Finite Dimensional Inner Product Spaces

Jason Kenyon

November 2022

Contents

Preface

The following is my attempt at a summary of the theory of finite dimensional inner product spaces. That is, finite dimensional vector spaces equipped with a sesquilinear innner product. I have tried my best to include only the essential developments required to understand such spaces and omit excessive details, which serve only to distract the reader. Additionally, I aim to motivate certain topics in functional analysis by treating the infinite dimensional case in some instances. When appropriate, I also comment on which results extend, or fail to extend, to infinite dimensional vector spaces, as well as provide some inuition as to why. Primarily, however, these are notes on linear algebra.

Throughout, I assume familiarity with the elements of Zermlo-Fraenkel-Choice set theory, and work within this context; however, I rarely appeal directly to the axioms thereof; rather, I take a naive approach. In addition, the reader should have a cursory understanding of algebraic structures such as groups and fields, and in particular, the properties of the complex field.

Chapter 1

Vector Spaces

1.1 Spaces and Subspaces

Definition 1.1.1. A vector space over a field $\mathbb{F}$ is a set V, along with two binary functions $+ : V \times V \to V$ and $\cdot : \mathbb{F} \times V \to V$ such that for all $a, b \in \mathbb{F}$ and $v, w, z \in V$

1. $a \cdot v + w \in V$.
2. $v + w = w + v$.
3. $v + (w + z) = (v + w) + z$.
4. $\mathbf{1}v = v$.
5. $(a \cdot b)x = a \cdot (bx)$.
6. $a \cdot (v + w) = av + aw$.
7. $(a + b)v = a \cdot v + b \cdot v$.
8. There exists an element $\mathbf{0} \in V$ such that $v + \mathbf{0} = v$.
9. There exists an element $-v \in V$ such that $v + (-v) = \mathbf{0}$.

The elements of V are called vectors, and the elements of $\mathbb{F}$ are called scalars. Any vector $\mathbf{0} \in V$ that satisfies condition 8 is called a **zero vector**, and any vector $-v \in V$ that satisfies 9 is called an **additive inverse** of

v. Moreover, we adopt the convention that any vector denoted by $\mathbf{0}$ satisfies condition 8 and any vector denoted by $-v$ for some $v \in V$ satisfies condition 9. The functions $+$ and $\cdot$ are called vector addition and scalar multiplication, respectively. We omit the $\cdot$ and do not explicitly apply $+$ for clarity. For the sake of brevity, we will refer to a vector space $(V, \mathbb{F}, +, \cdot)$ as V when there is no ambiguity. Unless otherwise specified, we assume that any vector space is over an arbitrary field $\mathbb{F}$. Note that $\mathbf{1}$ will denote the identity element in $\mathbb{F}$, and for all $a \in \mathbb{F}$ $-a$ will denote the additive inverse of a. We shall now establish some foundational properties of vector addition and scalar multiplication.

Theorem 1.1.1. *Let V be a vecor space. Then for all $x, y, z \in V$ and $a \in \mathbb{F}$*

(i) *If $x + z = y + z$ then $x = y$ for all $x, y, z \in V$.*

(ii) *The zero vector is unique.*

(iii) *The additive inverse of x is unique.*

(iv) $0x = \mathbf{0}$.

(v) $(-a)x = -(ax) = a(-x)$.

(vi) $a\mathbf{0} = \mathbf{0}$.

Definition 1.1.2. A subspace W of a vector space V over a field $\mathbb{F}$ is a subset $W \subseteq V$ that is a vector space over $\mathbb{F}$.

Definition 1.1.3. Let V be a vector space. A linear combination of a set $\{v_1, v_2, \ldots, v_n\} \subseteq V$ is a sum of the form

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

where $a_i \in \mathbb{F}$ for all $1 \leq i \leq n$. Furthermore, we define the span of such a set

$$\operatorname{span}(\{v_1, v_2, \ldots, v_n\})$$

to be the set of all linear combinations thereof.

Theorem 1.1.2. *Let V be a vector space. Then W is a subspace of V if and only if $\mathbf{0} \in W$ and $cx + y \in W$ for all $x, y \in W$ and $a \in \mathbb{F}$.*

Theorem 1.1.3. *Let S be a subset of a vector space V. Then* $\operatorname{span}(S)$ *is a subspace of V, and any subspace of V contains S must necessarily contain* $\operatorname{span}(S)$

The proof follows directly from the definition of span. The span is defined precisely to generate a subspace. Additionally, it should be clear that any linear combination of vectors in a subspace must be contained in that subspace, as this is the defining characteristic of a vector space.

1.2 Linear Independence

Definition 1.2.1. A set of vectors $\{v_1, v_2, \ldots, v_n\}$ is linearly dependent if

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = \mathbf{0}$$

for $a_1, a_2, \ldots, a_n \in \mathbb{F}$ not all zero. Similarly, a set of vectors is linearly independent if it is not linearly dependent.

Theorem 1.2.1. *If $S_1 \subseteq S_2 \subseteq V$ and S_1 is linearly dependent, then S_2 is linearly dependent as well. Similarly, if S_2 is linearly dependent, then S_1 is linearly dependent.*

The proof should be clear when considering the above definition.

Theorem 1.2.2. *Let S be a linearly independent subset of a vector space V, and let $v \in V \backslash S$. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \operatorname{span}(S)$.*

Proof. If $S \cup \{v\}$ is linearly dependent then there exist scalars $a_1, a_2, \ldots a_n, a_v \in \mathbb{F}$ not all zero such that

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n + a_vv = \mathbf{0}.$$

Therefore, $a_v \neq 0$, for otherwise we would contradict the linear independence of S. This implies that

$$v = -\frac{a_1s_1 + a_2s_2 + \cdots a_ns_n}{a_v}.$$

and hence $v \in \operatorname{span}(S)$. Conversely, if $v \in \operatorname{span}(S)$, then

$$v = a_1s_1 + a_2s_2 + \cdots a_ns_n$$

for some scalars $a_1, a_2, \cdots \in \mathbb{F}$ This implies that

$$1(v) - (a_1s_1 + \cdots a_ns_n) = \mathbf{0}.$$

which is a nontrivial solution, so the set $S \cup \{v\}$ is linearly dependent. ∎

1.3 Bases

Definition 1.3.1. A subset $\beta \subseteq V$ of a vector space V is a basis for V if it is a linearly independent set such that $\text{span}(\beta) = V$.

Definition 1.3.2. A partial order $\leq$ on a class X is a binary relation on X such that for all $x, y, z \in X$

1. $x \leq x$.
2. $x \leq y$ or $y \leq x$.
3. If $x \leq y$ and $y \leq x$ then $x = y$.

If, in addition, we have that

$$x \leq y \text{ and } y \leq z \text{ implies } x \leq z \text{ we call } \leq \text{ a total order on } X.$$

If there exists an element $n \in X$ such that for all $x \in X$ $x \leq n$ we call n a maximal element of X. If $Y \subseteq X$ and there exists an element $m \in X$ such that for all $y \in X$ $y \leq m$ we call m an upper bound of Y.

Notice that $\subseteq$ is a partial order on the class of all sets.

Theorem 1.3.1 (Zorn's Lemma)**.** *Let X be a partially ordered set such that any totally ordered $Y \subseteq X$ has an upper bound in X. Then X contains at least one maximal element.*

Theorem 1.3.2. *Every vector space has a basis.*

Proof. Consider the set L of all linearly independent subsets of a vector space V. Let $T \subseteq L$ be a chain. That is, for any two sets A and B in T either $A \subseteq B$ or $B \subseteq A$. Hence, any finite subset of $\bigcup T$ is in L. In other words, taking a union over a chain yields an upper bound under $\subseteq$ which must necessarily be in the set from whence it came. This ensures that T is totally ordered by $\subseteq$, for transitivity, reflexivity, and antisymmetry are already satisfied by definition of a subset. Therefore, Zorn's lemma implies that there exists a maximal element in L. That is, there exists an element $l \in L$ such that for all $A \in L$ $A \subseteq l$. Moreover, we know that l is linearly independent by assumption.

To show that l spans V, suppose that there were an element $v \in V$ such that $v \notin \text{span}(l)$. Then by theorem 1.2.2 $l \cup \{v\}$ would be a linearly independent set, in which case $l \cup \{v\} \in L$. But $l \cup \{v\} \not\subseteq l$, contradicting the fact that l is the maximal element of L. ∎

Theorem 1.3.3. *A subset $\beta = \{v_1, v_2, \ldots v_n\}$ of V is a basis for V if and only if for any vector $v \in V$*

$$v = a_1 v_1 + \cdots a_n v_n$$

for unique scalars $a_1, \ldots a_n \in \mathbb{F}$.

