): If 1 holds, so S is linearly independent and spans V, but S is not maximal, then $\exists v \in V\backslash S$ such that $S \cup \{v\}$ is also linearly independent. But $v \notin \text{span}(S)$ by 3, contradiction. Therefore S is maximal and 1 implies 5. Conversely, if S is a maximal linearly independent set, then S must span V, for otherwise there is a $v \in V\backslash S$ that isn't a linear combination of vectors in S, implying $S \cup \{v\}$ is linearly independent proper superset, violating maximality. Therefore $\text{span}(S) = V$, and 5 implies 1. ■

Corollary 1.14 *If V is a vector space, then $S = \{v_1, \ldots, v_n\} \subseteq V$ is a basis for V iff*

$$V = \text{span}(v_1) \oplus \text{span}(v_2) \oplus \cdots \oplus \text{span}(v_n) \tag{1.20}$$

■

The previous theorem did not say that any such basis exists for a given vector space. It merely gave the defining characteristics of a basis, should we ever come across one. The next theorem assures us that all vector spaces have a basis – a consequence of Zorns lemma.

Theorem 1.15 (Existence of a Basis) *If V is a nonzero vector space and $I \subseteq S \subseteq V$ with I linearly independent and $V = \text{span}(S)$, then $\exists B$ basis for V for which*

$$I \subseteq B \subseteq S \tag{1.21}$$

More specifically,

(1) *Any nonzero vector space has a basis.*

(2) *Any linearly independent set in V is contained in a basis.*

(3) *Any spanning set in V contains a basis.*

Proof: Let $\mathcal{A} \subseteq P(V)$ be the set of all linearly independent sets L such that $I \subseteq L \subseteq S$. Then $\mathcal{A}$ is non-empty because $I \subseteq I \subseteq S$, so $I \in \mathcal{A}$. Now, if

$$\mathcal{C} = \{I_k \mid k \in K\} \subseteq \mathcal{A}$$

is a chain in $\mathcal{A}$, that is a totally ordered (under set inclusion) subset of $\mathcal{A}$, then the union

$$U = \bigcup \mathcal{C} = \bigcup_{i \in K} I_i$$

is linearly independent and satisfies $I \subseteq U \subseteq S$, that is $U \in \mathcal{A}$. But by Zorn's lemma every chain has a maximal element B, so that $\exists B \in \mathcal{A}$ maximal which is linearly independent. But of course such a B is a basis for $V = \text{span}(S)$, for if any $s \in S$ is not a linear combination of elements in B, then $B \cup \{s\}$ is linearly independent and $B \subsetneq B \cup \{s\}$, contradicting the maximality of B. Therefore $S \subseteq \text{span}(B)$, and so $V = \text{span}(S) \subseteq \text{span}(B) \subseteq V$, or $V = \text{span}(B)$. This shows that (1) there is a basis B for V, (2) any linearly independent set I has $I \subseteq B$ for this B, and (3) any spanning set S has $B \subseteq S$ for this B. ■

Corollary 1.16 (All Subpaces Have Complements) *If S is a subspace of a vector space V, then there exists a subspace T of V such that*

$$V = S \oplus T \tag{1.22}$$

Proof: If $S = V$, then let $T = \{0\}$. Otherwise, if $V \backslash S \neq \varnothing$, then since V has a basis B and B is maximal, we must have $S = \text{span}(B \cap S)$ and $B \cap (V \backslash S) \neq \varnothing$. That is, there is a nonzero subspace $T = \text{span}(B \cap (V \backslash S))$, and by Corollary 1.8

$$V = \text{span}(B) = \text{span}((B \cap S) \cup (B \cap (V \backslash S))) = \text{span}(B \cap S) + \text{span}(B \cap (V \backslash S)) = S + T$$

But $S \cap T = \{0\}$ because of the essentially unique representation of vectors in V as linear combinations of vectors in B. Hence, $V = S \oplus T$, and T is a complement of S. ■

Theorem 1.17 (Replacement Theorem) *If V is a vector space such that $V = \text{span}(S)$ for some $S \subseteq V$ with $|S| = n$, and if $B \subseteq V$ is linearly independent with $|B| = m$, then*

(1) $m \leq n$

(2) $\exists C \subseteq S$ *with* $|C| = n - m$ *such that* $V = \text{span}(B \cup C)$

Proof: The proof is by induction on m, beginning with $m = 0$. For this m, $B = \varnothing$, so taking $C = S$, we have $B \cup C = \varnothing \cup S = S$, which generates V. Now suppose the theorem holds for some $m \geq 0$. For $m+1$, let $B = \{v_1, \ldots, v_{m+1}\} \subseteq V$ be linearly independent. By Corollary 1.8, $B' = \{v_1, \ldots, v_m\}$ is linearly independent, and so by the induction hypothesis $m \leq n$ and $\exists C' = \{u_1, \ldots, u_{n-m}\} \subseteq S$ such that $\{v_1, \ldots, v_n\} \cup C'$ generates V. This means $\exists a_1, \ldots, a_m, b_1, \ldots, b_{n-m} \in F$ such that

$$v_{m+1} = a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_{n-m} u_{n-m} \tag{1.23}$$

Note that if $n = m$, $C' = \varnothing$, so that $v_{m+1} \in \text{span}(B')$, contradicting the assumption that B is linearly independent. Therefore, $n > m$, or $n \geq m + 1$. Moreover, some b_i, say b_1, is nonzero, for otherwise we have $v_{m+1} = a_1 v_1 + \cdots + a_m v_m$, leading again to B being linearly dependent in contradiction to assumption. Solving (1.23) for u_1, we get

$$u_1 = (-b_1^{-1} a_1) v_1 + \cdots + (-b_1^{-1} a_m) v_m + b_1^{-1} v_{m+1} + (-b_1^{-1} b_2) u_2 + \cdots + (-b_1^{-1} b_{n-m}) u_{n-m}$$

Let $C = \{u_2, \ldots, u_{n-m}\}$. Then $u_1 \in \text{span}(B \cup C)$, and since $v_1, \ldots, v_m, u_2, \ldots, u_{n-m} \in B \cup C$, they are also in $\text{span}(B \cup C)$, whence

$$B' \cup C' \subseteq \text{span}(B \cup C)$$

Now, since $\text{span}(B' \cup C') = V$ by the induction hypothesis, we have by Corollary 1.8 that $\text{span}(B \cup C) = V$. So we have B linearly independent with $|B| = m+1$ vectors, $m+1 \leq n$, $\text{span}(B \cup C) = V$, and $C \subseteq S$ with $|C| = (n-m) - 1 = n - (m+1)$ vectors, so the theorem holds for $m+1$. Therefore the theorem is true for all $m \in \mathbb{N}$ by induction. ■

Corollary 1.18 *If V is a vector space with a finite spanning set, then every basis for V contains the same number of vectors.*

Proof: Suppose S is a finite spanning set for V, and let β and γ be two bases for V with $|\gamma| > |\beta| = k$, so that some subset $T \subseteq \gamma$ contains exactly $k+1$ vectors. Since T is linearly independent and β generates V, the replacement theorem implies that $k + 1 \leq k$, a contradiction. Therefore γ is finite and $|\gamma| \leq k$. Reversing the roles of β and γ shows that $k \leq |\gamma|$, so that $|\gamma| = |\beta| = k$. ■

Corollary 1.19 *If V is a finite-dimensional vector space with $\dim(V) = n$, then the following hold:*

1. *Any finite spanning set for V contains at least n vectors, and a spanning set for V that contains exactly n vectors is a basis for V.*
2. *Any linearly independent subset of V that contains exactly n vectors is a basis for V.*
3. *Every linearly independent subset of V can be extended to a basis for V.*

Proof: Let B be a basis for V. (1) Let $S \subseteq V$ be a finite spanning set for V. By Theorem 1.15, $\exists B \subseteq S$ that is a basis for V, and by Corollary 1.18 $|B| = n$ vectors. Therefore $|S| \geq n$, and $|S| = n \implies S = B$, so that S is a basis for V. (2) Let $I \subseteq V$ be linearly independent with $|I| = n$. By the replacement theorem $\exists T \subseteq B$ with $|T| = n - n = 0$ vectors such that $V = \text{span}(I \cup T) = \text{span}(I \cup \varnothing) = \text{span}(I)$. Since I is also linearly independent, it is a basis for V. (3) If $I \subseteq V$ is linearly independent with $|I| = m$ vectors, then the replacement theorem asserts that $\exists H \subseteq B$ containing exactly $n - m$ vectors such that $V = \text{span}(I \cup H)$. Now, $|I \cup H| \leq n$, so that by 1 $|I \cup H| = n$, and so $I \cup H$ is a basis for V extended from I. ■

Theorem 1.20 *Any two bases for a vector space V have the same cardinality.*

Proof: If V is finite-dimensional, then the previous corollary applies, so we need only consider bases that are infinite sets. Let $B = \{b_i \mid i \in I\}$ ba a basis for V and let C be any other basis for V. Then all $c \in C$ can be written as finite linear combinations of vectors in B, where all the coefficients are nonzero. That is, if $U_c = \{1, \ldots, n_c\}$, we have

$$c = \sum_{i=1}^{n_c} a_i b_i = \sum_{i \in U_c} a_i b_i$$

for unique $a_1, \ldots, a_{n_c} \in F$. But because C is a basis, we must have $I = \bigcup_{c \in C} U_c$, for if $\bigcup_{c \in C} \subsetneq I$, then all $c \in C$ can be expressed by a *proper* subset B' of B, so that $V = \text{span}(B')$, which is contradiction of the minimality of B as a spanning set. Now, for all $c \in C$ we have $|U_c| < \aleph_0$, which implies that

$$|B| = |I| = \left| \bigcup_{c \in C} U_c \right| \leq \aleph_0 |C| = |C|$$

Reversing the roles of B and C gives $|C| \leq |B|$, so by the Schröder-Bernstein theorem we have $|B| = |C|$. ■

Corollary 1.21 *If V is a vector space and S is a subspace of V:*

1. $\dim(S) \leq \dim(V)$
2. $\dim(S) = \dim(V) < \infty \implies S = V$
3. *V is infinite-dimensional iff it contains an infinite linearly independent subset.* ■

Theorem 1.22 *Let V be a vector space.*

1. *If B is a basis for V and $B = B_1 \cup B_2$, where $B_1 \cap B_2 = \varnothing$, then $V = \text{span}(B_1) \oplus \text{span}(B_2)$.*
2. *If $V = S \oplus T$ and B_1 is a basis for S and B_2 is a basis for T, then $B_1 \cap B_2 = \varnothing$ and $B = B_1 \cup B_2$ is a basis for V.*

