Finite Dimensional Inner Product Spaces

Jason Kenyon

November 2022

Contents

Preface **iii**

1 Vector Spaces **1**
1.1 Spaces and Subspaces 1
1.2 Linear Independence 3
1.3 Bases 4
1.4 Direct Sum and Projections 6

2 Linear Functions **8**
2.1 Linearity 8
2.2 Matrices 11

3 Linear Systems of Equations **16**
3.1 Rank 16
3.2 Form 21
3.3 Solution 23

4 The Determinant **25**
4.1 Permuations 25
4.2 Cofactor Expansion 25
4.3 Multilinear and Alternating 25
4.4 Properties 25
4.5 Measure 25

5 Eigenspaces **26**
5.1 Characteristic Polynomial 26
5.2 Diagonalization and Similarity 26
5.3 Dimension 26

6 Orthogonality **27**
6.1 Inner Products 27
6.2 The Adjoint 27
6.3 Orthogonal Projections 27
6.4 Normal and Unitary Operators 29
6.5 Definiteness 29

7 Matrix Decomposition **30**
7.1 Schur's Theorem 30
7.2 Spectral Theorem 30
7.3 Singular Value Decomposition and Pseudo-inverse 30

Preface

Hello

Chapter 1

Vector Spaces

1.1 Spaces and Subspaces

Definition 1.1.1. A vector space V over a field $\mathbb{F}$ is a set, along with a binary operation $+ : V^2 \to V$ and a binary operation $\cdot : \mathbb{F} \times V \to V$ that satisfy the following properties:

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^{-1} such that $v + v^{-1} = \mathbf{0}$

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

The elements of V are called vectors, and the elements of $\mathbb{F}$ are called scalars. The operations $+$ and $\cdot$ are called vector addition and scalar multiplication, respectively. We omit the $\cdot$ and do not explicitly apply $+$ for clarity. The $\mathbf{1}$ is the identity element of $\mathbb{F}$. $-a$ will denote the additive inverse of $a \in \mathbb{F}$ for the field $\mathbb{F}$, and $-v = v^{-1}$ will denote the additive inverse of a vector $v \in V$ under vector addition, while $-a(v)$ will denote multiplication of a vector by a scalar's additive inverse in $\mathbb{F}$.

Theorem 1.1.1. *Let x, y, and z be vectors in V. If $x + z = y + z$, then $x = y$. The zero element of V is unique. The additive inverse in V is unique for each vector in V.*

Theorem 1.1.2. *The following are true*

1. $0(x) = \mathbf{0}$
2. $(-a)x = -(ax) = a(-x)$
3. $a(\mathbf{0}) = \mathbf{0}$

Definition 1.1.2. A subspace W of a vector space V is a set $W \subseteq V$ that is itself a vector space.

Theorem 1.1.3. *Let V be a vector space with zero element $\mathbf{0}$. Then a subset $W \subseteq V$ 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 $c \in \mathbb{F}$.

Theorem 1.1.4. *Let S be a subset of a vector space V. Then* $\mathrm{span}(S)$ *is a subspace of V, and if any subspace of V contains S must necessarily contain* $\mathrm{span}(S)$

The proof follows directly from the definition of span. The span is defined precisely to generate a subspace in this way. 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

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

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.*

Proof. The proof should be clear when considering the above definition. ∎

Theorem 1.2.2. *Let S be a linearly independent subset of V and $v \in V$ such that $v \in S$. 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 $\beta \subseteq V$ is a basis for V if it is a linearly independent set such that $\text{span}(\beta) = V$.

Theorem 1.3.1. *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. ∎

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 linearly 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 $l \cup \{v\} \nsubseteq l$, contradicting the fact that l is the maximal element of L. ■

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 1.3.3. *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 $\text{span}(\{v_1, v_2, \ldots v_{m-1}\} \cup \{h_1, h_2, \cdots h_{n-(m-1)}\}) = V$. 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 1.3.3. ∎

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

Definition 1.3.2. 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 how they map basis elements.