Proof. Suppose that $\beta = \{v_1, \ldots v_n\}$ is a linearly independent generating set of V. Then $v = a_1 v_1 + \cdots a_n v_n$ for scalars $a_1, \ldots a_n \in \mathbb{F}$. Further, suppose that there exists another collection $b_1, \ldots b_n$ of scalars such that $v = b_1 v_1 + \cdots b_n v_n$. Subtracting, we have

$$(a_1 - b_1)v_1 \cdots (a_n - b_n)v_n = \mathbf{0}.$$

Since β is linearly independent, it follows that $a_i - b_i = 0$, and hence $a_i = b_i$ for all $1 \leq i \leq n$. Therefore, the linear combination $a_1 v_1 + \cdots a_n v_n$ is the unique representation of V for β. Similarly, if we know that $v = a_1 v_1 \cdots a_n v_n$ for unique scalars, then

$$(b_1)v_1 + \cdots (b_n)v_n = \mathbf{0} = v - v.$$

if and only if $b_i = a_i - a_i = 0$ for all $1 \leq i \leq n$. And certainly $V = \text{span}(\beta)$, so β is a basis for V. ■

Corollary 1.3.1. *If V is generated by a finite set, then there exists a finite basis for V contained within the generating set.*

Proof. Suppose that $\text{span}(S) = V$ for a finite set S. Consider an arbitrary linearly independent subset $\beta \subseteq S$ such that $\beta \cup \{v\}$ is linearly dependent for any $v \in S$ such that $v \notin \beta$. Such a set certainly exist because any set containing a single vector is linearly independent, and so we may continue to add vectors from S into β until another union results in a linearly dependent set. Hence if we demonstrate that $S \subseteq \text{span}(\beta)$ we will have that $\text{span}(S) \subseteq \text{span}(\beta)$, and we already know that $\text{span}(\beta) \subseteq V$. To show this, note that for any $v \in S$ if $v \in \beta$ then trivially $v \in \text{span}(\beta)$, and if $v \notin \beta$, then by assumption $\beta \cup \{v\}$ is linearly dependent, in which case $v \in \text{span}(\beta)$ by theorem 1.2.2. ■

Theorem 1.3.4. *Let V be a vector space generated by a set G containing n vectors, and $L \subseteq V$ be linearly independent containing m vectors. Then $m \leq n$ and there exists a subset $H \subseteq G$ containing $n - m$ vectors such that* $\text{span}(L \cup H) = V$.

Proof. We proceed by induction on m. For $m = 0$ $L = \emptyset \subseteq V$ and $0 \leq n$ for all $n \in \mathbb{N}$. Taking $H = G$ we are done. So suppose our theorem is true for any linearly independent set with $m-1$ vectors. Now consider an arbitrary linearly independent subset of V, $L = \{v_1, v_2, \ldots v_m\}$. The set $\{v_1, v_2, \ldots v_{m-1}\} \subseteq L$ is then linearly independent, and so by our induction hypothesis, $m - 1 \leq n$ and there is a subset $\{h_1, h_2, \cdots h_{n-(m-1)}\}$ of G such that $\operatorname{span}(\{v_1, v_2, \ldots v_{m-1}\} \cup \{h_1, h_2, \cdots h_{n-(m-1)}\}) = V$. That is

$$v_m = a_1 v_1 + \cdots a_{m-1} v_{m-1} + b_1 h_1 + \cdots b_{n-(m-1)} h_{n-(m-1)}$$

for $a_i, b_i \in \mathbb{F}$. And $n - (m-1) \neq 0$, for otherwise L would not be linearly independent by theorem 1.2.2. This means that $n - (m-1) > 0$, or, $n > (m-1)$, from which it follows that $m \leq n$. Moreover, there exists some $b_i \neq 0$ as otherwise we would, once again, contradict the linear independence of L. Without loss of generality, we have

$$h_1 = \frac{v_m - (a_1 v_1 + a_2 v_2 + \cdots a_{m-1} v_{m-1} + b_2 h_2 + \cdots b_{n-(m-1)} h_{n-(m-1)})}{b_1}.$$

It follows that $h_1 \in \operatorname{span}(L \cup \{h_2, \ldots, h_{n-(m-1)}\})$, in which case,

$$\{v_1, \ldots v_m, h_1, \ldots h_{n-(m-1)}\} \subseteq \operatorname{span}(L \cup \{h_2, \ldots h_{n-(m-1)}\}).$$

But by our induction hypothesis, $\operatorname{span}(\{v_1, \ldots v_m, h_1, \ldots h_{n-(m-1)}\}) = V$, and hence,

$$\operatorname{span}(L \cup \{h_2, \ldots h_{n-(m-1)}\}) = V.$$

since $\{h_2, \ldots h_{n-(m-1)}\}$ is a subset of G that contains $n-(m-1)-1 = n-m$ vectors, we have demonstrated the theorem for L with m vectors. ∎

Corollary 1.3.2. *If a vector space V is generated by a finite basis then any basis for V is finite and of equal cardinality.*

Proof. Let β and γ be bases for V with m and n vectors respectively. We have that $m \leq n$ and $n \leq m$ by theorem 1.3.4. ∎

Thus we may safely define the dimension of a vector space:

Definition 1.3.3. The dimension of a vectors space V, denoted $\dim(V)$, is the unique cardinality of any basis for V.

Corollary 1.3.3. *Suppose that V is a vector space with dimension n. Then any linearly independent subset of V containing n vectors is a basis for V. And any generating set for V contains at least n vectors. Additionally, any linearly independent subset of V can have at most n vectors.*

Corollary 1.3.4. *Let $W \subseteq V$ be a subspace. Then $\dim(W) \leq \dim(V)$, and if $\dim(W) = \dim(V)$ then $V = W$.*

1.4 Direct Sum and Projections

Definition 1.4.1. Let V be a vector space with subspaces W_1 and W_2 such that $W_1 \cap W_2 = \{\mathbf{0}\}$. Then we call the direct sum of W_1 and W_2

$$W_1 \oplus W_2 = \{w_1 + w_2 : w_1 \in W_1 \text{ and } w_2 \in W_2\}.$$

Theorem 1.4.1. *Let $V = W_1 \oplus W_2$ for some subspaces W_1, W_2 of V. If $\{v_1, \ldots v_k\}$ is a basis for W_1 and $\{w_1, \ldots w_l\}$ is a basis for W_2 then $\{v_1, \ldots v_k, w_1, \ldots w_l\}$ is a basis for V. Hence $\dim(V) = \dim(W_1) + \dim(W_2)$.*

Proof. Let $x \in V$ and $x = x_1 + x_2$ for $x_1 \in W_1$ and $x_2 \in W_2$. Then

$$x = a_1 v_1 + \cdots a_k v_k + b_1 w_1 + \cdots b_l w_l$$

If $x = \mathbf{0}$ then it must be that $x_1 = -x_2$. Hence $x_1 = x_2 = \mathbf{0}$ for otherwise $-x_2 \in W_1$ and $-x_2 \notin W_2$, which would contradict our assumption that W_2 is a subspace of V. Moreover, since $\{v_1, \ldots v_k\}$ and $\{w_1, \ldots w_l\}$ are both linearly indpendent, it must be that $a_i = 0$ for all $1 \leq i \leq k$ and $b_i = 0$ for all $1 \leq i \leq l$. ■

Definition 1.4.2. If $V = W_1 \oplus W_2$ then a projection of V on W_1 along W_2 is a linear function $T : V \to V$ such that for any $x \in V$ where $x = x_1 + x_2$ $x_1 \in W_1$ and $x_2 \in W_2$ $T(x) = x_1$.

Theorem 1.4.2. *A linear function $T : V \to V$ is a projection of V on $W_1 = \{x : T(x) = x\}$ along $\ker T$ if and only if $T = T^2$.*

Proof. We have, for all $x \in V$ $x = x_1 + x_2$ where $x_1 \in W_1$ and $x_2 \in \ker(T)$ so that

$$\begin{aligned} TT(x) &= TT(x_1 + x_2) && (1.1)\\ &= Tx_1 && (1.2)\\ &= x_1 && (1.3)\\ &= Tx && (1.4) \end{aligned}$$

Conversely, if $T^2 = T$, we know that for all $x \in V$ $Tx = Tx + (x - Tx)$. This implies that $T(Tx) = Tx + \mathbf{0}$. Since $T(Tx) = Tx$, $Tx \in \{y : Ty = y\}$. Furthermore, If $Tx = x$ and $Tx = \mathbf{0}$ for some $x \in V$ then $x = \mathbf{0}$. That is

$$\{y : Ty = y\} \cap \ker T = \{\mathbf{0}\}$$

hence $V = \{y : Ty = y\} \oplus \ker T$ and so for any $x \in V$ we have $x = x_1 + x_2$ for some $x_1 \in \{y : Ty = y\}$ and $x_2 \in \ker(T)$ whence

$$Tx = x_1$$

■

Chapter 2

Linear Functions

2.1 Linearity

Definition 2.1.1. A function $f : V \to W$ between two vector spaces V and W is linear if

$$f(ax + y) = af(x) + f(y)$$

for all $x, y \in V$ and $a \in \mathbb{F}$.

The following properties of linear functions go without saying:

1. $f(\mathbf{0}) = \mathbf{0}$
2. $f(\sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} a_i f(x_i)$

It follows that linear functions are unique up to their values on a given basis.

Corollary 2.1.1. *Let $f : V \to W$ and $g : V \to W$ be linear and $\{v_1, \ldots, v_n\}$ be a basis for V. Then $f = g$ if and only if $f(v_i) = g(v_i)$.*

Definition 2.1.2. For a linear function $f : V \to W$ we define

$$\mathrm{im}(f) = \{y : f(x) = y \text{ for some } x \in V\}$$

and

$$\ker(f) = \{x \in V : f(x) = \mathbf{0}\}.$$

Theorem 2.1.1. *Let $f : V \to W$ be a linear function. Then $\ker(f)$ and $\mathrm{im}(f)$ are subspaces of V and W respectively.*

Proof. We begin with $\ker(f)$. Surely, $\ker(f) \subseteq V$, so suppose that $x, y \in \ker(f)$ and $a \in \mathbb{F}$. We have

$$f(x) = f(y) = \mathbf{0}$$

hence

$$af(x) + f(y) = f(ax + y) = \mathbf{0}$$

by linearity. Additionally, we know that $f(\mathbf{0}) = \mathbf{0}$. Thus, $ax + y, \mathbf{0} \in \ker(f)$, so by 1.1.2 we are done.