Proof: (1) If $B_1 \cap B_2 = \varnothing$ and $B = B_1 \cup B_2$ is a basis for V, then $0 \notin B_1 \cup B_2$. But, if a nonzero vector $v \in \text{span}(B_1) \cap \text{span}(B_2)$, then $B_1 \cap B_2 \neq \varnothing$, a contradiction. Hence $\{0\} = \text{span}(B_1) \cap \text{span}(B_2)$. Moreover, since $B_1 \cup B_2$ is a basis for V, and since it is also a basis for $\text{span}(B_1) + \text{span}(B_2)$, we must have $V = \operatorname{span}(B_1) + \operatorname{span}(B_2)$, and hence $V = \operatorname{span}(B_1) \oplus \operatorname{span}(B_2)$. (2) If $V = S \oplus T$, then $S \cap T = \{0\}$, and since $0 \notin B_1 \cup B_2$, we have $B_1 \cap B_2 = \varnothing$. Also, since all $v \in V = S \oplus T$ have the form $a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n$ for $u_1, \ldots, u_m \in B_1$ and $v_1, \ldots, v_n \in B_2$, $v \in \operatorname{span}(B_1 \cup B_2)$, so $B_1 \cup B_2$ is a basis for V by Theorem 1.13. ■

Theorem 1.23 *If S and T are subspaces of a vector space V, then*

$$\dim(S) + \dim(T) = \dim(S + T) + \dim(S \cap T) \tag{1.24}$$

As a consequence, $V = S \oplus T$ iff $\dim(V) = \dim(S) + \dim(T)$.

Proof: If $B = \{b_i \mid i \in I\}$ is a basis for $S \cap T$, then we can extend this to a basis $A \cup B$ for S and to another basis $B \cup C$ for T, where $A = \{a_j \mid j \in J\}$, $C = \{c_k \mid k \in K\}$ and $A \cap B = \varnothing$ and $B \cap C = \varnothing$. Then $A \cup B \cup C$ is a basis for $S + T$: clearly $\operatorname{span}(A \cup B \cup C) = S + T$, so we need only verify that $A \cup B \cup C$ is linearly independent. To that end, suppose not, suppose $\exists c_1, \ldots, c_n \in F \backslash \{0\}$ and $v_1, \ldots, v_n \in A \cup B \cup C$ such that

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$$

Then some $v_i \in A \cap C$ since by our construction $A \cup B$ and $B \cup C$ are linearly independent. Isolating the vectors in A, say $v_1, \ldots, v_k$, on one side of the equality shows that there is a nonzero vector,

$$\begin{aligned} x &= \overbrace{a_1v_1 + \cdots + a_kv_k}^{\in \operatorname{span}(A)} \\ &= \underbrace{a_{k+1}v_{k+1} + \cdots + a_nv_n}_{\in \operatorname{span}(B \cup C)} \end{aligned} \quad \Longrightarrow \quad x \in \operatorname{span}(A) \cap \operatorname{span}(B \cup C) = \operatorname{span}(A) \cap T \subseteq S \cap T$$

since $\operatorname{span}(A) \subseteq S$. Consequently $x = 0$, and therefore $a_1 = \cdots = a_n = 0$ since A and $B \cup C$ are linearly independent sets, a contradiction. Thus $A \cup B \cup C$ is linearly independent and hence a basis for $S + T$. Moreover,

$$\begin{aligned} \dim(S) + \dim(T) &= |A \cup B| + |B \cup C| \\ &= |A| + |B| + |B| + |C| \\ &= |A \cup B \cup C| + |B| \\ &= \dim(S + T) + \dim(S \cap T) \end{aligned}$$

Of course if $S \cap T = \{0\}$, then $\dim(S \cap T) = 0$, and

$$\dim(S) + \dim(T) = \dim(S \oplus T) = \dim(V)$$

while if $\dim(V) = \dim(S) + \dim(T)$, then $\dim(S \cap T) = 0$, so $S \cap T = \{0\}$, and so $V = S \oplus T$. ■

Corollary 1.24 *If S and T are subspaces of a vector space V, then*

(1) $\dim(S + T) \leq \dim(S) + \dim(T)$

(2) $\dim(S + T) \leq \max\{\dim(S), \dim(T)\}$, *if S and T are finite-dimensional.* ■

Theorem 1.25 (Direct Sums) *If $\mathcal{F} = \{S_i \mid i \in I\}$ is a family of subspaces of a vector space V such that $V = \sum_{i\in I} S_i$, then the following statements are equivalent:*

(1) $V = \bigoplus_{i\in I} S_i$.

(2) $0 \in V$ *has a* **unique expression** *as a sum of vectors each from different S_i, namely as a sum of zeros: for any distinct $j_1, \dots, j_n \in I$, we have*

$$v_{j_1} + v_{j_2} + \cdots + v_{j_n} = 0 \text{ and } v_{j_i} \in S_{j_i} \text{ for each } i \implies v_{j_1} = \cdots = v_{j_n} = 0$$

(3) *Each $v \in V$ has a* **unique expression** *as a sum of distinct $v_{j_i} \in S_{j_i} \backslash \{0\}$,*

$$v = v_{j_1} + v_{j_2} + \cdots + v_{j_n}$$

(4) *If γ_i is a basis for S_i, then $\gamma = \bigcup_{i\in I} \gamma_i$ is a basis for V. If V is finite-dimensional, then this may be restated in terms of* ordered *bases γ_i and γ.*

Proof: (1) $\implies$ (2): Suppose 1 is true, that is $V = \bigoplus_{i\in I} S_i$, and choose $v_1, \dots, v_n \in V$ such that $v_i \in S_{k_i}$ and $v_1 + \cdots + v_k = 0$. Then for any j,

$$-v_j = \sum_{i\neq j} v_i \in \sum_{i\neq j} S_i$$

However, $-v_j \in S_j$, whence

$$-v_j \in W_j \cap \sum_{i\neq j} S_i = \{0\}$$

so that $v_j = 0$. This is true for all $j = 1, \dots, k$, so $v_1 = \cdots = v_k = 0$, proving 2.

(2) $\implies$ (3): Suppose 2 is true and let $v \in V = \sum_{i\in I} S_i$ be given by

$$\begin{aligned} v &= u_1 + \cdots + u_n \\ &= w_1 + \cdots + w_m \end{aligned}$$

for some $u_i \in S_{k_i}$ and $w_j \in S_{k_j}$. Then, grouping terms from the same subspaces

$$0 = v - v = (u_{i_1} - w_{i_1}) + \cdots + (u_{i_k} - w_{i_k}) + u_{i_{k+1}} + \cdots + u_{i_n} - w_{i_{k+1}} - \cdots - w_{i_m}$$
$$\implies \begin{cases} & u_{i_1} = w_{i_1}, \cdots, u_{i_k} = w_{i_k} \\ \text{and} & u_{i_{k+1}} = \cdots = u_{i_n} = w_{i_{k+1}} = \cdots = w_{i_m} = 0 \end{cases}$$

proving uniqueness.

(3) $\implies$ (4): Suppose each vector $v \in V = \sum_{i\in I}^n S_i$ can be uniquely written as $v = v_1 + \cdots + v_k$, for $v_i \in S_{j_i}$. For each $i \in I$ let γ_i be an ordered basis for S_i, so that since $V = \sum_{i\in I}^n S_i$, we must have that $V = \text{span}(\gamma) = \text{span}\left(\bigcup_{i=1}^k \gamma_i\right)$, so we only need to show that γ is linearly independent. To that end, let

$$\begin{aligned} v_{11}, v_{12}, \dots, v_{1n_1} &\in \gamma_{i_1} \\ &\vdots \\ v_{m1}, v_{m2}, \dots, v_{mn_m} &\in \gamma_{i_m} \end{aligned}$$

for any bases γ_{i_j} for $S_{i_j} \in \mathcal{F}$, then let $a_{ij} \in F$ for $i = 1, \dots, m$ and $j = 1, \dots, n_i$, and let $w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}$. Then, suppose

$$w_1 + \cdots + w_m = \sum_{i,j} a_{ij} v_{ij} = 0$$

since $0 \in \bigcap_{i=1}^{m} S_{j_i}$ and $0 = 0 + \cdots + 0$, by the assumed uniqueness of expression of $v \in V$ by $w_i \in S_i$ we must have $w_i = 0$ for all i, and since all the γ_i are linearly independent, we must have $a_{ij} = 0$ for all i, j.

(4) $\implies$ (1): if 4 is true, then there exist ordered bases γ_i the S_i such that $\gamma = \bigcup_{i \in I} \gamma_i$ is an ordered basis for V. By Corollary 1.8 and Theorem 1.22,

$$V = \operatorname{span}(\gamma) = \operatorname{span}(\bigcup_{i \in I} \gamma_i) = \sum_{i \in I} \operatorname{span}(\gamma_i) = \sum_{i \in I} S_i$$

Choose any $j \in I$ and suppose for a nonzero vector $v \in V$ we have $v \in S_j \cap \sum_{i \neq j} S_i$. Then,

$$v \in S_j = \operatorname{span}(\gamma_j) \cap \operatorname{span}(\bigcup_{i \neq j} \gamma_i) = \sum_{i \neq j} S_i$$

which means v is a nontrivial linear combination of elements in γ_i and elements in $\bigcup_{i \neq j} \gamma_i$, so that v can be expressed as a linear combination of $\bigcup_{i \in I} \gamma_i$ in more than one way, which contradicts uniqueness of representation of vectors in V in terms of a basis for V, Theorem 1.13. Consequently, any such v must be 0. That is,

$$S_j \cap \sum_{i \neq j} S_i = \{0\}$$

and the sum is direct, i.e. $V = \bigoplus_{i \in I} S_i$, proving 1. ■

2 Linear Transformations

2.1 Definitions

2.1.1 Linear Transformations, or Vector Space Homomorphisms

If V and W are vector spaces over the same field F, a function $T : V \to W$ is called a **linear transformation** or **vector space homomorphism** if it preserves the vector space structure, that is if for all vectors $x, y \in V$ and all scalars $a, b \in F$ we have

$$\begin{aligned} T(x+y) &= T(x)+T(y) \\ T(ax) &= aT(x) \end{aligned} \qquad \text{or equivalently} \qquad T(ax+by) = aT(x)+bT(y) \tag{2.1}$$

The **set of all linear transformations from V to W** is denoted

$$\mathcal{L}(V,W) \qquad \text{or} \qquad \mathrm{Hom}(V,W) \tag{2.2}$$

Linear transformation, like group and ring homomorphisms, come in different types:

1. **linear operator (vector space endomorphism)** on V – if V is a vector space over F, a linear operator, or vector space endomorphism, is a linear transformation $T \in \mathcal{L}(V,V)$. The **set of all linear operators**, or **vector space endomorphisms**, is denoted

$$\mathcal{L}(V) \qquad \text{or} \qquad \mathrm{End}(V) \tag{2.3}$$

2. **monomorphism (embedding)** – a 1-1 or injective linear transformation.
3. **epimorphism** – an onto or surjective linear transformation.
4. **isomorphism (invertible linear transformation)** from V to W – a bijective (and therefore invertible) linear transformation $T \in \mathcal{L}(V,W)$, i.e. a monomorphism which is also an epimorphism. If such an isomorphism exists, V is said to be **isomorphic to** W, and this relationship is denoted

$$V \cong W \tag{2.4}$$

The **set of all linear isomorphomorphisms from V to W** is denoted

$$\mathrm{GL}(V,W) \tag{2.5}$$

by analogy with the general linear group.

