Linear Algebra

Xin Wen

Contents

I **Part One**

1 Vector Space 9

1.1 Definitions 9

1.1.1 Invertible Linear Maps 16

1.1.2 Duality 21

2 Polynomials 27

3 Eigenvalues, Eigenvectors, and Invariant Subspaces 31

4 Inner Product Spaces 39

5 Operators on Inner Product Spaces 49

6 Operators on Complex Vector Spaces 63

6.1 Jordan Form 71

7 Operators on Real Vector Spaces 75

7.1 Operators on Real Inner Product Spaces 79

8 Trace and Determinant 85

Bibliography 89

Articles 89

Part One

1 **Vector Space** 9
1.1 Definitions

2 **Polynomials** 27

3 **Eigenvalues, Eigenvectors, and Invariant Subspaces** 31

4 **Inner Product Spaces** 39

5 **Operators on Inner Product Spaces** 49

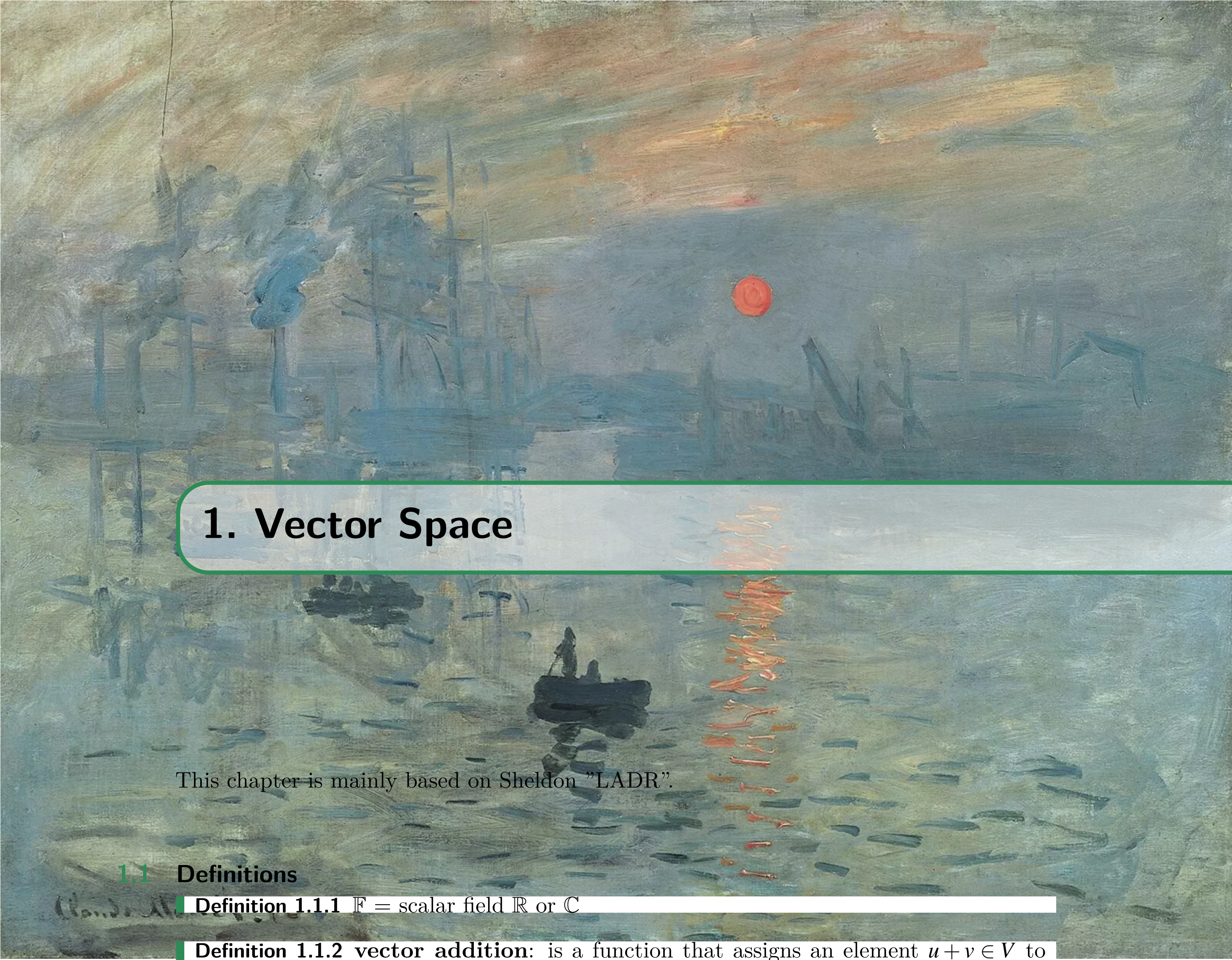
6 **Operators on Complex Vector Spaces** .. 63
6.1 Jordan Form

7 **Operators on Real Vector Spaces** 75
7.1 Operators on Real Inner Product Spaces

8 **Trace and Determinant** 85

Bibliography 89
Articles
Books

the first part is about LADR



1. Vector Space

This chapter is mainly based on Sheldon "LADR".