Now suppose that $x, y \in \operatorname{im}(f)$ and $a \in \mathbb{F}$ Then for some $x_0, y_0 \in V$, $f(x_0) = x$ and $f(y_0) = y$. Therefore,

$$af(x_0) + f(y_0) = f(ax_0 + y_0) = ax + y \in \operatorname{im}(f).$$

Furthermore,

$$f(\mathbf{0}) = \mathbf{0} \in \operatorname{im}(f).$$

■

Theorem 2.1.2. *Let $f : V \to W$ be linear and $\beta = \{v_1, v_2, \ldots v_n\}$ be a basis for V. Then*

$$\operatorname{im}(f) = \operatorname{span}(f(\beta)).$$

Proof. Let $x \in V$. We have $x = \sum_{i=1}^{n} a_i v_i$ for $a_i \in \mathbb{F}$ and $f(x) = \sum_{i=1}^{n} a_i f(v_i)$. That is, for an arbitrary element $f(x) \in \operatorname{im}(f)$ $f(x) \in \operatorname{span}(f(\beta))$. The converse containment follows by the same logic. ■

Theorem 2.1.3 (Dimension Theorem)**.** *Let $f : V \to W$ be linear. Then*

$$\dim(\ker(f)) + \dim(\operatorname{im}(f)) = \dim(V).$$

Proof. Let $\{v_1, \ldots v_k\}$ be a basis for $\ker(f)$. Then we may extend this basis to a basis $\{v_1, v_2, \ldots v_k, v_{k+1}, \ldots v_n\}$ for V. Now, by 2.1.2

$$\operatorname{im}(f) = \operatorname{span}(f(\{v_1, \ldots v_n\}))$$

but since $\{v_1, \ldots v_k\} \subseteq \ker(f)$ we have

$$\operatorname{im}(f) = \operatorname{span}(f(v_{k+1}, \ldots v_n)).$$

To show that this set is, indeed a basis, for $\mathrm{im}(f)$, suppose that

$$\sum_{i=k+1}^{n} a_i f(v_i) = \mathbf{0}.$$

The linearity of f yields

$$f(\sum_{i=k+1}^{n} a_i v_i) = \mathbf{0}$$

which is to say that

$$\sum_{i=k+1}^{n} a_i v_i \in \ker(f).$$

Thus we may represent this vector in the basis of $\ker(f)$. We have

$$\sum_{i=k+1}^{n} a_i v_i - \sum_{i=1}^{k} b_i v_i = \mathbf{0}$$

which implies that $a_i = 0$ because we know that $\{v_1, \dots v_n\}$ is a basis for V. Therefore $\dim(\mathrm{im}(f)) = \dim(V) - \dim(\ker(f))$. ■

Theorem 2.1.4. *Let $f : V \to W$ be linear. Then f is injective if and only if* $\ker(f) = \{\mathbf{0}\}$.

Proof. Suppose that f is injective and that $f(x) = \mathbf{0}$ for some $x \in V$. We have that $f(x) = f(\mathbf{0}) = \mathbf{0}$ so $x = \mathbf{0}$. Conversely, suppose $\ker(f) = \{\mathbf{0}\}$. Then if $f(x) = f(y)$ we know that $f(x - y) = \mathbf{0}$, and hence $x - y = \mathbf{0}$. ■

Theorem 2.1.5. *Let $f : V \to W$ be linear. If* $\dim(V) = \dim(W)$ *then the following statements are equivalent:*

1. *f is injective*
2. *f is surjective*
3. $\dim(\mathrm{im}(f)) = \dim(V)$

Proof. Applying, theorem 2.1.3 and theorem 2.1.4 we have f is injective if and only if $\ker(f) = \{\mathbf{0}\}$ if and only if $\dim(\ker(f)) = 0$ if and only if $\dim(\mathrm{im}(f)) = \dim(V) = \dim(W)$. And by corollary 1.3.4 $\mathrm{im}(f) = W$. ■

2.2 Matrices

Definition 2.2.1. Let V be a vector space with a basis $\{v_1, \dots v_n\}$. An ordered basis for V is a permutation of the $n - tuple$ $(v_1, \dots v_n)$.

Definition 2.2.2. Let $f : V \to W$ be a linear function between two vector spaces V and W and let $\beta = (v_1, \dots v_n)$ and $\gamma = (w_1, \dots w_m)$ be ordered bases for V and W respectively. Suppose that $f(v_j) = \sum_{i=1}^{m} a_{ij} w_i$. Then we call the $m \times n$ array with the scalar a_{ij} in the i^{th} row and j^{th} column thereof the matrix representation of f with respect to ordered bases β and γ. We denote this by

$$[f]_\beta^\gamma.$$

Furthermore, given $x \in V$ with $x = \sum_{i=1}^{n} b_i v_i$ we call the $n \times 1$ matrix whose i^{th} row is b_i the column vector of x with respect to β. Analogously we may define the row vector of x with respect to β.

Corollary 2.2.1. *Let A be an $m \times n$ matrix over $\mathbb{F}$. The mapping $L_A : \mathbb{F}^n \to \mathbb{F}^m$ defined by $A \to Ax$ for $x \in \mathbb{F}^n$ is linear and*

$$[L_A]_{e_n}^{e_m} = A$$

where e_n and e_m are the standard ordered bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively.

Proof. The linearity of L_A follows from theorem 2.2.3 The j^{th} column of $[L_A]_{e_n}^{e_m}$ is $L_A(e_j) = Ae_j$ which is the j^{th} column of A. ∎

Definition 2.2.3. $\delta_{ij} : X \to \{0, 1\}$ is the map with

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

e_j is the column vector with

$$(e_j)_{ij} = \delta_{ij}.$$

The tuple $(e_1, e_2, \dots e_n)$ is called the standard ordered basis for the vector space $\mathbb{F}^n$. The $m \times n$ matrix

$$[Id_V]_\beta^\beta$$

is called the $n \times n$ identity matrix.

Note that the above definition is well founded as one may easily verify that $\mathbb{F}^n$ forms a vector space in the natural way, with vector addition and scalar multiplication defined coordinate-wise. Additionally, one can easily verify that the $n \times n$ identity matrix is the matrix whose j^{th} column is e_j.

Theorem 2.2.1. *Let V and W be vector spaces with ordered bases β and γ respectively, and let $f, g \in \hom(V, W)$. Then the following hold:*

1. $[f+g]^\gamma_\beta = [f]^\gamma_\beta + [g]^\gamma_\beta$
2. $[af]^\gamma_\beta = a[f]^\gamma_\beta$

The proof follows from a direct application of the definition of a matrix.

Definition 2.2.4. Let A be an $m \times n$ matrix and B be an $n \times p$ matrix, then the product of A and B denoted AB is the $n \times p$ matrix defined by

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

Theorem 2.2.2. *Let $f : V \to W$ be a linear function between two vector spaces V and W and $g : W \to Z$ be a linear function between W and a vector space Z. Let $\alpha\beta\gamma$ be ordered bases for VWZ respectively. Then*

$$[f(g)]^\gamma_\alpha = [f]^\gamma_\beta [g]^\beta_\alpha.$$

Proof. Let $\alpha = (v_1, \ldots v_n)$, $\beta = (w_1, \ldots w_m)$ and $\gamma = (z_1, \ldots z_p)$. We have

$$f(g(v_j)) = f(\sum_{k=1}^{m} B_{kj} w_k) = \sum_{k=1}^{m} B_{kj} f(w_k) = \sum_{k=1}^{m} B_{kj} (\sum_{i=1}^{p} A_{ik} z_i) = \sum_{i=1}^{p} (\sum_{k=1}^{m} A_{ik} B_{kj}) z_i$$

■

Corollary 2.2.2. *Let $f : V \to W$ be linear and β and γ be ordered basis for V and W respectively. Then for all $v \in V$*

$$[f(v)]_\gamma = [f]^\gamma_\beta [v]_\beta.$$

Proof. Define $g : \mathbb{F} \to V$ and $h : \mathbb{F} \to W$ by $f(a) = av$ and $h(a) = af(v)$ for all $a \in \mathbb{F}$. Let $\alpha = \{\mathbf{1}\}$. Then by theorem 2.2.2

$$[f(v)]_\gamma = [h(\mathbf{1})]_\gamma = [h]^\gamma_\alpha = [f(g)]^\gamma_\alpha = [f]^\gamma_\beta [g]^\beta_\alpha = [f]^\gamma_\beta [g(1)]_\beta = [f]^\gamma_\beta [v]_\beta.$$

■

Theorem 2.2.3. *Let A be an $m \times n$ matrix B and C be $n \times p$ matrices and D and E be $q \times m$ matrices. Then*

1. $A(B + C) = AB + AC$ *and* $(D + E)A = DA + EA$
2. $a(AB) = (aA)B = A(aB)$
3. $I_m A = A = AI_n$

The proof follows from a direct application of the definition 2.2.4.

Theorem 2.2.4. *Let A B and C be matrices such that $A(BC)$. Then $AB(C)$ is defined and*

$$A(BC) = AB(C).$$

The proof follows by a direct application of matrix multiplication. Analogous results hold for linear functions, all of which follow from the definition of a linear function.