5. **automorphism** – a bijective linear operator. The **set of all automorphisms of V** is denoted

$$\mathrm{Aut}(V) \qquad \text{or} \qquad \mathrm{GL}(V) \tag{2.6}$$

Some common and important linear transformations are:

1. The **identity transformation**, $I_V : V \to V$, given by $I_V(x) = x$ for all $x \in V$. It's linearity is obvious: $I_V(ax+by) = ax+by = aI_V(x)+bI_V(y)$.
2. The **zero transformation**, $T_0 : V \to W$, given by $T_0(x) = 0$ for all $x \in V$. Sometimes this is denoted by 0 if the context is clear, like the zero polynomial $0 \in F[x]$. This is also clearly a linear transformation.
3. An **idempotent operator** $T \in \mathcal{L}(V)$ satisfies $T^2 = T$.

A crucial concept, which will allow us to classify and so distinguish operators, is that of similarity of operators. Two linear operators $T, U \in \mathcal{L}(V)$ are said to be **similar**, denoted

$$T \sim U \tag{2.7}$$

if there exists an automorphism $\phi \in \mathrm{Aut}(V)$ such that

$$T = \phi \circ U \circ \phi^{-1} \tag{2.8}$$

Similarity is an equivalence relation on $\mathcal{L}(V)$:

1. $T \sim T$, because $T = I \circ T \circ I^{-1}$
2. $T \sim U$ implies $T = \phi \circ U \circ \phi^{-1}$ for some automorphism $\phi \in \mathcal{L}(V)$, so that $U = \phi^{-1} \circ T \circ \phi$, or $U = \psi \circ T \circ \psi^{-1}$, where $\psi = \phi^{-1}$ is an automorphism. Thus $U \sim T$.
3. $S \sim T$ and $T \sim U$ implies $S = \phi \circ T \circ \phi^{-1}$ and $T = \psi \circ U \circ \psi^{-1}$ where ϕ and ψ are automorphisms, so
$$S = \phi \circ T \circ \phi^{-1} = \phi \circ (\psi \circ U \circ \psi^{-1}) \circ \phi^{-1} = (\phi \circ \psi) \circ U (\phi \circ \psi)^{-1}$$
and since $\phi \circ \psi$ is an automorphism of V, we have $S \sim U$.

This partitions $\mathcal{L}(V)$ into equivalence classes. As we will see, each equivalence class corresponds to a single matrix representation, depending only on the choice of basis for V, which maybe different for each operator.

The **kernel** or **null space** of a linear tranformation $T \in \mathcal{L}(V, W)$ is the pre-image of $0 \in W$ under T:

$$\ker(T) = T^{-1}(0) \tag{2.9}$$

i.e. $\ker(T) = \{x \in V \mid T(x) = 0\}$. As we will show below, $\ker(T)$ is a subspace of V.

The **range** or **image** of a linear tranformation $T \in \mathcal{L}(V, W)$ is the range of T as a set function: it is variously denoted by $R(T)$, $T(V)$ or $\mathrm{im}(T)$,

$$R(T) = \mathrm{im}(T) = T(V) := \{T(x) \mid x \in V\} \tag{2.10}$$

As we show below, $R(T)$ is a subspace of W. The **rank** of a linear transformation $T \in \mathcal{L}(V, W)$ is the dimension of its range:

$$\mathrm{rank}(T) := \dim\big(R(T)\big) \tag{2.11}$$

The **nullity** of $T \in \mathcal{L}(V, W)$ is the dimension of its kernel:

$$\mathrm{null}(T) := \dim\big(\ker(T)\big) \tag{2.12}$$

Given an operator $T \in \mathcal{L}(V)$, a subspace S of V is said to be T-**invariant** if $v \in S \implies T(v) \in S$, that is if $T(S) \subseteq S$. The restriction of T to a T-invariant subspace S is denoted T_S, and of course $T_S \in \mathcal{L}(S)$ is a linear operator on S. In the case that $V = R \oplus S$ decomposes as a direct sum of T-invariant subspaces R and S, we say the pair (R, S) **reduces** T, or is a **reducing pair** for T, and we say that T is the **direct sum of operators** T_R and T_S, denoting this by

$$T = T_R \oplus T_S \tag{2.13}$$

The expression $T = \tau \oplus \sigma \in \mathcal{L}(V)$ means there are subspaces R and S of V for which (R, S) reduces T and $\tau = T_R$ and $\sigma = T_S$.

If V is a direct sum of subspaces $V = R \oplus S$, then the function $\pi_{R,S} : V \to V$ given by

$$\pi_{R,S}(v) = \pi_{R,S}(r + s) = r \tag{2.14}$$

where $r \in R$ and $s \in S$, is called the **projection onto** R **along** S.

Example 2.1 *The projection on the y-axis along the x-axis is the map $\pi_{\mathbb{R}_y,\mathbb{R}_x} : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\pi_{\mathbb{R}_y,\mathbb{R}_x}(\mathbf{x}) = y$.* ■

Example 2.2 *The projection on the y-axis along the line $L = \{(s,s) \mid s \in \mathbb{R}\}$ is the map $\pi_{\mathbb{R}_y,L} : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\pi_{\mathbb{R}_y,L}(\mathbf{x}) = \pi_{\mathbb{R}_y,L}(s, s+y) = y$.* ■

Any projection $\pi_{R,S}$ is linear, i.e. $\pi_{R,S} \in \mathcal{L}(V)$:

$$\begin{aligned}
\pi_{R,S}(au+bv) &= \pi_{R,S}(a(r_1+s_1)+b(r_2+s_2)) \\
&= \pi_{R,S}((ar_1+br_2)+(as_1+bs_2)) \\
&= ar_1+br_2 \\
&= a\pi_{R,S}(u)+b\pi_{R,S}(v)
\end{aligned}$$

And clearly

$$R = \operatorname{im}(\pi_{R,S}) \qquad \text{and} \qquad S = \ker(\pi_{R,S}) \tag{2.15}$$

Two projections $\rho, \sigma \in \mathcal{L}(V)$ are said to be **orthogonal**, denoted

$$\rho \perp \sigma \tag{2.16}$$

if

$$\rho \circ \sigma = \sigma \circ \rho = 0 \equiv T_0 \tag{2.17}$$

This is of course by analogy with orthogonal vectors $v \perp w$, for which we have $\langle v, w \rangle = \langle w, v \rangle = 0$.

We use projections to define a **resolution of the identity**, a sum of the form $\pi_1 + \cdots + \pi_k = I \in \mathcal{L}(V)$, where the π_i are pairwise orthogonal projections, that is $\pi_i \perp \pi_j$ if $i \neq j$. This occurs when $V = S_1 \oplus S_2 \oplus \cdots \oplus S_k$, $\pi_i = \pi_{S_i, \bigoplus_{j \neq i} S_j}$, and $\pi_i \perp \pi_j$ for $i \neq j$. Resolutions of the identity will be important in establishing the spectral resolution of a linear operator.

2.2 Basic Properties of Linear Transformations

Theorem 2.3 *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V,W)$, then* $\ker(T)$ *is a subspace of V and* $\mathrm{im}(T)$ *is a subspace of W. Moreover, if $T \in \mathcal{L}(V)$, then* $\{0\}$, $\ker(T)$, $\mathrm{im}(T)$ *and V are T-invariant.*

Proof: (1) If $x, y \in \ker(T)$ and $a, b \in F$, we have $T(ax+by) = aT(x) + bT(y) = a0 + b0 = 0$ so $ax + by \in \ker(T)$, and the statement follows by Theorem 1.3. If $x, y \in \mathrm{im}(T)$ and $a, b \in F$, then $\exists u, v \in V$ such that $T(u) = x$ and $T(v) = y$, whence $ax+by = aT(u)+bT(v) = T(au+bv) \in \mathrm{im}(T)$. (2) Clearly $\{0\}$ and V are T-invariant, the first because $T(0) = 0$, the second because the codomain of T is V. If $x \in \mathrm{im}(T)$, then of course $x \in V$, so that $T(x) \in \mathrm{im}(T)$, while if $x \in \ker(T)$, then $T(x) = 0 \in \ker(T)$. ■

Theorem 2.4 *If V and W are vector spaces and $B = \{v_i \mid i \in I\}$ is a basis for V, then for any $T \in \mathcal{L}(V,W)$ we have*

$$\mathrm{im}(T) = \mathrm{span}(T(B)) \tag{2.18}$$

Proof: $T(v_i) \in \mathrm{im}(T)$ for each $i \in I$ by definition, and because $\mathrm{im}(T)$ is a subspace, we have $\mathrm{span}(T(B)) \subseteq \mathrm{im}(T)$. Moreover, $\mathrm{span}(T(B))$ is a subspace by Theorem 1.6. For the reverse inclusion, choose any $w \in \mathrm{im}(T)$j. Then $\exists v \in V$ such that $w = T(v)$, so that by Theorem 1.13 $\exists! a_1, \ldots, a_n \in F$ such that $v = a_1v_1 + \cdots + a_nv_n$, where by v_i we mean $v_i \equiv v_{j_i}$. Consequently,

$$w = T(v) = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n) \in \mathrm{span}(T(B)) \qquad ■$$