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 1.1.3 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 $\{v_1, \ldots v_n\}$ 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 $\mathrm{im}(f) = W$. ∎

Theorem 2.1.6. *Let V and W be vector spaces, $\{v_1, v_2, \ldots v_n\}$ be a basis for V and $\{w_1, w_2, \ldots w_n\} \subseteq W$. Then there exists a unique linear map $f : V \to W$ such that $f(v_i) = w_i$.*

Proof. For $x \in V$ let $x = \sum_{i=1}^{n} a_i v_i$ and define $f : V \to W$ by

$$f(\sum_{i=1}^{n} a_i v_i) = \sum_{i=1}^{n} a_i w_i.$$

Trivially f is linear, and if we let $a_i = 0$ for all $a_i \neq a_j$ and $a_j = 1$, we have $f(v_j) = w_j$. To show uniqueness, suppose that there exists another linear function $g : V \to W$ such that $g(v_i) = w_i$. Then for $x \in V$ we have $x = \sum_{i=1}^{n} a_i v_i$ and $g(x) = \sum_{i=1}^{n} a_i g(v_i) = \sum_{i=1}^{n} a_i w_i = f(x)$. ■

2.2 Matrices

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

Definition 2.2.2. Let $f : V \to W$ be a linear function between two vector spaces V and W and let $\beta = (v_1, \ldots v_n)$ and $\gamma = (w_1, \ldots w_m)$ 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.

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 $(e_1, e_2, \ldots e_n)$ 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 $\mathcal{L}(V, W)$ be the set of all linear functions between two vector spaces V and W over a field $\mathbb{F}$. For $f, g \in \mathcal{L}(V, W)$ and all $x \in V$ define*

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

and

$$af = af(x)$$

for all $a \in \mathbb{F}$. Additionally, define $\mathbf{0}(x) = \mathbf{0}$. Then $\mathcal{L}(V, W)$ forms a vector space.

Proof. Trivially, for any two linear functions $f, g \in \mathcal{L}$ and scalar $a \in \mathbb{F}$ $af + g$ is a linear function. ∎

Theorem 2.2.2. *Let V and W be vector spaces with ordered bases β and γ respectively, and let $f, g \in \mathcal{L}(V, W)$. Then the following hold:*

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

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 2.2.3. *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*

$$[f(g)]^{\gamma}_{\alpha} = [f]^{\beta}_{\alpha}[g]^{\gamma}_{\beta}.$$

Proof. Let $\alpha = (v_1, \dots v_n)$, $\beta = (w_1, \dots w_m)$ and $\gamma = (z_1, \dots z_p)$. 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 2.2.1. *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*

$$[L_A]^{e'}_{e} = A$$

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

Proof. The linearity of L_A follows from theorem 2.2.4 The j^{th} column of $[L_A]^{e'}_{e}$ is $L_A(e_j) = Ae_j$ which is the j^{th} column of A. ■

Theorem 2.2.4. *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 2.2.5. *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.

Theorem 2.2.6. *Let V and W be vector spaces over $\mathbb{F}$ with dimension n and m respectively. Let β and γ be ordered bases for V and W respectively. Then there exists an isomorphism between the space $\mathcal{L}(V, W)$ and the space $M_{m\times n}(\mathbb{F})$ of $m \times n$ matrices with entries in $\mathbb{F}$.*

Proof. It should be clear that $M_{m\times n}(\mathbb{F})$ is in fact a vector space itself. This space is defined analogously to $\mathbb{F}^m$ for an element thereof is nothing but a column vector, and hence a matrix. The *standard* basis for this space is also defined in a similar way, with 1 in a single entry and 0 everywhere else. By theorem 2.2.2 this map is linear. By theorem 2.1.6 there exists a unique linear function $T : V \to W$ with $T(v_i) = w_i$ for all $1 \leq i \leq n$. Therefore, there is a unique map $T' : V \to W$ such that

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

That is to say that $[T]_\beta^\gamma = A$ for some $m \times n$ matrix 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 2.2.2. *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 2.2.7. *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

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

Conversely, if $\dim(V) = \dim(W)$ and $\beta = \{v_1, \dots v_n\}$ and $\gamma = \{w_1, \dots w_n\}$. By theorem 2.1.6 there is a unique linear map such that $T(v_i) = w_i$. Furthermore by theorem 1.1.4

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

■

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.