2.3 Isomorphism

Definition 2.3.1. Let $(X, f_1, f_2, \ldots, f_n)$ and $(Y, g_1, g_2, \ldots, g_n)$ be two algebraic structures and $h : X \to Y$ be a function between them. We say that f is a homomorphism if for all $x, y \in X$ and $1 \leq i \leq n$

$$h(f_i(x, y)) = g_i(h(x), h(y)).$$

If, in addition, h is bijective, then we say that f is an isomorphism.

Definition 2.3.2. Let V and W be vector spaces over $\mathbb{F}$. The set of all homomorphisms between V and W is defined

$$\hom(V, W).$$

and the dual space of V is defined as

$$V^* = \hom(V, \mathbb{F}).$$

Note that $\hom(V, W)$ is precisely the set of all linear maps between V and W and forms a vector space by virtue of the linearity of each of its elements.

Theorem 2.3.1. *Let V and W be vector spaces over $\mathbb{F}$ with dimension n and m respectively. Let $\beta = \{v_1, v_2, \ldots v_n\}$ and $\gamma = \{w_1, w_2, \ldots w_m\}$ be ordered bases for V and W respectively. Then the map $\psi_{\beta,\gamma} : \hom(V, W) \to M_{m\times n}(\mathbb{F})$ defined by*

$$\psi_{\beta,\gamma}(T) = [T]_\beta^\gamma$$

is an ismorphism.

Proof. Let $A \in M_{m\times n}$ By theorem 2.2.1, $\psi_{\beta,\gamma}$ is linear. Since linear maps are unique up to their values attained on a basis, there exists a unique linear map $T : V \to W$ such that

$$T(v_j) = \sum_{i=1}^{m} A_{ij} w_i$$

for all $1 \leq j \leq n$. That is to say that for any matrix $A \in M_{m\times n}(\mathbb{F})$ there exists a unique linear map $T \in \hom(V, W)$ such that $\psi_{\beta,\gamma}(T) = [T]_\beta^\gamma = A$. ∎

It follows that we can uniquely associate arrays with matrices, and matrices with linear functions. Thus, we may phrase all of our results on linear functions in terms of multiplying arrays of numbers.

Corollary 2.3.1. *Let $f : V \to W$ be linear. Then $[f]_\beta^\gamma$ is invertible if and only if f is invertible. And $([f]_\beta^\gamma)^{-1} = [f^{-1}]_\gamma^\beta$*

Proof. If $[f]_\beta^\gamma$ is invertible then $[f]_\beta^\gamma A = A[f]_\beta^\gamma = I_n$ And for some linear function S $B = [S]_\gamma^\beta$ so

$$[f]_\beta^\gamma [S]_\gamma^\beta = [f(s)]_\gamma = I_n = [Id_W]_\gamma$$

and

$$[S]_\gamma^\beta [f]_\beta^\gamma = [S(f)]_\beta = I_n = [Id_V]_\beta.$$

That is, $f(S) = Id_W$ and $S(f) = Id_V$, whence $S = f^{-1}$.

Conversely if f is invertible, then

$$f^{-1}(f) = Id_V$$

so

$$[f^{-1}(f)]_\beta = [Id_V]_\beta = I_n = [f^{-1}]_\gamma^\beta [f]_\beta^\gamma.$$

Similarly,

$$[f]_\beta^\gamma [f^{-1}]_\gamma^\beta = I_n.$$

∎

Theorem 2.3.2. *Let V and W be vector spaces over a field. Then V is isomorphic to W if and only if* $\dim(V) = \dim(W)$.

Proof. If $T : V \to W$ is an isomorphism then T is, by definition, bijective and linear, and hence by theorem 2.1.3

$$\dim(\mathrm{im}(T)) = \dim(W) = \dim(V).$$

Conversely, if $\dim(V) = \dim(W)$ and $\beta = \{v_1, \ldots v_n\}$ and $\gamma = \{w_1, \ldots w_n\}$. By theorem **??** there is a unique linear map such that $T(v_i) = w_i$. Furthermore by theorem 1.1.3

$$\mathrm{im}(T) = \mathrm{span}(T(\beta)) = \mathrm{span}(\gamma) = W.$$

∎

Therefore, we have shown that an arbitrary vector space is isomorphic to some $\mathbb{F}^n$, the cannonical, most intuitive vector space there is, hence demystifying the idea of an abstract vector space. This is a common theme in linear algebra.

2.4 Change of Basis

Theorem 2.4.1. *Let β and β' be bases for a vector space V. And let*

$$Q = [Id_V]_{\beta'}^{\beta}.$$

Then Q is invertible and

$$[v]_\beta = Q[v]_{\beta'}$$

for any $v \in V$.

Proof. By theorem 2.3.1 Q is invertible. Moreover

$$[v]_\beta = [Id_V(v)]_\beta = [Id_V]_{\beta'}^{\beta}[v]_{\beta'} = Q[v]_{\beta'}$$

by corollary 2.2.2. ∎

Corollary 2.4.1. *Let $A \in M_n(\mathbb{F})$ and $\gamma = (v_1, v_2, \ldots v_n)$ be an ordered basis for $\mathbb{F}^n$.If $Q \in M_n(\mathbb{F})$ is the matrix whose j^{th} column is v_j then*

$$[L_A]_\gamma = Q^{-1}AQ.$$

In this case, we say that the matrix A is similar to the matrix Q.

Chapter 3

Linear Systems of Equations

3.1 Rank

Definition 3.1.1. An elementary row or column operation on an $m \times n$ matrix A is defined as one of the following:

1. Interchanging any two rows or columns of A
2. Scaling each entry in a row or or column of A
3. Adding a multiple of one row or column to another row or column of A

An elementary matrix is the result of applying one of the above to the $n \times n$ identity matrix.

Theorem 3.1.1. *Suppose that B is the result of applying an elementary row operation to A. Then there exists an elementary matrix E such that $B = EA$. Furthermore, E is the matrix obtained by performing the same elementary row operation to I_n as was performed to convert A into B. Similarly, if B is the result of applying an elementary column operation to A, then there exits an elementary matrix E such that $B = AE$, and E is the result of applying the same elementary column operation to I_m as was applied to A.*

The proof is a tedious verification of cases; the elementary matrices are defined precisely for this to work.

Definition 3.1.2. The rank of a matrix A is defined as the rank of the linear function $L_A = Ax$

Theorem 3.1.2. *Let $T : V \to W$ be an isomorphism and $V_0 \subseteq V$ be a subspace of V. Then $T(V_0) \subseteq W$ is a subspace of W. Moreover* $\dim(V_0) = \dim(T(V_0))$

Proof. If $V_0 \subseteq V$ is a subspace of V then $T(V_0)$ is a subspace of W because T is linear. Further, we may consider the map $T' : V_0 \subseteq T(V_0)$ such that $T'(x) = T(x)$ for all $x \in V_0$. By theorem 2.1.3 we have

$$\dim(\ker(T')) + \dim(\mathrm{im}(T')) = 0 + \dim(T(V_0)) = \dim(V_0).$$

■

Theorem 3.1.3. *Let $T : V \to W$ be linear and $A = [T]_\beta^\gamma$. Then* $\mathrm{rank}(T) = \mathrm{rank}(L_A)$

Proof. Consider the map $\phi_\beta : V \to \mathbb{F}^n$. That is, the function mapping a vector to its representation in coordinates. This is linear by definition and invertible as we know that any basis represents a vector uniquely as a linear combination of its elements. We have

$$L_A(\mathbb{F}^n) = L_A\phi_\beta(V) = \phi_\gamma(T(V)).$$

It follows, by theorem 3.1.2, that

$$\dim(\mathrm{im}(L_A)) = \dim(\mathrm{im}(T))$$

because ϕ_γ is an isomorphism. ■

Theorem 3.1.4. *Let A be an $m \times n$. Let P and Q be invertible $m \times m$ and $n \times n$ matrices, respectively. Then*

1. $\mathrm{rank}(AQ) = \mathrm{rank}(A)$
2. $\mathrm{rank}(PA) = \mathrm{rank}(A)$
3. $\mathrm{rank}(PAQ)$

Proof.

$$\begin{aligned} \mathrm{im}(L_{AQ}) &= \mathrm{im}(L_A L_Q) && (3.1)\\ &= L_A L_Q(\mathbb{F}^n) && (3.2)\\ &= L_A(L_Q((\mathbb{F}^n)) && (3.3)\\ &= L_A(\mathbb{F}^n) && (3.4)\\ &= \mathrm{im}(L_A) && (3.5) \end{aligned}$$

Thus, $\operatorname{rank}(L_{AQ}) = \operatorname{rank}(L_A)$. Similarly, $\operatorname{im}(L_P L_A) = L_P(\operatorname{im}(L_A)) = \operatorname{im}(L_A)$ and so $\dim(\operatorname{im}(L_P L_A)) = \dim(\operatorname{im}(L_A))$ since P is an isomorphism. It follows, by applying the previous two results that $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$. ■

Theorem 3.1.5. *Let*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{mn} \end{pmatrix}.$$

Then $\operatorname{rank}(A) = \dim \left(\operatorname{span} \left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}, \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\} \right)$

Proof.

$$\begin{aligned} \operatorname{im}(L_A) &= L_A(\mathbb{F}^n) && (3.6) \\ &= L_A(\operatorname{span}\{e_1, \dots e_n\}) && (3.7) \\ &= \operatorname{span}\{Ae_1, \dots, Ae_n\} && (3.8) \\ &= \operatorname{span} \left\{ \begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}, \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\} && (3.9) \end{aligned}$$

Furthermore, $\dim(\operatorname{span}(X))$ is nothing but the number of linearly independent vectors in X for any set of vectors X. Thus we have shown that the rank of a matrix is nothing but the number of linearly independent vectors in its columns. ■

Theorem 3.1.6. *Let A be an $m \times n$ matrix. Then a finite composition of elementary row and column operations applied to A results in a matrix of the form*

$$\begin{pmatrix} I_{\operatorname{rank}(A)} & O_1 \\ O_2 & O_3 \end{pmatrix}$$

where O_1, O_2, O_3 are zero matrices.