1.1 Definitions

Definition 1.1.1 $\mathbb{F}$ = scalar field $\mathbb{R}$ or $\mathbb{C}$

Definition 1.1.2 vector addition: is a function that assigns an element $u+v \in V$ to each pair of elements $u, v \in V$. A **scalar multiplication** on a set V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in F$ and each $v \in V$.

Definition 1.1.3 Vector Space is a set **V** containing at least two distinct elements called 0 and 1 over $\mathbb{F}$, together with two operations vector addition and scalar multiplication, if the following properties for all vectors x, y, z $\in V$ and all scalars c, d $\in \mathbb{F}$.

(1) $x+y=y+x$
(2) $x+(y+z)=(x+y)+z$
(3) There is a unique vector 0 such that $x+0=x$ for all x .
(4) $x+(-1)x=0$
(5) $1x=x$
(6) $c(dx)=(cd)x$
(7) $(c+d)x=cx+dx$
(8) $c(x+y)=cx+cy$

Proposition 1.1.1 1. A vector space has a unique additive identity.
2. Every element in a vector space has a unique additive inverse.

Definition 1.1.4 A subset U of V is called a **subspace** of V if U is also a vector space (using the same addition and scalar multiplication as on V).

Theorem 1.1.2 — Conditions for a subspace. A subset U of V is a subspace of V if and only if U satisfies the following three conditions:

1. additive identity: $0 \in U$
2. closed under addition: $u, w \in U$ implies $u + v \in U$
3. closed under scalar multiplication: $a \in \mathbb{F}$ and $u \in U$ implies $au \in U$

R The empty set is not a subspace of V because a subspace must be a vector space and hence must contain at least one element, namely, an additive identity.

Definition 1.1.5 — Sum of subsets.

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}$$

Proposition 1.1.3 Suppose $U_1, ..., U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, ..., U_m$.

Definition 1.1.6 — Direct sum. The sum $U_1 + \cdots + U_m$ is called a **direct sum** if each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$,where each $u_j \in U_j$. Denote $U_1 \oplus \cdots \oplus U_m$

Theorem 1.1.4 — Condition for a direct sum. Suppose $U_1, ..., U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each $u_j \in U_j$, is by taking each u_j equal to 0.

Theorem 1.1.5 — Direct sum of two subspaces. Suppose U and W are subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = 0$.(Proof by 1.1.4)

Definition 1.1.7 — Span. The set of all linear combinations of a list of vectors $v_1, \cdots, v_m$ in V is called the *span* of $v_1, \cdots, v_m$, denoted $span(v_1, \cdots, v_m)$. In other words,

$$\operatorname{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbf{F}\}$$

The span of the empty list is defined to be $\{0\}$. If $span(v_1, \cdots, v_m)$ equals V, we say that $v_1, \cdots, v_m$ spans V.

Proposition 1.1.6 The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Definition 1.1.8 — Polynomial. A function $p : \mathbb{F} \mapsto \mathbb{F}$ is called a *polynomial* with coefficients in $\mathbb{F}$ if there exist $a_0, \cdots, a_m \in \mathbb{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbf{F}$, $\mathscr{P}(F)$ is a vector space.

Definition 1.1.9 — Linearly independent. A list $v_1, \cdots, v_m$ of vectors in V is called **linearly independent** if the only choice of $a_0, \cdots, a_m \in \mathbb{F}$ that makes $a_1 v_1 + \cdots + a_m v + m$ equal 0 is $a_1 = \cdots = a_m = 0$. $v_1, \cdots, v_m$ is linearly independent if and only if each vector in $span(v_1, \cdots, v_m)$ has only one representation as a linear combination of $v_1, \cdots, v_m$.

Definition 1.1.10 — Linearly dependent. A list of vectors in V is called **linearly dependent** if it is not linearly independent. In other words, a list $v_1,\cdots,v_m$ of vectors in V is linearly dependent if there exist $a_0,\cdots,a_m \in \mathbb{F}$, not all 0, such that $a_1 = \cdots = a_m = 0$.

■ **Example 1.1** A list v of one vector $v \in V$ is linearly independent if and only if $v \neq 0$. A list of two vectors in V is linearly independent if and only if neither vector is a scalar multiple of the other.