Theorem 2.5 (A Linear Transformation Is Defined by Its Action on a Basis) *Let V and W be vector spaces over the same field F. If $B = \{v_i \mid i \in I\}$ is a basis for V, then we can define a unique linear transformation $T \in \mathcal{L}(V,W)$ by arbitrarily specifying the values $w_i = T(v_i) \in W$ for each $i \in I$ and extending T by linearity, i.e. specifying that for all $a_1, \ldots, a_n \in F$ our T satisfy*

$$\begin{aligned} T(a_1v_1 + \cdots + a_nv_n) &= a_1T(v_1) + \cdots + a_nT(v_n) \\ &= a_1w_1 + \cdots + a_nw_n \end{aligned} \tag{2.19}$$

Proof: By Theorem 1.13: $\forall v \in V$, $\exists a_1, \ldots, a_n \in F$ such that $v = a_1v_1 + \cdots + a_nv_n$. Defining $T \in W^V$ by (2.19) we have the following results:

(1) $T \in \mathcal{L}(V,W)$: For any $u, v \in V$, there exist unique $a_1, \ldots, a_n, b_1, \ldots, b_m \in F$ and $u_1, \ldots, u_n$, $v_1, \ldots, v_m \in B$ such that

$$u = a_1u_1 + \cdots + a_nu_n \qquad v = b_1v_1 + \cdots + b_mv_m$$

If $w_1, \ldots, w_n, z_1, \ldots, z_m \in W$ are the vectors for which we have specified

$$\begin{aligned} T(u) &= a_1T(u_1) + \cdots + a_nT(u_n) \\ &= a_1w_1 + \cdots + a_nw_n \end{aligned} \qquad \begin{aligned} T(v) &= b_1T(v_1) + \cdots + b_mT(v_m) \\ &= b_1z_1 + \cdots + b_mz_m \end{aligned}$$

then, for all $a, b \in F$ we have

$$\begin{aligned} T(au+bv) &= T\big(a(a_1u_1 + \cdots + a_nu_n) + b(b_1v_1 + \cdots + b_mv_m)\big) \\ &= aa_1T(u_1) + \cdots + aa_nT(u_n) + bb_1T(v_1) + \cdots + bb_mT(v_m) \\ &= aT(a_1u_1 + \cdots + a_nu_n) + bT(b_1v_1 + \cdots + b_mv_m) \\ &= aT(u) + bT(u) \end{aligned}$$

(2) T is unique: suppose $\exists U \in \mathcal{L}(V,W)$ such that $U(v_i) = w_i = T(v_i)$ for all $i \in I$. Then for any $v \in V$ we have $v = \sum_{i=1}^n a_i v_i$ for unique $a_i \in F$, so that

$$U(v) = \sum_{i=1}^n a_i U(v_i) = \sum_{i=1}^n a_i w_i = T(v)$$

and $U = T$. ■

Theorem 2.6 *If V, W, Z are vector spaces over the same field F and $T \in \mathcal{L}(V,W)$ and $U \in \mathcal{L}(W,Z)$, then $U \circ T \in \mathcal{L}(V,Z)$.*

Proof: If $v, w \in V$ and $a, b \in F$, then

$$\begin{aligned}(U \circ T)(ac + bw) = U(T(ac + bw)) = U(aT(v) + bT(w))& \\ = aU(T(v)) + bU(T(w)) = a(U \circ T)(x) + b(U \circ T)(y)&\end{aligned}$$

so $U \circ T \in \mathcal{L}(V,Z)$. ■

Theorem 2.7 *If V and W are vector spaces over F, then $\mathcal{L}(V,W)$ is a vector space over F under ordinary addition and scalar multiplication of functions.*

Proof: First, note that if $T, U \in \mathcal{L}(V,W)$ and $a, b, c, d \in F$, then $\forall v, w \in V$

$$\begin{aligned}(aT + bU)(cv + dw) &= aT(cv + dw) + bU(cv + dw) \\ &= acT(v) + adT(w) + bcU(v) + bdU(w) \\ &= c\big(aT(v) + bU(v)\big) + d\big(aT(w) + bU(w)\big) \\ &= c(aT + bU)(v) + d(aT + bU)(w)\end{aligned}$$

so $(aT + bU) \in \mathcal{L}(V,W)$, and $\mathcal{L}(V,W)$ is closed under addition and scalar multiplication. Moreover, $\mathcal{L}(V,W)$ satisfies VS1-8 because W is a vector space, and addition and scalar multiplication in $\mathcal{L}(V,W)$ are defined pointwise for values which any functions $S, T, U \in \mathcal{L}(V,W)$ take in W. ■

2.3 Monomorphisms and Isomorphisms

Theorem 2.8 *If V, W, Z are vector spaces over the same field F and $T \in \mathrm{GL}(V,W)$ and $U \in \mathrm{GL}(W,Z)$ are isomorphisms, then*

(1) $T^{-1} \in \mathcal{L}(W,V)$

(2) $(T^{-1})^{-1} = T$, *and hence* $T^{-1} \in \mathrm{GL}(W,V)$.

(3) $(U \circ T)^{-1} = T^{-1} \circ U^{-1} \in \mathrm{GL}(Z,V)$.

Proof: (1) If $w, z \in W$ and $a, b \in F$, since T is bijective $\exists! u, v \in V$ such that $T(u) = w$ and $T(v) = z$. Hence $T^{-1}(w) = u$ and $T^{-1}(z) = v$, so that

$$\begin{aligned} T^{-1}(aw + bz) = T^{-1}\big(aT(u) + bT(v)\big) &= T^{-1}\big(T(au + bv)\big) \\ &= au + bv = aT^{-1}(w) + bT^{-1}(z) \end{aligned}$$

(2) $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$ shows that T is the inverse of T^{-1}, that is $T = (T^{-1})^{-1}$.

(3) First, any composition of bijections is a bijection, so $U \circ T$ is bijective. Moreover,

$$\begin{aligned} (U \circ T) \circ \underline{(U \circ T)^{-1}} = \underline{I_Z} = U \circ U^{-1} &= U \circ I_W \circ U^{-1} \\ &= U \circ (T \circ T^{-1}) \circ U^{-1} = (U \circ T) \circ \underline{(T^{-1} \circ U^{-1})} \end{aligned}$$

and

$$\begin{aligned} \underline{(U \circ T)^{-1}} \circ (U \circ T) = \underline{I_V} = T^{-1} \circ T &= T^{-1} \circ I_W \circ T \\ &= T^{-1} \circ (U^{-1} \circ U) \circ T = \underline{(U^{-1} \circ T^{-1})} \circ (U \circ T) \end{aligned}$$

so by the cancellation law and the fact that $\mathcal{L}(V,Z)$ is a vector space, we must have $(U \circ T)^{-1} = T^{-1} \circ U^{-1} \in \mathcal{L}(Z,V)$ and by (1) it's an isomorphism. Alternatively, we can get (3) from (2) via Theorem 2.5:

$$U \circ T : V \to Z, \quad \text{so} \quad (U \circ T)^{-1} : Z \to V$$

and of course

$$T^{-1} \circ U^{-1} : Z \to V$$

Let α, β and γ be bases for V, W and Z, respectively, such that

$$T(\alpha) = \beta \quad \text{and} \quad U(\beta) = U(T(\alpha)) = (U \circ T)(\alpha) = \gamma, \quad \text{so that} \quad (U \circ T)^{-1}(\gamma) = \alpha$$

which is possible by Theorem 2.10. But we also have

$$U^{-1}(\gamma) = \beta \quad \text{and} \quad T^{-1}(\beta) = \alpha$$

so that

$$(T^{-1} \circ U^{-1})(\gamma) = \alpha$$

By Theorem 2.5 (uniqueness) $(U \circ T)^{-1} = T^{-1} \circ U^{-1}$ and by (1) it's an isomorphism. ■

Theorem 2.9 *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V,W)$, then T is injective iff* $\ker(T) = \{0\}$.

Proof: If T is 1-1 and $x \in \ker(T)$, then $T(x) = 0 = T(0)$, so that $x = 0$, whence $\ker(T) = \{0\}$. Conversely, if $\ker(T) = \{0\}$ and $T(x) = T(y)$, then,

$$0 = T(x) - T(y) = T(x - y) \quad \Longrightarrow \quad x - y = 0$$

or $x = y$, and so T is 1-1. ■

Theorem 2.10 *Let V and W be vector spaces over the same field and $S \subseteq V$. Then for any $T \in \mathcal{L}(V, W)$,*

(1) *T is injective iff it carries linearly independent sets into linearly independent sets.*

(2) *T is injective implies $S \subseteq V$ is linearly independent iff $T(S) \subseteq W$ is linearly independent.*

Proof: (1) If T is 1-1 and $S \subseteq V$ is linearly independent, then for any $v_1, \dots, v_n \in S$ we have $0_V = a_1v_1 + \cdots + a_nv_n \implies a_1 = \cdots a_n = 0$, so

$$0_W = T(0_V) = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n)$$

whence $\{T(v_1), \dots, T(v_n)\}$ is linearly independent, and since this is true for all $v_i \in S$, $T(S)$ is linearly independent, so T carries linearly independent sets into linearly independent sets. Conversely, if T carries linearly independent sets into linearly independent sets, let B be a basis for V and suppose $T(u) = T(v)$ for some $u, v \in V$. Since $u = a_1u_1 + \cdots + a_nu_n$ and $v = b_1v_1 + \cdots + b_mv_m$ for unique $a_i, b_i \in F$ and $u_1, \dots, u_n, v_1, \dots, v_m \in B$, we have

$$\begin{aligned}
0 &= T(u) - T(v) \\
&= T(u - v) \\
&= T\big((a_{i_1} - b_{j_1})v_{i_1} + \cdots + (a_{i_k} - b_{j_k})v_{i_k} + a_{i_{k+1}}u_{i_{k+1}} + \cdots + a_{i_n}u_{i_n} \\
&\qquad + b_{i_{k+1}}v_{i_{k+1}} + \cdots + b_{i_m}v_{i_m}\big) \\
&= (a_{i_1} - b_{j_1})T(v_{i_1}) + \cdots + (a_{i_k} - b_{j_k})T(v_{i_k}) \\
&\qquad + a_{i_{k+1}}T(u_{i_{k+1}}) + \cdots + a_{i_n}T(u_{i_n}) + b_{i_{k+1}}T(v_{i_{k+1}}) + \cdots + b_{i_m}T(v_{i_m})
\end{aligned}$$

so that $a_{i_{k+1}} = \cdots = a_{i_n} = b_{i_{k+1}} = \cdots = b_{i_m} = 0$, which means $m = n = k$ and therefore $a_{i_1} = b_{j_1}$, $\dots, a_{i_k} = b_{j_k}$. Consequently

$$u = a_{i_1}v_{i_1} + \cdots + a_{i_k}v_{i_k} = v$$

whence T is 1-1.