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

Theorem 3.1.5. *Let*

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

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

Proof.

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

Furthermore, $\dim(\text{span}(X))$ 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*

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

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 $\text{rank}(A) = 0$, 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, $\dim(\text{im}(TU)) \leq \dim(\text{im}(U))$. 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

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

■

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, $\dim(\ker(A)) = n - \text{rank}(L_A)$. Additionally, we know that $\text{rank}(A)$ is nothing but the number of linearly independent vectors defined by its rows which certainly cannot exceed m. Therefore $\text{rank}(A) \leq m < n$, in which case $n - \text{rank}(A) = \dim(\ker(A)) > 0$, 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* $\operatorname{rank}(A) = \operatorname{rank}(A|b)$.

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

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 $\operatorname{rank}(A) = \operatorname{rank}(A|b)$. 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

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

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*

$$Ax = b and C Ax = Cb$$

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 Characteristic Polynomial

5.2 Diagonalization and Similarity

5.3 Dimension

Chapter 6

Orthogonality

6.1 Inner Products

Hello

6.2 The Adjoint

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 $\{w_1, \dots w_n\}$ 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

$$\begin{aligned}
\langle x^\perp, w_j \rangle &= \langle x - x_W, w_j \rangle && (6.1)\\
&= \langle x - \sum_{i=1}^n \langle x, w_i \rangle w_i, w_j \rangle && (6.2)\\
&= \langle x, w_j \rangle - \sum_{i=1}^n \langle x, w_i \rangle \langle w_i, w_j \rangle && (6.3)\\
&= 0 && (6.4)
\end{aligned}$$

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.

$$\begin{aligned} ||x-y||^2 &= ||x_W + x^\perp - y||^2 && (6.5)\\ &= ||(x_W - y) + x^\perp||^2 && (6.6)\\ &= ||x_W - y||^2 + ||x^\perp||^2 && (6.7)\\ &\geq ||x^\perp||^2 = ||x - x_W||. && (6.8) \end{aligned}$$

∎

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 $W \subseteq V$ $x \in V$ and $\beta = \{v_1, \ldots v_n\}$ be an orthonormal basis for W and A be the matrix whose j^{th} column is v_j. Then the orthogonal projection of x on W $x_W = AA^*x$.*

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 $\operatorname{im}(A) = \operatorname{span}\beta$. Therefore, for some vector c $Ac = x_W$, 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 $x_W = Ac$. Furthermore, since β is orthonormal, A must be unitary, in which case

$$Ac = AA^*x = x_W.$$

■

Corollary 6.3.2. *AA^* is the unique projection of V on $W = \{x \in V : AA^*x = x\}$ along $W^\perp = \ker(AA^*)$. Additionally, AA^* is self adjoint.*

Proof. Surely AA^* is linear, and since we know that $x = x_W + x_{W^\perp}$ for all $x \in V$ it follows that $(AA^*)^2x = AA^*x_W = x_w = AA^*x$. 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 AA^* is unique. Furthermore,

$$\begin{aligned} (AA^*)^* &= (A^*)^*A^* && (6.9)\\ &= AA^*. && (6.10) \end{aligned}$$

■

6.4 Normal and Unitary Operators

self adjoint iff orthogonal projection all unitary operators are rotations

6.5 Definiteness

pos definite iff inner product

Chapter 7

Matrix Decomposition

7.1 Schur's Theorem

7.2 Spectral Theorem

7.3 Singular Value Decomposition and Pseudo-inverse