Proof. First, note that if A is a zero matrix, then by theorem 3.1.5 $\operatorname{rank}(A) = 0$, and so $A = I_0$, the degenerate case of our claim. Suppose otherwise. We proceed by induction on m, the number of rows of A. In the case that $m = 1$, we may convert A to a matrix of the form

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

by first making the leftmost entry 1 and adding the corresponding additive inverses of the others to the other columns. Clearly the rank of the above matrix is 1 and is of the form

$$\begin{pmatrix} I_1 & O \end{pmatrix}$$

This is another degenerate case, as it lacks zeros below the identity. Now suppose that our theorem holds when A has $m - 1$ rows.

To demonstrate that our theorem holds when A is an $m \times n$ matrix, notice that when $n = 1$, we can argue that our theorem holds as before, but using row operations instead of column operations. This is another degenerate case. For $n > 0$, note that there exists an entry $A_{ij} \neq 0$ and by applying at most an elementary row and column operation, we can move A_{ij} to position $1, 1$. Additionally, we may transform A_{ij} to value 1, and as before, transform all of the entries in row and column 1 besides A_{ij} to 0. Thus we have a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_{11} & \cdots & x_{1\ n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_{m-1\ 1} & \cdots & x_{m-1\ n-1} \end{pmatrix}$$

■

The submatrix defined by x_{ij} is of dimension $m - 1 \times n - 1$ and so must have rank $\text{rank}(A) - 1$ as elementary opertations preserve rank and deleting a row and column of a matrix reduces its rank by 1. Furthermore, by our induction hypothesis the above matrix may be converted via a finite number of elementary operations to a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & I_{\text{rank}(A)-1} & O_1 & \\ \vdots & & & \\ 0 & O_2 & O_3 & \end{pmatrix}$$

Therefore, for an $m \times n$ matrix A, a finite number of elementary operations converts it into a matrix of the form

$$\begin{pmatrix} I_{\mathrm{rank}(A)} & O_1 \\ O_2 & O_3 \end{pmatrix}$$

.

Theorem 3.1.7. *For any matrix A,* $\mathrm{rank}(A^T) = \mathrm{rank}(A)$.

Proof. By theorem 3.1.6, we may convert A to a matrix $D = BAC$ where $B = E_1 \cdots E_p$ and $C = G_1 \cdots G_q$ where E_i and G_i are elementary row and column matrices respectively. It follows that $D^T = C^T A^T B^T$, whence $\mathrm{rank}(A^T) = \mathrm{rank}(D^T)$ by theorem (insert) because elementary matrices are invertible, and so is the transpose of the compositions thereof. Further, D^T must be of the same form as D since the only nonzero entries of D are along the diagonal from entry $1, 1$ to entry $\mathrm{rank}(A), \mathrm{rank}(A)$. Hence, we have $\mathrm{rank}(A)$ linearly independent columns in the matrix D^T.

Since the columns of D^T are the rows of D, we see that the number of linearly independent columns of A is equal to the number of linearly independent columns of A^T. In other words, the dimension of the space generated by the columns of A is equal to the dimension of the space generated by its rows. ∎

Theorem 3.1.8. *Let A be an invertible $n \times n$ matrix. Then A is a product of elementary matrices.*

Proof. By the dimension theorem, if A is invertible, then $\mathrm{rank}(A) = n$. So by theorem 3.1.6 A may converted into a matrix of the form $I_n = E_1 \cdots E_p A G_1 \cdots G_q$, whence $A = E_1^{-1} \cdots E_p^{-1} I_n G_1^{-1} \cdots G_q^{-1}$. ∎

Theorem 3.1.9. *Let $T : V \to W$ and $U : W \to Z$. Then*

1. $\mathrm{rank}(TU) \leq \mathrm{rank}(U)$

2. $\mathrm{rank}(TU) \leq \mathrm{rank}(T)$

Proof. We have

$$\begin{aligned} \mathrm{rank}(TU) &= \dim(\mathrm{im}(TU)) && (3.10) \\ &= \dim(\mathrm{im}(T(U(V)))) && (3.11) \\ &\subseteq U(W) && (3.12) \\ &= \mathrm{im}(U) && (3.13) \end{aligned}$$

Therefore, $\dim(\mathrm{im}(TU)) \leq \dim(\mathrm{im}(U))$. Next, let β, γ, ϕ be ordered bases for V, W, and Z, respectively; and let $A = [T]_\beta^\gamma$ and $B = [U]_\gamma^\phi$. By theorem 3.1.7

$$\begin{aligned}
\dim(\mathrm{im}(TU)) &= \dim(\mathrm{im}(AB)) && (3.14)\\
&= \dim(\mathrm{im}((AB)^T)) && (3.15)\\
&= \dim(\mathrm{im}(B^T A^T)) && (3.16)\\
&\leq \dim(\mathrm{im}(A^T)) && (3.17)\\
&= \dim(\mathrm{im}(A)) && (3.18)\\
&= \dim(\mathrm{im}(T)) && (3.19)
\end{aligned}$$

■

3.2 Form

We now apply the fruits of our investigation into vector spaces and linearity to solve systems of linear equations.

Definition 3.2.1. A linear system of equations is a collection of m equations of the form:

$$a_1 x_1 + \cdots + a_n x_n = b$$

where $a_i, x_i, b \in \mathbb{F}$ for $1 \leq i \leq n$. Equivalently, we may say $Ax = b$ for an $m \times n$ matrix A, where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$. If $b = \mathbf{0}$, the linear system is said to be homogenous.

Definition 3.2.2. A solution to a linear system is a vector $s \in \mathbb{F}^n$ such that $As = b$

Theorem 3.2.1. *Let A be an $m \times n$ matrix over $\mathbb{F}$. If $m < n$, then the homogenous system $Ax = 0$ has a nontrivial solution.*

Proof. Notice that, the solution set to the system $Ax = 0$ is $\ker(L_A)$, so by the dimension theorem, $\dim(\ker(A)) = n - \mathrm{rank}(L_A)$. Additionally, we know that $\mathrm{rank}(A)$ is nothing but the number of linearly independent vectors defined by its rows which certainly cannot exceed m. Therefore $\mathrm{rank}(A) \leq m < n$, in which case $n - \mathrm{rank}(A) = \dim(\ker(A)) > 0$, and so $\ker(A) \neq \{0\}$. ■

Theorem 3.2.2. *For any solution s to the linear system $Ax = b$,*

$$\{s + s_0 : As_0 = \mathbf{0}\}$$

is its solution set.

Proof. Suppose that $As = b$ and $As' = b$. Then $A(s' - s) = As' - As = b - b = 0$. It follows that $s + (s' - s) \in S$. Conversely, if $y \in S$, then $y = s + s'$, in which case $Ay = A(s + s') = As + As' = b + 0 = b$. That is, $Ay = b$. ∎

Theorem 3.2.3. *Let $Ax = b$ for an $n \times n$ matrix A. If A is invertible, then the system has a single solution $A^{-1}b$. If the system has a single solution, then A is invertible.*

Proof. Suppose A is invertible. Then $A(A^{-1}b) = AA^{-1}(b) = b$. Furthermore, if $As = b$ for some $s \in \mathbb{F}^n$, then $A^{-1}(As) = A^{-1}b$ and so $s = A^{-1}b$. Next, suppose that the system has a unique solution s. Then by theorem 3.2.2, we know that the solution set $S = \{s + s_0 : As_0 = 0\}$. But this is only the case if $\ker(A) = \{0\}$, lest s not be unique. And so, by the dimension theorem, A is invertible. ∎

Theorem 3.2.4. *The linear system $Ax = b$ has a nonempty solution set if and only if* $\text{rank}(A) = \text{rank}(A|b)$.