(2) If T is 1-1 and $S \subseteq V$ is linearly independent, then by (1) $T(S)$ is linearly independent. Conversely, if T is 1-1 and $T(S)$ is linearly independent, then for $K = \{v_1, \dots, v_k\} \subseteq S \subseteq V$ we have that

$$T(0) = 0 = a_1T(v_1) + \cdots + a_kT(v_k) = T(a_1v_1 + \cdots + a_kv_k)$$

implying simultaneously that $a_1v_1 + \cdots + a_kv_k = 0$ and $a_1 = \cdots = a_k = 0$, so K is linearly independent. But this is true for all such $K \subseteq S$, so S is linearly independent. ■

Theorem 2.11 *If V and W are vector spaces over the same field F and $T \in \mathrm{GL}(V, W)$, then for any $S \subseteq V$ we have*

(1) *$V = \mathrm{span}(S)$ iff $W = \mathrm{span}(T(S))$.*

(2) *S is linearly independent in V iff $T(S)$ is linearly independent in W.*

(3) *S is a basis for V iff $T(S)$ is a basis for W.*

Proof: (1) $V = \mathrm{span}(S) \iff W = \mathrm{im}(T) = \overbrace{T(\mathrm{span}(S)) = \mathrm{span}(T(S))}^{T \in \mathrm{GL}(V,W)}$.

(2) S linearly independent means for any $s_1, \dots, s_n \in S$ we have $\sum_{i=1}^{n} a_is_i = 0 \iff$ for all $a_i = 0$, which implies

$$\begin{aligned}
T\Big(\sum_{i=1}^{n} a_is_i\Big) = \sum_{i=1}^{n} a_iT(s_i) = 0 = T(0) \quad &\overset{T \in \mathrm{GL}(V,W)}{\Longrightarrow} \quad \sum_{i=1}^{n} a_is_i = 0 \\
&\overset{S \text{ lin. indep.}}{\Longrightarrow} \quad a_1 = \cdots = a_n = 0
\end{aligned}$$

so $T(S)$ is linearly independent, since this is true for all $s_i \in S$. Conversely, if $T(S)$ is linearly independent we have for any $T(s_1), \dots, T(s_n) \in T(S)$

$$0 = \sum_{i=1}^{n} a_i T(s_i) = T\Big(\sum_{i=1}^{n} a_i s_i\Big) = T(0) \quad \overset{T \in \mathrm{GL}(V,W)}{\Longrightarrow} \quad \sum_{i=1}^{n} a_i s_i = 0$$
$$\overset{T(S)\ \text{lin. indep.}}{\Longrightarrow} \quad a_1 = \cdots = a_n = 0$$

so S is linearly independent.

(3) S basis for V $\overset{(1),\,(2)}{\Longleftrightarrow}$ $T(S)$ linearly independent in W and $W = \mathrm{span}(T(S)) \implies T(S)$ is a basis for W. We could also say that since T^{-1} is an isomorphism by Theorem 2.8, we have that since $T(S)$ is a basis for W, $S = T^{-1}(T(S))$ is a basis for V. ■

Theorem 2.12 (Isomorphisms Preserve Bases) *If V and W are vector spaces over the same field F and B is a basis for V, then $T \in \mathrm{GL}(V,W)$ iff $T(B)$ is a basis for W.*

Proof: If $T \in \mathrm{GL}(V,W)$ is an isomorphism, then T is bijective, so by Theorem 2.11 $T(B)$ is a basis for W. Conversely, if $T(B)$ is a basis for W, then $\forall u \in V$, $\exists! a_1, \dots, a_n \in F$ and $v_1, \dots, v_n \in B$ such that $u = a_1 v_1 + \cdots + a_n v_n$. Therefore,

$$0 = T(u) = T(a_1 v_1 + \cdots + a_n v_n) = a_1 T(v_1) + \cdots + a_n T(v_n) \implies a_1 = \cdots = a_n = 0$$

so $\ker(T) = \{0\}$, whence by Theorem 2.9 T is injective. But since $\mathrm{span}(T(B)) = W$, we have for all $w \in W$ that $\exists! a_1, \dots a_n \in F$ such that $w = a_1 T(v_1) + \cdots + a_n T(v_n) = T(a_1 v_1 + \cdots + a_n v_n)$, so $\exists u = a_1 v_1 + \cdots + a_n v_n \in V$ such that $w = T(u)$, and $W \subseteq \mathrm{im}(T)$. We obviously have $\mathrm{im}(T) \subseteq W$, and so T is surjective. Therefore it is bijective and an isomorphism. ■

Theorem 2.13 (Isomorphisms Preserve Dimension) *If V and W are vector spaces over the same field F, then*

$$V \cong W \iff \dim(V) = \dim(W) \tag{2.20}$$

Proof: $V \cong W \implies \exists T \in \mathrm{GL}(V,W)$, so B basis for $V \implies T(B)$ basis for W, and $\dim(V) = |B| = |T(B)| = \dim(W)$. Conversely, if $\dim(V) = |B| = |C| = \dim(W)$, where C is a basis for W, then $\exists T : B \to C$ bijective. Extending T to V by linearity defines a unique $T \in \mathcal{L}(V,W)$ by Theorem 2.5 and T is an isomorphism because it is surjective, $\mathrm{im}(T) = W$, and injective, $\ker(T) = \{0\}$, so $V \cong W$. ■

Corollary 2.14 *If V and W are vector spaces over the same field and $T \in \mathcal{L}(V,W)$, then $\dim(V) < \dim(W) \implies T$ cannot be onto and $\dim(V) > \dim(W) \implies T$ cannot be 1-1.* ■

Corollary 2.15 *If V is an n-dimensional vector space over F, where $n \in \mathbb{N}$, then*

$$V \cong F^n \tag{2.21}$$

If κ is any cardinal number, B is a set of cardinality κ and V is a κ-dimensional vector space over F, then

$$V \cong (F^B)_0 \tag{2.22}$$

Proof: As a result of the fact that $\dim(F^n) = n$ and $\dim\big((F^B)_0\big) = \kappa$, by the previous theorem. ■

Corollary 2.16 *If V and W are vector spaces over the same field F and $T \in \mathrm{GL}(V,W)$, then for any subspace S of V we have that $T(S)$ is a subspace of W and $\dim(S) = \dim(T(S))$.*

Proof: If S is a subspace of V, then $\forall x, y \in S$ and $a, b \in F$, we have $ax + by \in S$, and $T(x), T(y) \in S$ and $T(ax + by) = aT(x) + bT(y) \in T(S)$ imply that S is a subspace of W, while $\dim(S) = \dim(T(S))$ follows from the theorem. ■

2.4 Rank-Nullity Theorem

Lemma 2.17 *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$, then any complement of the kernel of T is isomorphic to the range of T, that is*

$$V = \ker(T) \oplus \ker(T)^c \quad \Longrightarrow \quad \ker(T)^c \cong \operatorname{im}(T) \tag{2.23}$$

where $\ker(T)^c$ *is any complement of* $\ker(T)$.

Proof: By Theorem 1.23 $\dim(V) = \dim\big(\ker(T)\big) + \dim\big(\ker(T)^c\big)$. Let

$$T^c = T|_{\ker(T)^c} : \ker(T)^c \to \operatorname{im}(T)$$

and note that T is injective by Theorem 2.9 because

$$\ker(T^c) = \ker(T) \cap \ker(T^c) = \{0\}$$

Moreover T^c is surjective because $T^c(V) = \operatorname{im}(T)$: first we obviously have $T^c(V) \subseteq \operatorname{im}(T)$, while for the reverse inclusion suppose $v \in \operatorname{im}(T)$. Then by Theorem 1.25 $\exists! s \in \ker(T)$ and $t \in \ker(T)^c$ such that $v = s + t$, so that

$$T(v) = T(s+t) = T(s) + \cancelto{0}{T(t)} = T(s) = T^c(s) \in T^c(V)$$

so $\operatorname{im}(T) = T^c(V)$ and T^c is surjective, so that T^c is an isomorphism, whence

$$\ker(T)^c \cong \operatorname{im}(T)$$

■

Theorem 2.18 (Rank-Nullity Theorem) *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$, then*

$$\dim(V) = \operatorname{rank}(T) + \operatorname{null}(T) \tag{2.24}$$

Proof: By the lemma and we have $\ker(T)^c \cong \operatorname{im}(T)$, which by the previous theorem implies $\dim\big(\ker(T)^c\big) = \dim\big(\operatorname{im}(T)\big) = \operatorname{rank}(T)$, while by Theorem 1.23 $V = \ker(T) \oplus \ker(T)^c$ implies

$$\begin{aligned} \dim(V) &= \dim\big(\ker(T)\big) + \dim\big(\ker(T)^c\big) \\ &= \dim\big(\ker(T)\big) + \dim\big(\operatorname{im}(T)\big) \\ &= \operatorname{null}(T) + \operatorname{rank}(T) \end{aligned}$$

which completes the proof. ■

Corollary 2.19 *Let V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$. If* $\dim(V) = \dim(W)$, *then the following are equivalent:*

(1) *T is injective.*

(2) *T is surjective.*

(3) $\operatorname{rank}(T) = \dim(V)$

Proof: By the Rank-Nullity Theorem, $\operatorname{rank}(T) + \operatorname{null}(T) = \dim(V)$, and by Theorem 2.9 we have

$$\begin{aligned} T \text{ is 1-1} &\overset{\text{Thm 2.9}}{\iff} \ker(T) = \{0\} \\ &\iff \operatorname{null}(T) = 0 \\ &\overset{\text{R-N Thm}}{\iff} \dim(\operatorname{im}(T)) = \operatorname{rank}(T) = \dim(V) \overset{\text{assump.}}{=} \dim(W) \\ &\overset{\text{Thm 1.21.1}}{\iff} \operatorname{im}(T) = W. \\ &\iff T \text{ is onto} \end{aligned}$$

which completes the proof. ■

Theorem 2.20 *For any linear operator $T \in \mathcal{L}(V)$ we have that*

$$\ker(T) \subseteq \ker(T^2) \subseteq \cdots \tag{2.25}$$

Moreover, if there is an $m \in \mathbb{N}$ for which $\mathrm{rank}(T^m) = \mathrm{rank}(T^{m+1})$, then for all $n \in \mathbb{N}$ we have

$$\mathrm{rank}(T^m) = \mathrm{rank}(T^{m+n}) \qquad\qquad \ker(T^m) = \ker(T^{m+n}) \tag{2.26}$$

Proof: (1) Let $n \in \mathbb{N}$ and $v \in \ker(T^n)$, so that $T^n(v) = 0$. Then, $0 = T(0) = T(T^n(v)) = T^{n+1}(v)$, so $v \in \ker(T^{n+1})$, which proves (2.25).

(2) If there is an $m \in \mathbb{N}$ such that $\mathrm{rank}(T^m) = \mathrm{rank}(T^{m+1})$, then by the Rank-Nullity Theorem $\mathrm{null}(T^m) = \mathrm{null}(T^{m+1})$, which by (1) shows that $\ker(T^m) = \ker(T^{m+1})$ and hence $T^m(V) = T^{m+1}(V)$. But then $T|_{\ker(T^m)} = T_0 \in \mathcal{L}(\ker(T^m))$ and $T|_{T^m(V)}$ must be bijective. Consequently, $T|_{T^m(V)}(v) \neq 0$ if $v \neq 0$, so $T|^n_{T^m(V)}(v) \neq 0$ if $v \neq 0$ for all $n \in \mathbb{N}$. This shows that $T|^n_{T^m(V)}(v) = 0$ iff $v \in \ker(T^m)$ for all $n \in \mathbb{N}$, so $\ker(T^{m+n}) = \ker(T^m)$ and $T^m(V) = T^{m+n}(V)$ for all $n \in \mathbb{N}$, proving (2.26). ■

3 Matrix Representations of Linear Transformations

3.1 Definitions

If $\beta = (v_1, \ldots, v_n)$ is an ordered basis for a finite-dimensional vector space V over F, then by Theorem 1.13 we know that for any vector $x \in V$ there are unique scalars $a_1, \ldots, a_n \in F$ such that $x = a_1 v_1 + \cdots + a_n v_n$. The **coordinate vector** of $x \in V$ with respect to (relative to) β is defined to be the (column) vector in F^n consisting of the scalars a_i:

$$[x]_\beta := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{3.1}$$

and the **coordinate map**, also called the **standard representation** of V with respect to β,

$$\phi_\beta : V \to F^n \tag{3.2}$$

is given by

$$\phi_\beta(x) = [x]_\beta \tag{3.3}$$

Let V and W be finite dimensional vector spaces over the same field F with ordered bases $\beta = (v_1, \ldots, v_n)$ and $\gamma = (w_1, \ldots, w_m)$, respectively. If there exist (and there do exist) unique scalars $a_{ij} \in F$ such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \qquad \text{for } j = 1, \ldots, n \tag{3.4}$$

then the **matrix representation of a linear transformation** $T \in \mathcal{L}(V, W)$ **in the ordered bases** β **and** γ is the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$,

$$\begin{aligned} A = [T]_\beta^\gamma &:= \Big[[T(v_1)]_\gamma \ \ [T(v_2)]_\gamma \ \cdots \ [T(v_n)]_\gamma \Big] \\ &= \left[\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right] = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \end{aligned}$$

Here are some examples:

1. If $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ is given by $T(x_1, x_2) = (x_1 + 3x_2,\ 0,\ 2x_1 - 4x_2)$, and $\beta = (\mathbf{e}_1, \mathbf{e}_2)$ and $\gamma = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ are the standard ordered bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively, then the matrix representation of T is found as follows:

$$\left.\begin{aligned} T(\mathbf{e}_1) = T(1,0) = (1,0,2) = 1\mathbf{e}_1 + 0\mathbf{e}_2 + 2\mathbf{e}_3 \\ T(\mathbf{e}_2) = T(0,1) = (3,0,-4) = 3\mathbf{e}_1 + 0\mathbf{e}_2 - 4\mathbf{e}_3 \end{aligned}\right\} \quad \Longrightarrow \quad [T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}$$

2. If $T \in \mathcal{L}(\mathbb{R}_3[x], \mathbb{R}_2[x])$ is given by $T(f(x)) = f'(x)$, and $\beta = (1, x, x^2, x^3)$ and $\gamma = (1, x, x^2)$ are the standard ordered bases for $\mathbb{R}_3[x]$ and $\mathbb{R}_2[x]$, respectively, the matrix representation of T is found as follows:

$$\left.\begin{aligned} T(1) &= 0 \cdot 1 + 0x + 0x^2 \\ T(x) &= 1 \cdot 1 + 0x + 0x^2 \\ T(x^2) &= 0 \cdot 1 + 2x + 0x^2 \\ T(x^3) &= 0 \cdot 1 + 0x + 3x^2 \end{aligned}\right\} \quad \Longrightarrow \quad [T]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

3.2 Matrix Representations of Linear Transformations

Theorem 3.1 *If $A \in M_{m,n}(F)$, $B, C \in M_{n,p}(F)$, $D, E \in M_{q,m}(F)$, and $a \in F$, then*

(1) $A(B+C) = AB + AC$ *and* $(D+E)A = DA + EA$

(2) $a(AB) = (aA)B = A(aB)$

(3) $I_m A = A = AI_n$

Proof: For each case we show the result by showing that it holds for each ijth entry as a result of the properties of the field F:

$$\begin{aligned}[A(B+C)]_{ij} &= \sum_{k=1}^{n} A_{ik}(B+C)_{kj} = \sum_{k=1}^{n} A_{ik}(B_{kj}+C_{kj}) = \sum_{k=1}^{n}(A_{ik}B_{kj} + A_{ik}C_{kj}) \\ &= \sum_{k=1}^{n} A_{ik}B_{kj} + \sum_{k=1}^{n} A_{ik}C_{kj} = (AB)_{ij} + (AC)_{ij} = [AB+AC]_{ij}\end{aligned}$$

$$\begin{aligned}[(D+E)A]_{ij} &= \sum_{k=1}^{m}(D+E)_{ik}A_{kj} = \sum_{k=1}^{m}(D_{ik}+E_{ik})A_{kj} = \sum_{k=1}^{m}(D_{ik}A_{kj} + E_{ik}A_{kj}) \\ &= \sum_{k=1}^{m} D_{ik}A_{kj} + \sum_{k=1}^{m} E_{ik}A_{kj} = (DA)_{ij} + (EA)_{ij} = [DA+EA]_{ij}\end{aligned}$$

$$\begin{aligned}\underline{[a(AB)]_{ij}} &= a\sum_{k=1}^{n} A_{ik}B_{kj} = \sum_{k=1}^{n} aA_{ik}B_{kj} = \underline{[(aA)B]_{ij}} = \sum_{k=1}^{n} aA_{ik}B_{kj} \\ &= \sum_{k=1}^{n} A_{ik}aB_{kj} = \underline{[A(aB)]_{ij}}\end{aligned}$$

$$\underline{(I_mA)_{ij}} = \sum_{k=1}^{m}(I_m)_{ik}A_{kj} = \sum_{k=1}^{m}\delta_{ik}A_{kj} = \underline{A_{ij}} = \sum_{k=1}^{n} A_{ik}\delta_{kj} = \sum_{k=1}^{n} A_{ik}(I_n)_{kj} = \underline{(AI_n)_{ij}}$$

■

Theorem 3.2 *If $A \in M_{m,n}(F)$ and $B \in M_{n,p}(F)$, and $\mathbf{b}_j$ is the jth column of B, then*

$$\begin{aligned}AB &= A[\mathbf{b}_1 \;\; \mathbf{b}_2 \;\; \cdots \;\; \mathbf{b}_p] && (3.5)\\ &= [A\mathbf{b}_1 \;\; A\mathbf{b}_2 \;\; \cdots \;\; A\mathbf{b}_p] && (3.6)\end{aligned}$$

That is if $\mathbf{u}_j$ is the jth column of AB, then $\mathbf{u}_j = A\mathbf{b}_j$. As a result, if $A \in M_{m,n}(F)$ and $\mathbf{a}_j$ is the jth column of A, then

$$A = AI_n = I_mA \tag{3.7}$$

Proof: First, we have

$$\mathbf{u}_j = \begin{pmatrix}(AB)_{1j}\\ \vdots \\ (AB)_{mj}\end{pmatrix} = \begin{pmatrix}\sum_{k=1}^{n} A_{1k}B_{kj}\\ \vdots \\ \sum_{k=1}^{n} A_{mk}B_{kj}\end{pmatrix} = A\begin{pmatrix}B_{1j}\\ \vdots \\ B_{nj}\end{pmatrix} = A\mathbf{v}_j$$

As a result

$$A = [\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n] = [A\mathbf{e}_1 \;\; A\mathbf{e}_2 \;\; \cdots \;\; A\mathbf{e}_n] = AI_n$$

and

$$I_mA = I_m[\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n] = [I_m\mathbf{a}_1 \;\; I_m\mathbf{a}_2 \;\; \cdots \;\; I_m\mathbf{a}_n] = [\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n] = A$$

■

Theorem 3.3 *If V is a finite-dimensional vector space over a field F with ordered basis $\beta = (b_1, \ldots, b_n)$, the coordinate map $\phi_\beta : V \to F^n$ is an isomorphism,*

$$\phi_\beta \in \mathrm{GL}(V, F^n) \tag{3.8}$$

Proof: First, $\phi_\beta \in \mathcal{L}(V, F^n)$ since $\forall x, y \in V$ and $a, b \in F$ we have by the ordinary rules of matrix addition and scalar multiplication (or alternatively by the rules of addition and scalar multiplication on n-tuples)

$$\phi_\beta(ax + by) = [ax + by]_\beta = a[x]_\beta + [y]_\beta = a\phi_\beta(x) + \phi_\beta(y)$$

Second, because $\phi_\beta(b_1) = \mathbf{e}_1, \ldots, \phi_\beta(b_n) = \mathbf{e}_n$, it takes a basis into a basis. Therefore it is an isomorphism by Theorem 2.12. ■

Theorem 3.4 *Let V and W be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let $T \in \mathcal{L}(V, W)$. Then for all $v \in V$ we have*

$$[T(v)]_\gamma = [T]_\beta^\gamma [v]_\beta \tag{3.9}$$

In other words, if $A = [T]_\beta^\gamma$ is the matrix representation of T, if $T_A \in \mathcal{L}(F^n, F^m)$ is the matrix multiplication map, and if $\phi_\beta \in \mathrm{GL}(V, F^n)$ and $\phi_\gamma \in \mathrm{GL}(W, F^n)$ are the respective coordinate maps, then

$$\phi_\gamma \circ T = T_A \circ \phi_\beta \tag{3.10}$$

and the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\;T\;} & W \\ \phi_\beta \big\downarrow & & \big\downarrow \phi_\gamma \\ F^n & \xrightarrow[\;T_A\;]{} & F^m \end{array}$$