Proof. If the system has a solution, then $b \in \text{im}(L_A)$. Additionally, $\text{im}(L_A) = L_A(F^n)$ and $L_A(e_i) = Ae_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$. Therefore, since $L_A(\mathbb{F}^n) = \text{span}\{Ae_1, \dots Ae_n\}$, $\text{im}(L_A) = \text{span}\{A_1, \dots A_n\}$, where A_i is the i^{th} column of A. Certainly, $b \in \text{span}\{A_1, \dots A_n\}$ if and only if $\text{span}\{A_1, \dots A_n\} = \text{span}\{A_1, \dots A_n, b\}$, which is to say $\dim(\text{im}(\text{span}\{A_1, \dots A_n\})) = \dim(\text{im}(\text{span}\{A_1, \dots A_n, b\}))$, or, $\text{rank}(A) = \text{rank}(A|b)$. ∎

Corollary 3.2.1. *Let $Ax = b$ be a linear system of m equations in n variables. Then its solution set is either, empty, of one element, or of infinitely many elements (provided that $\mathbb{F}$ is not a finite field).*

Proof. By theorem 3.2.4 $Ax = b$ has a nonempty solution set if and only if $\text{rank}(A) = \text{rank}(A|b)$. Therefore, it may be that our linear system has no solutions; however, supposing that this is not the case, by theorem 3.2.3 it has a unique solution if and only if A is invertible. Finally, assume that our linear system has neither no solution nor a single solution. This yields

$$Ax_1 = Ax_2 = b \tag{3.20}$$

for $x_1, x_2 \in \mathbb{F}^n$, which implies

$$\begin{aligned} Ax_1 - Ax_2 &= \mathbf{0} && (3.21)\\ &= A(x_1 - x_2) && (3.22)\\ &= nA(x_1 - x_2) && (3.23)\\ &= A(n(x_1 - x_2)) && (3.24)\\ & && (3.25) \end{aligned}$$

where $n \in \mathbb{F}$. Thus, by theorem 3.2.2

$$A(x_1 + n(x_1 - x_2)) = b.$$

■

3.3 Solution

Definition 3.3.1. A matrix of the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

is said to be in reduced echelon form if

1. $a_{ii} \neq 0$ implies that $a_{ij} = 1$
2. $a_{ij} \neq 1$ implies that $a_{ij} = 0$
3. $a_{ij} = 0$ for all $1 \leq j \leq n$ implies that $i < r$ for all nonzero rows $\begin{pmatrix} a_{r1} & \cdots & a_{rn} \end{pmatrix}$

Theorem 3.3.1. *Any matrix can be converted into reduced echelon form via a finite number of elementary row operations.*

Proof. This is a restatement of theorem 3.1.6. ■

This form is of particular interest because reducing an augmented matrix is equivalent to solving a linear system of equations. We now have a procedure for solving arbitrary systems of linear equations. For example, we may now demonstrate that a set of vectors is linearly dependent by finding a nontrivial solution to a linear system of equations; similarly we may apply theorem 3.2.4 to demonstrate that a set of vectors is linearly dependent. In the following chapter, we will also see that computing the elements of an eigenspace is made possible by reducing a matrix. It follows that

Corollary 3.3.1. *For any invertible $n \times n$ matrix A.*

$$A^{-1}(A|I_n) = E_1 \cdots E_p(A|I_n) = (I_n|A^{-1})$$

where $E_1, \ldots, E_p$ are elementary matrices.

Notice that the above elementary matrices may be either row or column matrices; however, since we are left multiplying, the product will result in a row operation. Thus we now have a procedure for finding the inverse of any matrix: perform row operations to convert it into the identity matrix, while accounting for each change. Additionally,

Corollary 3.3.2. *Let A be an $m \times n$ matrix and C be an invertible $n \times n$ matrix. Then the solutions sets to the linear systems*

$$Ax = bandCAx = Cb$$

are equal.

This follow directly from the invertibility, and fits with our intuition: as we row reduce a linear system, its solutions do not change.

Chapter 4

The Determinant

4.1 Permuations

define determinant show equal to cofactor expansion

4.2 Cofactor Expansion

deduce enough properties to define the determinat more formally

4.3 Multilinear and Alternating

demonstrate cofactor expansion is unqiue multilinear alternating etc hence permutation=cofactor=unique such function

4.4 Properties

det of block matrix
deduce remaining important properties need invertible iff det nonzero

4.5 Measure

Chapter 5

Eigenspaces

5.1 Diagonalization and Similarity

Definition 5.1.1. Let T be a linear operator on a vector space V. We say T is diagonalizable if there exists an orderd basis β for V such that $[T]_\beta$ is a diagonal matrix. We say a matrix $A \in M_n(\mathbb{F})$ is diagonalizable if L_A is diagonalizable.

Definition 5.1.2. Let T be a linear operator on a vector space V. If

$$Tv = \lambda v$$

for some $v \in V$ and $\lambda in \mathbb{F}$ we call λ an eigenvalue of T and v an eigenvector of T.

Theorem 5.1.1. *If T is diagonalizable and β is the basis of V for which $D = [T]_\beta$ is diagonal, then β is a set of eigenvectors of T and the diagonal entries of D are eigenvalues of T. Similarly, if A is diagonalizable then*

$$A = QDQ^{-1}$$

for a diagonal matrix D and matrix Q whose columns are eigenvectors of A.

This is hardly a theorem. Our definitions are designed such that this is the case.

Theorem 5.1.2. *Let $A \in M_n(\mathbb{F})$. Then λ is an eigenvalue of A if and only if*

$$\det(A - \lambda I) = 0.$$

Proof. If $Av = \lambda v$ for some nonzero $v \in \mathbb{F}^n$ then $(A - \lambda I)(v) = \mathbf{0}$. That is to say that $\ker(A - \lambda I) \neq \{\mathbf{0}\}$ and hence $A - \lambda I$ is not invertible. However, this is only the case if $\det(A - \lambda I) = 0$. ■

Definition 5.1.3. We call

$$f(t) = \det(A - tI)$$

the characteristic polynomial of A.

Theorem 5.1.3. *Let $A \in M_n(\mathbb{F})$. Then the characteristic polynomial of A is a polynomial of degree n.*

This follows from a simple induction proof, applying cofactor expansion along the first row of A.

Corollary 5.1.1. *Any matrix $A \in M_n(\mathbb{F})$ has at most n distinct eigenvalues.*

This follows directly from the fundamental theorem of algebra.

Theorem 5.1.4. *Let T be a linear operator on a vector space V with distinct eigenvalues $\lambda_1, \dots \lambda_k$. Let S_i be a finite set of eigenvectors corresponding to λ_i. If each S_i is linearly independent then $\bigcup_{i=1}^k S_i$ is linearly independent.*

Proof. We proceed by induction on k. If $k = 1$ we are done. So suppose that our theorem holds for $k - 1$ and that there exist k distinct eigenvalues $\lambda_1, \dots \lambda_k$ of T. For each i let

$$S_i = \{v_{i1}, v_{i2}, \dots v_{in_i}\}$$

be a finite set of n_i eigenvectors corresponding to λ_i. Suppose

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} v_{ij} = \mathbf{0}$$

for some scalars $a_{ij} \in \mathbb{F}$. Applying $T - \lambda_k I$ to both sides yields

$$\sum_{i=1}^{k-1} \sum_{j=1}^{n_i} a_{ij}(\lambda_i - \lambda_k) v_{ij} = \mathbf{0}.$$

By the induction hypothesis, $a_{ij}(\lambda_i - \lambda_k) = 0$ for $1 \leq i \leq k-1$ and $1 \leq j \leq n_i$. Moreover, since λ_i and λ_k are distinct it must be that $a_{ij} = 0$ for $1 \leq i \leq k-1$ and $1 \leq j \leq n_i$. But this implies that $\cup_{i=1}^k S_i$ is linearly independent because S_k is linearly independent by assumption. ■

Corollary 5.1.2. *If T has n distinct eigenvalues then T is diagonalizable.*

Definition 5.1.4. Let λ be an eigenvalue of an operator. The multiplicity of λ is the largest integer k such that $(t-\lambda)^k$ is a factor of the characteristic polynomial of that operator.

Definition 5.1.5. Let λ be an eigenvalue of an operator T. Then the eigenspace of λ is defined as

$$E_\lambda = \{v : Tv = \lambda v\}.$$

Theorem 5.1.5. *Let T be a linear operator on V. And suppose that λ is an eigenvalue of T with multiplicity m. Then*

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

Proof. Let $\{v_1, \dots v_p\}$ be a basis for for E_λ and extend it to a basis $\beta = \{v_1, \dots v_p, \dots v_n\}$ for V. Since v_i is an eigenvector for T for $1 \leq i \leq p$ corresponding to λ

$$[T]_\beta = \begin{pmatrix} \lambda I_p & B \\ 0 & C \end{pmatrix}$$

The characteristic polynomial of T is then

$$\det((\lambda - t))I_p \cdot \det(C - tI_{n-p}) = (\lambda - t)^p \cdot \det(C - tI_{n-p}).$$

Hence the multiplicity of λ is at least p and so

$$\dim(E_\lambda) = p \leq m.$$

■

Theorem 5.1.6. *Let T be a linear operator on V with distinct eigenvalues $\lambda_1, \dots \lambda_k$ with multiplicities m_i. And let $\dim(V) = n$ Then*

1. *T is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for $1 \leq i \leq k$.*

2. *T is diagonalizable and if β_i is a basis for E_{λ_i} then $\bigcup_{i=1}^k \beta_i$ is a basis for V consisting of eigenvectors of T. In other words, $\bigoplus_{i=1}^k E_{\lambda_i} = V$.*

Proof. Firstly, suppose T is diagonalizable.Let β be a basis for V consisting of eigenvectors of T. Let $\beta_i = \beta \cap E_{\lambda_i}$. Certainly, $|\beta_i| \leq \dim(E_{\lambda_i})$ as β_i is a collection of linearly independent vectors of a vector space E_{λ_i}. Additionally, both the $|\beta_i|$ and m_i must sum to n for β_i are linearly independent and the degree of the characteristic polynomial is equal to the sum of the multiplicities of the eigenvalues of T. It follows that

$$n = \sum_{i=1}^{k} |\beta_i| \leq \sum_{i=1}^{k} \dim(E_{\lambda_i}) \leq \sum_{i=1}^{k} m_i = n.$$

Hence, $m_i = \dim(E_{\lambda_i})$. Conversely, if $\dim(E_{\lambda_i}) = m_i$ then if we let $\beta = \bigcup_{i=1}^{k} \beta_i$ for bases β_i of E_{λ_i} β is necessarily linearly independent by theorem 5.1.4. Moreover,

$$\sum_{i=1}^{k} \dim(E_{\lambda_i}) = \sum_{i=1}^{k} m_i = n = |\beta|.$$

Thus β is a basis for V consisting of eigenvectors of T. ■

Therefore, to prove that an operator T is diagonalizable, it suffices to show that the multiplicity of λ is equal to $\dim(\ker(T - \lambda I))$ for all eigenvalues λ of T.