Proof: If $\beta = (v_1, \ldots, v_n)$ is an odered basis for V and $\gamma = (w_1, \ldots, w_m)$ is an ordered basis for W, then let

$$[T]_\beta^\gamma = \Big[[T(v_1)]_\gamma \;\cdots\; [T(v_n)]_\gamma\Big] = \left[\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right] = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Now, $\forall u \in V$, $\exists! b_1, \ldots, b_n \in F$ such that $u = b_1 v_1 + \cdots + b_n v_n$. Therefore,

$$\begin{aligned} T(u) &= T(b_1 v_1 + \cdots + b_n v_n) \\ &= b_1 T(v_1) + \cdots + b_n T(v_n) \end{aligned}$$

so that

$$\begin{aligned} [T(u)]_\gamma = \phi_\gamma\big(T(u)\big) &= \phi_\gamma\big(b_1 T(v_1) + \cdots + b_n T(v_n)\big) \\ &= b_1 \phi_\gamma\big(T(v_1)\big) + \cdots + b_n \phi_\gamma\big(T(v_n)\big) \\ &= b_1 [T(v_1)]_\gamma + \cdots + b_n [T(v_n)]_\gamma \\ &= [T]_\beta^\gamma [u]_\beta \end{aligned}$$

which completes the proof. ■

Theorem 3.5 *If V, W, Z are finite-dimensional vector spaces with ordered bases α, β and γ, respectively, and $T \in \mathcal{L}(V, W)$ and $U \in \mathcal{L}(W, Z)$, then*

$$[U \circ T]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta \tag{3.11}$$

Proof: By the previous two theorems we have

$$\begin{aligned}
[U \circ T]_\alpha^\gamma &= \Big[[(U \circ T)(v_1)]_\gamma \ \cdots \ [(U \circ T)(v_n)]_\gamma \Big] \\
&= \Big[[U(T(v_1))]_\gamma \ \cdots \ [U(T(v_n))]_\gamma \Big] \\
&= \Big[[U]_\beta^\gamma [T(v_1)]_\beta \ \cdots \ [U]_\beta^\gamma [T(v_n)]_\beta \Big] \\
&= [U]_\beta^\gamma \Big[[T(v_1)]_\beta \ \cdots \ [T(v_n)]_\beta \Big] \\
&= [U]_\beta^\gamma [T]_\alpha^\beta
\end{aligned}$$

which completes the proof. ■

Corollary 3.6 *If V is an n-dimensional vector space over F with an ordered basis β and $I \in \mathcal{L}(V)$ is the identity operator, then $[I]_\beta = I_n \in M_n(F)$.*

Proof: By Theorem 3.5, given any $T \in \mathcal{L}(V)$, $T = I \circ T$, so that $[T]_\beta = [I \circ T]_\beta = [I]_\beta [T]_\beta = I_n [T]_\beta$, so that $[I]_\beta = I_n$ by the cancellation law. ■

Theorem 3.7 *If V and W are finite-dimensional vector spaces of dimensions n and m, respectively, with ordered bases β and γ, respectively, and $T \in \mathcal{L}(V, W)$, then*

$$\operatorname{rank}(T) = \operatorname{rank}([T]_\beta^\gamma) \qquad \text{and} \qquad \operatorname{null}(T) = \operatorname{null}([T]_\beta^\gamma) \tag{3.12}$$

Proof: This follows from Theorem 3.4, since ϕ_γ and ϕ_β are isomorphisms, and so carry bases into bases, and $\phi_\gamma \circ T = T_A \circ \phi_\beta$, so that

$$\begin{aligned}
\operatorname{rank}(T) = \dim\big(\operatorname{im}(T)\big) = \dim(\phi_\gamma(\operatorname{im}(T))\big) = \operatorname{rank}(\phi_\gamma \circ T) \\
= \operatorname{rank}(T_A \circ \phi_\beta) = \dim\big(\operatorname{im}(T_A \circ \phi_\beta))\big) = \dim\big(\operatorname{im}(T_A)\big) \\
= \operatorname{rank}(T_A) = \operatorname{rank}(A) = \operatorname{rank}([T]_\beta^\gamma)
\end{aligned}$$

while the second equality follows from the rank nullity theorem along with the fact that the ranks are equal. ■

Theorem 3.8 *If V and W are finite-dimensional vector spaces over F with ordered bases $\beta = (b_1, \ldots, b_n)$ and $\gamma = (c_1, \ldots, c_m)$, then the function*

$$\Phi : \mathcal{L}(V, W) \to M_{m,n}(F) \tag{3.13}$$

$$\Phi(T) = [T]_\beta^\gamma \tag{3.14}$$

is an isomorphism,

$$\Phi \in \operatorname{GL}\big(\mathcal{L}(V, W),\ M_{m,n}\big) \tag{3.15}$$

and consequently

$$\mathcal{L}(V, W) \cong M_{m,n}(F) \tag{3.16}$$

and

$$\dim\big(\mathcal{L}(V, W)\big) = \dim\big(M_{m,n}(F)\big) = mn \tag{3.17}$$

Proof: First, $\Phi \in \mathcal{L}\big(\mathcal{L}(V,W), M_{m,n}\big)$: there exist scalars $r_{ij}, s_{ij} \in F$, for $i = 1, \dots, m$ and $j = 1, \dots, n$, such that for any $T \in \mathcal{L}(V,W)$ we have

$$\begin{array}{ccc} T(b_1) = r_{11}c_1 + \cdots + r_{m1}c_m & & U(b_1) = s_{11}c_1 + \cdots + s_{m1}c_m \\ \vdots & \text{and} & \vdots \\ T(b_n) = r_{1n}c_1 + \cdots + r_{mn}c_m & & U(b_n) = s_{1n}c_1 + \cdots + s_{mn}c_m \end{array}$$

Hence, for all $k, l \in F$ and $j = 1, \dots, n$ we have

$$(kT + lU)(b_j) = kT(b_j) + lU(b_j) = k\sum_{i=1}^{m} r_{ij}c_i + l\sum_{i=1}^{m} s_{ij}c_i = \sum_{i=1}^{m}(kr_{ij} + ls_{ij})c_i$$

As a consequence of this and the rules of matrix addition and scalar multiplication we have

$$\begin{aligned} \Phi(kT + lU) &= [kT + lU]_\beta^\gamma \\ &= \begin{pmatrix} kr_{11} + ls_{11} & \cdots & kr_{1n} + ls_{1n} \\ \vdots & \ddots & \vdots \\ kr_{m1} + ls_{m1} & \cdots & kr_{mn} + ls_{mn} \end{pmatrix} \\ &= k\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix} + l\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \\ &= k[T]_\beta^\gamma + l[U]_\beta^\gamma \\ &= k\Phi(T) + l\Phi(U) \end{aligned}$$

Moreover, Φ is bijective by Theorem 2.5, since $\forall A \in M_{m,n},\ \exists! T \in \mathcal{L}(V,W)$ such that $\Phi(T) = A$, because there exists a unique $T \in \mathcal{L}(V,W)$ such that

$$T(b_j) = A_{1j}c_1 + \cdots + A_{nj}c_m \qquad \text{for } \ j = 1, \dots, n$$

This makes Φ onto, and also 1-1 because $\ker(\Phi) = \{T_0\}$ because for $O \in M_{m\times n}(F)$ there is only $T_0 \in \mathcal{L}(V,W)$ such that $\Phi(T_0) = O$, because

$$T(b_j) = 0c_1 + \cdots + 0c_m = 0 \qquad \text{for } \ j = 1, \dots, n$$

defines a unique transformation, T_0. ■

Theorem 3.9 *If V and W are finite-dimensional vector spaces over the same field F with ordered bases $\beta = (b_1, \dots, b_n)$ and $\gamma = (c_1, \dots, c_n)$, respectively, then $T \in \mathcal{L}(V,W)$ is an isomporhpism iff $[T]_\beta^\gamma$ is invertible, in which case*

$$[T^{-1}]_\gamma^\beta = \big([T]_\beta^\gamma\big)^{-1} \tag{3.18}$$

Proof: If T is an isomorphism and $\dim(V) = \dim(W) = n$, then $T^{-1} \in \mathcal{L}(W,V)$, and $T \circ T^{-1} = I_W$ and $T^{-1} \circ T = I_V$, so that by Theorem 3.5, Corollary 3.6 and Theorem 3.1 we have

$$[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = [T \circ T^{-1}]_\gamma = [I_W]_\gamma = I_n = [I_V]_\beta = [T^{-1} \circ T]_\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma$$

so that $[T]_\beta^\gamma$ is invertibe with inverse $[T^{-1}]_\gamma^\beta$, and by the uniqueness of the multiplicative inverse in $M_n(F)$, which follows from the uniqueness of $T_A^{-1} \in \mathcal{L}(F^n)$, we have

$$[T^{-1}]_\gamma^\beta = \big([T]_\beta^\gamma\big)^{-1}$$

Conversely, if $A = [T]_\beta^\gamma$ is invertible, there is a $n \times n$ matrix B such that $AB = BA = I_n$, so by Theorem 2.5 there is a unique $U \in \mathcal{L}(W, V)$ such that $U(c_j) = v_j = \sum_{i=1}^n B_{ij} b_j$. Hence $B = [U]_\gamma^\beta$. To show that $U = T^{-1}$, note that

$$[U \circ T]_\beta = [U]_\gamma^\beta [T]_\beta^\gamma = BA = I_n = [I_V]_\beta \quad \text{and} \quad [T \circ U]_\gamma = [T]_\beta^\gamma [U]_\gamma^\beta = AB = I_n = [I_W]_\gamma$$

But since $\Phi \in \big(\mathcal{L}(V, W), M_{m,n}(F)\big)$ is an isomorphism, and therefore 1-1, we must have that

$$U \circ T = I_V \quad \text{and} \quad T \circ U = I_W$$

By the uniqueness of the inverse, however, $U = T^{-1}$, and T is an isomorphism. ■