5.2 Dimension

Chapter 6

Orthogonality

6.1 Inner Products

Definition 6.1.1. Let V be a vector space over a field $\mathbb{F}$. An inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is function such that for all $x, y, z \in V$ and $a \in \mathbb{F}$

1. $\langle ax + z, y \rangle = a\langle x, y \rangle + \langle z, y \rangle$.
2. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
3. $\langle x, x \rangle > 0$ whenever $x \neq \mathbf{0}$.

Theorem 6.1.1. *Let V be an inner product space. Then for all $x, y, z \in V$ and $c \in \mathbb{F}$*

1. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
2. $\langle x, cy \rangle = \bar{c}\langle x, y \rangle$.
3. $\langle \mathbf{0}, x \rangle = \langle x, \mathbf{0} \rangle = 0$.
4. *If* $\langle x, y \rangle = \langle x, z \rangle$ *then* $y = z$.

Proof.

$$\begin{aligned} \langle x, y + z \rangle &= \overline{\langle y + z, x \rangle} && (6.1) \\ &= \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} && (6.2) \\ &= \langle x, y \rangle + \langle x, z \rangle && (6.3) \end{aligned}$$

$$\langle x, cy\rangle = \overline{\langle cy, x\rangle} \tag{6.4}$$
$$= \overline{c\langle y, x\rangle} \tag{6.5}$$
$$= \bar{c}\overline{\langle y, x\rangle} \tag{6.6}$$
$$= \bar{c}\langle x, y\rangle \tag{6.7}$$

$$\langle \mathbf{0}, x\rangle = \langle 0(\mathbf{0}), x\rangle \tag{6.8}$$
$$= 0\langle \mathbf{0}, x\rangle \tag{6.9}$$
$$= 0 \tag{6.10}$$
$$= \overline{\langle x, 0(\mathbf{0})} \tag{6.11}$$
$$= \overline{0\langle x, \mathbf{0}\rangle} \tag{6.12}$$
$$\tag{6.13}$$

$\langle x, y-z\rangle = 0$ for all $x \in V$ so $\langle y-z, y-z\rangle = 0$, which implies $y-z=0$ and $y=z$. ■

Definition 6.1.2. A subset $\{v_1, \ldots v_n\}$ of an inner product space V is called orthogonal if $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$ .Moreover, if $\langle v_j, v_j\rangle = 1$ it is called orthonormal.

Definition 6.1.3. Let V be a vector space over a field $\mathbb{F}$.
A norm $||\cdot|| : V \to \mathbb{R}$ is a function such that for all $x, y \in V$ and $a \in \mathbb{F}$

1. $||x|| \geq 0$ and $||x|| = 0$ if and only if $x = \mathbf{0}$.
2. $||ax|| = |a| \cdot ||x||$.
3. $||x+y|| \leq ||x|| + ||y||$.

It should be clear that an inner product induces a norm

$$||x|| = \sqrt{\langle x, x\rangle}.$$

Theorem 6.1.2. *Let V be an inner product space and $\{v_1, \ldots v_n\} \subseteq V$ be orthogonal. Then for all $y \in \operatorname{span}(\{v_1, \ldots v_n\})$*

$$y = \sum_{i=1}^{n} \frac{\langle y, v_i\rangle}{||v_i||^2} v_i.$$

Proof. For all $y = \sum_{i=1}^{n} a_i v_i$

$$\langle y, v_j \rangle = \sum_{i=1}^{n} a_i \langle v_i, v_j \rangle = a_j \langle v_j, v_j \rangle = a_j ||v_j||^2.$$

■

Corollary 6.1.1. *Let V be an inner product space and $S \subseteq V$ be orthogonal. Then S is linearly independent.*

Theorem 6.1.3 (Graham-Schmidt Process)**.** *Let V be an inner product space and $S = \{w_1, \ldots w_n\}$ be a linearly indepedent subset of V. Let $S' = \{v_1, \ldots v_n\}$ where $v_1 = w_1$ and*

$$v_n = w_n - \sum_{j=1}^{n-1} \frac{\langle w_n, v_j \rangle}{||v_j||^2}.$$

Then S' is an orthogonal set of nonzero vectors such that $\operatorname{span}(S) = \operatorname{span}(S')$.

Proof. Suppose $n = 1$, then we are done. So suppose the theorem holds for $n - 1$. We now show that $S' = \{v_1, \ldots v_{n-1}, v_n\}$ satisfies the theorem where v_n is defined as stated above. If $v_n = \mathbf{0}$ then $w_n \in \operatorname{span}\{v_1, \ldots v_{n-1}\} = \operatorname{span}\{w_1, \ldots w_{n-1}\}$ contradicting our assumption that S is linearly independent. Moreover, for all $1 \leq i \leq n - 1$

$$\langle v_n, v_i \rangle = \langle w_n, v_i \rangle - \sum_{j=1}^{n-1} \frac{\langle w_n, v_j \rangle}{||v_j||^2} \langle v_j, v_i \rangle = \langle w_n, v_i \rangle - \frac{\langle w_k, v_i \rangle}{||v_i||^2} ||v_i||^2 = 0$$

by the induction hypothesis. Finally, by assumption, $\operatorname{span}(S') \subseteq \operatorname{span}(S)$. The orthogonality of S' implies that it S' is linearly independent so that $\dim(\operatorname{span}(S')) = \dim(\operatorname{span}(S)) = n$. Thus, $\operatorname{span}(S) \subseteq \operatorname{span}(S')$ and $\operatorname{span}(S') = \operatorname{span}(S)$ by corollary 1.3.4. ■

6.2 The Adjoint

Definition 6.2.1. Let $A \in M_{m \times n}(\mathbb{F})$. The matrix A^* defined by $(A^*)_{ij} = \overline{A}_{ji}$ is called the hermitian conjugate of A.

6.3 Orthogonal Projections

Definition 6.3.1. Let $W \subseteq V$. The orthogonal complement of W is defined as $W^{\perp} = \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W\}$.

Theorem 6.3.1. *Let $W \subseteq V$. Then for any $x \in V$ there exist unique vectors $x_w \in W$ and $x^{\perp} \in W^{\perp}$ such that $x = x_W + x^{\perp}$. In other words $V = W \oplus W^{\perp}$.*

Proof. Let $\{w_1, \ldots w_n\}$ be an orthonormal basis for W $x_W = \sum_{i=1}^{n} \langle x, w_i \rangle w_i$ and $x^{\perp} = x - x_W$. Certainly $x_W \in W$ and $x = x_W + x^{\perp}$. To show that $x^{\perp} \in W^{\perp}$ we have

$$\begin{aligned}
\langle x^{\perp}, w_j \rangle &= \langle x - x_W, w_j \rangle && (6.14)\\
&= \langle x - \sum_{i=1}^{n} \langle x, w_i \rangle w_i, w_j \rangle && (6.15)\\
&= \langle x, w_j \rangle - \sum_{i=1}^{n} \langle x, w_i \rangle \langle w_i, w_j \rangle && (6.16)\\
&= 0 && (6.17)
\end{aligned}$$

For uniqueness, suppose that $x = y + z$ for $y \in W$ and $z \in W^{\perp}$. Then $x_W + x^{\perp} = y + z$ and so

$$x_W - y = z - x^{\perp} \in W \cap W^{\perp}.$$

But $W \cap W^{\perp} = \{\mathbf{0}\}$ so $x_W = y$ and $x^{\perp} = z$. ■

Corollary 6.3.1. *For all $y \in W$*

$$||x - x_W|| \leq ||x - y||$$

Proof.

$$\begin{aligned}
||x - y||^2 &= ||x_W + x^{\perp} - y||^2 && (6.18)\\
&= ||(x_W - y) + x^{\perp}||^2 && (6.19)\\
&= ||x_W - y||^2 + ||x^{\perp}||^2 && (6.20)\\
&\geq ||x^{\perp}||^2 = ||x - x_W||^2. && (6.21)
\end{aligned}$$

■

Theorem 6.3.2. *The following statments are true*

1. $W^\perp$ *is a subspace of* V
2. $\dim(W^\perp) = \dim(V) - \dim(W)$

Proof. Firstly, note that $\langle \mathbf{0}, w\rangle = \mathbf{0}$ for all $w \in W$, so $\mathbf{0} \in W^\perp$. Furthermore, if $\langle w, c\rangle = 0$ for some $w \in W$ then $\langle aw, c\rangle = a\langle w, c\rangle = 0$ by linearity. Similarly, if $\langle w, a\rangle = 0$ and $\langle b, c\rangle = 0$ then $\langle w, a\rangle + \langle b, c\rangle = \langle w + b, c\rangle = 0$. Secondly, $V = W^\perp \oplus W$ implies that $\dim(V) = \dim(W^\perp) + \dim(W)$. ∎

Theorem 6.3.3. *Let* W *be a subspace of* $\mathbb{F}^n$ *with basis* $\beta = \{v_1, \dots v_m\} \subseteq \mathbb{F}^n$. *Let* $x \in \mathbb{F}^n$ *and* A *be the* $m \times n$ *matrix whose* j^{th} *column is* v_j. *Then the orthogonal projection of* x *on* W

$$x_W = A(A^*A)^{-1}A^*x.$$

Proof. We begin by demonstrating that $W^\perp = \ker A^*$. We have

$$A^*x = \begin{pmatrix} v_1^*x \\ \vdots \\ v_n^*x \end{pmatrix} = \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_n, x\rangle \end{pmatrix}.$$

Certainly $A^*x = \mathbf{0}$ if and only if $\langle v_i, x\rangle = 0$ for all $1 \le i \le n$. But that is to say $x \in W^\perp$, and so

$$\ker(A^*) = W^\perp.$$

Let $x = x_W + x_\perp$ be the orthogonal decomposition of x with respect to W. Note that $\operatorname{im}(A) = \operatorname{span}(\beta) = W$. Therefore, for some vector $c \in \mathbb{F}^n$ $Ac = x_W$, which means that $x - x_W = x - Ac \in W^\perp$. It follows that $A^*(x - Ac) = 0$ and so

$$A^*Ac = A^*x.$$

Thus, we see that $x_W = Ac$.

Now, we will show that A^*A is invertible. Suppose that $A^*Ac = \mathbf{0}$. Then by the above result, we have $A^*Ac = A^*\mathbf{0} = \mathbf{0}$. But this implies that $Ac = \mathbf{0}_W = \mathbf{0}$ and hence $c \in \ker(A)$. However, we know that the columns of A are linearly independent, so the only solution to $Ac = \mathbf{0}$ is the trivial solution $c = \mathbf{0}$. Therefore

$$\ker(A^*A) = \{\mathbf{0}\}.$$

Knowing this we may solve for c yielding

$$x_W = A(A^*A)^{-1}A^*x.$$

■

Definition 6.3.2. Let W be a subspace of $\mathbb{F}^n$ with basis $\{v_1, \dots v_m\}$. Let A be the $m \times n$ matrix whose j^{th} column is v_j. We call the orthogonal projection operator on W

$$P_W = A(A^*A)^{-1}A^*$$

Corollary 6.3.2. *P_W is the unique projection of V on $\mathrm{im}(P_W) = W = \{x \in V : P_W x = x\}$ along $W^\perp = \ker P_W$.*

Proof. Surely P_W is linear, and since we know that $x = x_W + x_{W^\perp}$ for all $x \in V$ it follows that $(P_W)^2 x = P_W x_W = x_w = P_W x$. Thus the orthogonal projection is, in fact, a projection on $W = \{x \in V : AA^*x = x\}$ along $W^\perp = \ker(AA^*)$, by theorem 1.4.2 ($V = W \oplus W^\perp$). Since we know that the orthogonal decomposition of a vector with respect to a given subspace is unique, it follows that P_W is unique. Furthermore, we have

$$(W^\perp)^\perp = \ker(P_W)^\perp = W = \mathrm{im}(P_W).$$

■

Theorem 6.3.4. *Let T be an operator on an inner product space V. If T is an orthogonal projection, then*

$$T^2 = T = T^*$$

Proof. If T is in orthogonal projection, we know that $T^2 = T$. Thus, we must only show that T^* exists and $T = T^*$. We have $V = \mathrm{im}(T) \oplus \ker(T)$ and $\mathrm{im}(T)^\perp = \ker(T)$. Hence for $x = x_1 + x_2$ and $y = y_1 + y_2$ with $x_1, y_1 \in \mathrm{im}(T)$ and $x_2, y_2 \in \ker(T)$

$$\begin{aligned}
\langle x, Ty \rangle &= \langle x_1 + x_2, y_1 \rangle && (6.22)\\
&= \langle x_1, y_y \rangle + \langle x_2, y_2 \rangle && (6.23)\\
&= \langle x_1, y_1 \rangle && (6.24)\\
&= \langle Tx, y \rangle && (6.25)
\end{aligned}$$

■

6.4 Normal and Unitary Operators

Chapter 7

Matrix Decomposition

7.1 Schur's Theorem

7.2 Spectral Theorem

7.3 Singular Value Decomposition and Pseudo-inverse

7.4 Definiteness

Definition 7.4.1. An $n \times n$ matrix A with complex entries is called positive-definite if

$$x^* A x > 0$$

for all $x \in \mathbb{C}^n \backslash \{\mathbf{0}\}$. Similarly, A is called negative-definite if

$$x^* A x < 0$$

for all $x \in \mathbb{C}^n \backslash \{\mathbf{0}\}$. We replace positive-definite and negative-definite with positive semi-definite and negative semi-definite in the case that the inequality is not strict.

Theorem 7.4.1. *Let $A \in M_n(\mathbb{F})$. Then the following statements are equivalent*

1. *A is positive definite*

2. *All the eigenvalues of A are real and positive*

3. $A = B^*B$ *for some* $B \in M_n(\mathbb{F})$

Proof. Suppose A is positive definite, and let $v \in \mathbb{F}^n$ be an eigenvector of A with eigenvalue λ. Then we have

$$0 < \langle v, Av\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\langle v, v\rangle.$$

Therefore $\bar{\lambda} \in \mathbb{R}$ and hence $\lambda = \bar{\lambda} > 0$.

Now suppose that $A \in M_n(\mathbb{F})$ has all positive eigenvalues. By the Spectral theorem, $A = PDP^*$ for some unitary matrix $P \in M_n(\mathbb{F})$ and diagonal matrix $D \in M_n(\mathbb{F})$. Furthermore the eigenvalues of A are along the diagonal of D and so we may consider the matrix $\sqrt{D}$ defined by taking the square root of the eigenvalues of A, which are positive by assumption. Thus, we have

$$A = P\sqrt{D}\sqrt{D}P^*.$$

Letting $B = \sqrt{D}P^*$, we are done.

Finally, suppose that $A = B^*B$ for some $B \in M_n(\mathbb{F})$. We have

$$v^*Av = v^*B^*Bv = \langle Bv, Bv\rangle > 0$$

for all $v \in \mathbb{F}^n \backslash \{\mathbf{0}\}$. ■

Corollary 7.4.1. *Let* $A, B \in M_n(\mathbb{F})$ *be positive definite and* $c > 0$. *Then* $A + cB$ *and* A^{-1} *are positive definite.*

Proof. We have

$$0 < \langle v, A(v)\rangle + \langle v, cB(v)\rangle = \langle v, A(v) + cB(v)\rangle = \langle v, (A + cB)(v)\rangle.$$

Firstly, note that the Spectral theorem implies that A is a product of invertible matrices, and hence invertible. Define $y = Av$ for $v \in \mathbb{F}^n \backslash \{\mathbf{0}\}$. Then

$$y^*A^{-1}y = v^*A^*A^{-1}Av = v^*Av > 0.$$

And since A is invertible, we know that A^{-1} is onto. Hence the above inequality holds for any nonzero vector in $\mathbb{F}^n$. ■

Theorem 7.4.2. *Let* $\beta = \{v_1, \ldots v_n\}$ *be an orthonormal basis for an inner product space* V. *Then*

$$\langle v, w\rangle = [v]_\beta \cdot [w]_\beta.$$

Proof. Consider the isomorphism $T : V \to \mathbb{F}^n$ defined by $Tv_j = [v_j]_\beta = e_j$. Notice that

$$\langle v_i, v_j \rangle = \delta_{ij} = e_i \cdot e_j = [v_i]_\beta \cdot [v_j]_\beta.$$

Therefore, the sequlinearity of $\langle \cdot, \cdot \rangle$ implies that $\langle v, w \rangle = [v]_\beta \cdot [w]_\beta$. ∎

Theorem 7.4.3. *The function $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ is an inner product if and only if*

$$\langle v, w \rangle = v^* A w$$

for some positive definite matrix A.

Proof. By the above theorem, $\langle v, w \rangle = [v]_\beta \cdot [w]_\beta$. If we let P be the change of basis matrix converting the standard basis into β then

$$\langle v, w \rangle = Pv \cdot Pw = v^*(P^*P)w.$$

But we know that P^*P is positive definite. Conversely, suppose that $\langle v, w \rangle = v^*Aw$ for some positive definite matrix A. We have

$$\begin{aligned}
\langle v, w \rangle &= v^*Aw && (7.1)\\
&= v \cdot Aw && (7.2)\\
&= \overline{Aw \cdot v} && (7.3)\\
&= \overline{(Aw)^* v} && (7.4)\\
&= \overline{w^*A^*v} && (7.5)\\
&= \overline{\langle w, v \rangle} && (7.6)
\end{aligned}$$

Next

$$\begin{aligned}
\langle ax + b, y \rangle &= (ax + b)^* y && (7.7)\\
&= ((ax)^* + (b)^*)y && (7.8)\\
&= \bar{a}x^* + \langle b, y \rangle && (7.9)\\
&= \bar{a}x \cdot Ay + \langle b, y \rangle && (7.10)\\
&= (\bar{a}x)^* Ay + \langle b, y \rangle && (7.11)\\
&= ax^* Ay + \langle b, y \rangle && (7.12)\\
&= a\langle x, y \rangle + \langle b, y \rangle && (7.13)
\end{aligned}$$

And finally, $\langle v, v \rangle = v^*Av > 0$ for all nonzero $v \in \mathbb{F}^n$ by assumption and $\langle \mathbf{0}, \mathbf{0} \rangle = \mathbf{0}^* A \mathbf{0} = 0$. ∎