4 Change of Coordinates

4.1 Definitions

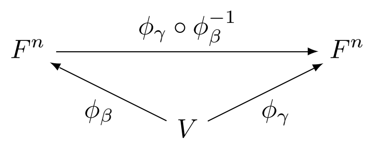
We can also now define the **change of coordinates** or **change of basis, operator**. If V is a finite-dimensional vector space, say $\dim(V) = n$, and $\beta = (b_1, \ldots, b_n)$ and $\gamma = (c_1, \ldots, c_n)$ are two ordered bases for V, then the coordinate maps $\phi_\beta, \phi_\gamma \in \mathcal{L}(V, F^n)$, which are isomorphisms by Theorem 2.15, may be used to define a change of coordinates operator $\phi_{\beta,\gamma} \in \mathcal{L}(F^n)$ changing β coordinates into γ coordinates, that is having the property

$$\phi_{\beta,\gamma}([v]_\beta) = [v]_\gamma \tag{4.1}$$

We define the operator by

$$\phi_{\beta,\gamma} := \phi_\gamma \circ \phi_\beta^{-1} \tag{4.2}$$

The relationship between these three functions is illustrated in the following commutative diagram:

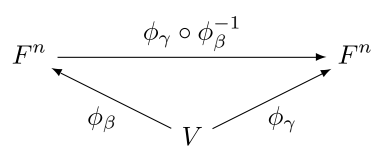


The change of coordinates operator has a matrix representation

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho = [\phi_\gamma]_\beta^\rho = \begin{bmatrix} [b_1]_\gamma & [b_2]_\gamma & \cdots & [b_n]_\gamma \end{bmatrix} \tag{4.3}$$

4.2 Change of Coordinate Maps and Matrices

Theorem 4.1 (Change of Coordinate Matrix) *Let V be a finite-dimensional vector space over a field F and $\beta = (b_1, \ldots, b_n)$ and $\gamma = (c_1, \ldots, c_n)$ be two ordered bases for V. Since ϕ_β and ϕ_γ are isomorphisms, the following diagram commutes,*



and the change of basis operator $\phi_{\beta,\gamma} := \phi_\gamma \circ \phi_\beta^{-1} \in \mathcal{L}(F^n)$, changing β coordinates into γ coordinates, is an automorphism of F^n. It's matrix representation,

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho \in M_n(F) \tag{4.4}$$

where $\rho = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the standard ordered basis for F^n, is called the **change of coordinate matrix***, and it satisfies the following conditions:*

1. $$\begin{aligned} M_{\beta,\gamma} &= [\phi_{\beta,\gamma}]_\rho \\ &= [\phi_\gamma]_\beta^\rho \\ &= \begin{bmatrix} [b_1]_\gamma & [b_2]_\gamma & \cdots & [b_n]_\gamma \end{bmatrix} \end{aligned}$$
2. $[v]_\gamma = M_{\beta,\gamma}[v]_\beta, \quad \forall v \in V$
3. $M_{\beta,\gamma}$ *is invertible and* $\quad M_{\beta,\gamma}^{-1} = M_{\gamma,\beta} = [\phi_\beta]_\gamma^\rho$

Proof: The first point is shown as follows:

$$\begin{aligned} M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho &= \Big[\phi_{\beta,\gamma}(\mathbf{e}_1) \;\; \phi_{\beta,\gamma}(\mathbf{e}_2) \;\cdots\; \phi_{\beta,\gamma}(\mathbf{e}_n)\Big] \\ &= \Big[(\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_1) \;\; (\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_2) \;\cdots\; (\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_n)\Big] \\ &= \Big[(\phi_\gamma \circ \phi_\beta^{-1})([b_1]_\beta) \;\; (\phi_\gamma \circ \phi_\beta^{-1})([b_2]_\beta) \;\cdots\; (\phi_\gamma \circ \phi_\beta^{-1})([b_n]_\beta)\Big] \\ &= \Big[\phi_\gamma(b_1) \;\; \phi_\gamma(b_2) \;\cdots\; \phi_\gamma(b_n)\Big] \\ &= \Big[[b_1]_\gamma \;\; [b_2]_\gamma \;\cdots\; [b_n]_\gamma\Big] \\ &= [\phi_\gamma]_\beta \end{aligned}$$

or by Theorem 3.5 and Theorem 3.9 we have that

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho = [\phi_\gamma \circ \phi_\beta^{-1}]_\rho = [\phi_\gamma]_\beta^\rho [\phi_\beta^{-1}]_\rho^\beta = [\phi_\gamma]_\gamma^\rho ([\phi_\beta]_\beta^\rho)^{-1} = [\phi_\gamma]_\beta^\rho I_n^{-1} = [\phi_\gamma]_\beta^\rho I_n = [\phi_\gamma]_\beta^\rho$$

The second point follows from Theorem 3.4, since

$$\phi_\gamma(v) = (\phi_\gamma \circ I)(v) = (\phi_\gamma \circ (\phi_\beta^{-1} \circ \phi_\beta))(v) = ((\phi_\gamma \circ \phi_\beta^{-1}) \circ \phi_\beta)(v)$$

implies that

$$[v]_\gamma = [\phi_\gamma(v)]_\rho = [((\phi_\gamma \circ \phi_\beta^{-1}) \circ \phi_\beta)(v)]_\rho = [\phi_\gamma \circ \phi_\beta^{-1}]_\rho [\phi_\beta(v)]_\rho = [\phi_{\beta,\gamma}]_\rho [v]_\beta = M_{\beta,\gamma}[v]_\beta$$

And the last point follows from the fact that ϕ_β and ϕ_γ are isomorphism, so that $\phi_{\beta,\gamma}$ is an isomorphism, and hence $\phi_{\beta,\gamma}^{-1} \in \mathcal{L}(F^n)$ is an isomorphism, and because the diagram above commutes we must have

$$\phi_{\beta,\gamma}^{-1} = (\phi_\gamma \circ \phi_\beta^{-1})^{-1} = \phi_\beta \circ \phi_\gamma^{-1} = \phi_{\beta,\gamma}$$

so that by 1

$$M_{\beta,\gamma}^{-1} = [\phi_{\beta,\gamma}^{-1}]_\rho = [\phi_{\gamma,\beta}]_\rho = M_{\gamma,\beta}$$

or alternatively by Theorem 3.9

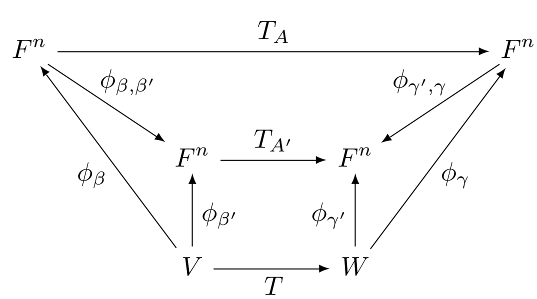
$$M_{\beta,\gamma}^{-1} = ([\phi_{\beta,\gamma}]_\rho)^{-1} = [\phi_{\beta,\gamma}^{-1}]_\rho = [\phi_{\gamma,\beta}]_\rho = [\phi_\beta]_\gamma^\rho = M_{\gamma,\beta}$$ ■

Corollary 4.2 (Change of Basis) *Let V and W be finite-dimensional vector spaces over the same field F and let (β,γ) and (β',γ') be pairs of ordered bases for V and W, respectively. If $T \in \mathcal{L}(V,W)$, then*

$$[T]_{\beta'}^{\gamma'} = M_{\gamma,\gamma'}[T]_\beta^\gamma M_{\beta',\beta} \tag{4.5}$$

$$= M_{\gamma,\gamma'}[T]_\beta^\gamma M_{\beta,\beta'}^{-1} \tag{4.6}$$

where $M_{\gamma,\gamma'}$ and $M_{\beta',\beta}$ are change of coordinate matrices. That is, the following diagram commutes



Proof: This follows from the fact that if $\beta = (b_1, \ldots, b_n)$, $\beta' = (b'_1, \ldots, b'_n)$, $\gamma = (c_1, \ldots, c_m)$, $\gamma = (c'_1, \ldots, c'_m)$, then for each $i = 1, \ldots, n$ we have

$$[T(b'_i)]_{\gamma'} = [(\phi_{\gamma,\gamma'} \circ T \circ \phi^{-1}_{\beta,\beta'})(b'_i)]_{\gamma'}$$

so

$$\begin{aligned}
[T]^{\gamma'}_{\beta'} &= [\phi_{\gamma,\gamma'} \circ T \circ \phi^{-1}_{\beta,\beta'}]^{\gamma'}_{\beta'} \\
&= [\phi_{\gamma,\gamma'}]_{\rho_m} [T]^{\gamma}_{\beta} [\phi^{-1}_{\beta,\beta'}]_{\rho_n} \\
&= [\phi_{\gamma,\gamma'}]_{\rho_m} [T]^{\gamma}_{\beta} ([\phi_{\beta,\beta'}]_{\rho_n})^{-1} \\
&= M_{\gamma,\gamma'} [T]^{\gamma}_{\beta} M^{-1}_{\beta,\beta'}
\end{aligned}$$

which completes the proof. ■

Corollary 4.3 (Change of Basis for a Linear Operator) *If V is a finite-dimensional vector space with ordered bases β and γ, and $T \in \mathcal{L}(V)$, then*

$$[T]_{\gamma} = M_{\beta,\gamma} [T]_{\beta} M^{-1}_{\beta,\gamma} \tag{4.7}$$

where $M_{\beta,\gamma}$ is the change of coordinates matrix. ■

Corollary 4.4 *If we are given any two of the following:*

(1) *$A \in M_n(F)$ invertible*

(2) *an ordered basis β for F^n*

(3) *an ordered basis γ for F^n*

the third is uniquely determined by the equation $A = M_{\beta,\gamma}$, where $M_{\beta,\gamma}$ is the change of coordinates matrix of the previous theorem.

Proof: If we have $A = M_{\beta,\gamma} = [\phi_\beta]_\gamma = \Big[[b_1]_\gamma \;\; [b_2]_\gamma \;\; \cdots \;\; [b_n]_\gamma\Big]$, suppose we know A and γ. Then, $[b_i]_\gamma$ is given by A, so $b_i = A_{i1}c_1 + \cdots + A_{in}c_n$, so β is uniquely determined. If β and γ are given, then by the previous theorem $M_{\beta,\gamma}$ is given by $M_{\beta,\gamma} = \Big[[b_1]_\gamma \;\; [b_2]_\gamma \;\; \cdots \;\; [b_n]_\gamma\Big]$. Lastly, if A and β are given, then γ is given by the first case applied to $A^{-1} = M_{\gamma,\beta}$. ■