Theorem 1.1.7 — Linear Dependence Lemma. Suppose $v_1,\cdots,v_m$ is a linearly dependent list in V. Then there exists $j \in \{1,2,\cdots,m\}$ such that the following hold:

1. $v_j \in span(v_1,\cdots,v_{j-1})$;
2. if the j^{th} term is removed from $v_1,\cdots,v_m$, the span of the remaining list equals $span(v_1,\cdots,v_m)$.

Theorem 1.1.8 — independent $\leq$ spanning . In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof. Suppose $u_1,\cdots,u_m$ is linearly independent in V. Suppose also that $w_1,\cdots,w_n$ spans V. We need to prove that $m \leq n$.

Let B be the list $w_1,\cdots,w_n$, which spans V. Thus adjoining any vector in V to this list produces a linearly dependent list (because the newly adjoined vector can be written as a linear combination of the other vectors). In particular, the list

$$u_1,w_1,\cdots,w_n$$

is linearly dependent. Thus by the Linear Dependence Lemma 1.1.7, we can remove one of the w's so that the new list B (of length n) consisting of u_1 and the remaining w's spans V. So finally, we can add all the $u's$, so $m \leq n$.

■

■ **Example 1.2** The list $(1,0,0),(0,1,0),(0,0,1)$ spans $\mathbb{R}^3$. Thus no list of length larger than 3 is linearly independent in $\mathbb{R}^3$.

The list $(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)$ is linearly independent in $\mathbb{R}^4$. Thus no list of length less than 4 spans $\mathbb{R}^4$. ■

Proposition 1.1.9

1. If some vectors are removed from a linearly independent list, the remaining list is also linearly independent.
2. Every list of vectors in V containing the 0 vector is linearly dependent.

Definition 1.1.11 — Basis. A *basis* of V is a list of vectors in V that is linearly independent and spans V.

Theorem 1.1.10 — Criterion for basis. A list $v_1,\cdots,v_m$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form

$$v = a_1v_1 + \cdots + a_nv_n$$

where $a_1,\ldots,a_n \in \mathbb{F}$

Corollary 1.1.11 — Spanning list contains a basis . Every spanning list in a vector space can be reduced to a basis of the vector space.

Corollary 1.1.12 Every finite-dimensional vector space has a basis.

Corollary 1.1.13 — Linearly independent list extends to a basis. Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.(Indep+basis, then reduce some elements)

Corollary 1.1.14 — Every subspace of V is part of a direct sum equal to V. Suppose V is finite-dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.

Theorem 1.1.15 — Basis length does not depend on basis. Any two bases of a finite-dimensional vector space have the same length.

Definition 1.1.12 The *dimension* of a finite-dimensional vector space is the length of any basis of the vector space, denoted by $\dim V$

Proposition 1.1.16
1. If V is finite-dimensional and U is a subspace of V, then $\dim U \leq \dim V$.
2. Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length $\dim V$ is a basis of V.
3. Suppose V is finite-dimensional. Then every spanning list of vectors in V with length $\dim V$ is a basis of V.

Theorem 1.1.17 — Dimension of a sum. If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$dim\,(U_1 + U_2) = dim\ U_1 + dim\ U_2 - dim\,(U_1 \cap U_2) \tag{1.1}$$

Proof. Let $u_1, \cdots, u_m$ be a basis of $U_1 \cap U_2$ thus $dim(U_1 \cap U_2) = m$. Because $u_1, \cdots, u_m$ is a basis of $U_1 \cap U_2$, it is linearly independent in U_1. Hence this list can be extended to a basis $u_1, \cdots, u_m, v_1, \cdots, v_j$ of U_1. Thus $dim\ U_1 = m + j$. Also extend $u_1, \cdots, u_m$ to a basis $u_1, \cdots, u_m, w_1, \cdots, w_k$ of U_2; thus $dim\ U_2 = m + k$.

We will show that

$$u_1, \cdots, u_m, v_1, \cdots, v_j, w_1, \cdots, w_k$$

is a basis of $U_1 + U_2$. This will complete the proof, because then we will have

$$\begin{aligned}\dim\,(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim\,(U_1 \cap U_2)\end{aligned}$$

Clearly $\operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k)$ contains U_1 and U_2 and hence equals $U_1 + U_2$. So to show that this list is a basis of $U_1 + U_2$ we need only show that it is linearly independent. To prove this, suppose

$$a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k = 0$$

where all the a 's, b 's, and c 's are scalars. We need to prove that all the a 's, b's, and c's equal 0. The equation above can be rewritten as

$$c_1w_1+\cdots+c_kw_k=-a_1u_1-\cdots-a_mu_m-b_1v_1-\cdots-b_jv_j$$

which shows that $c_1w_1+\cdots+c_kw_k \in U_1$. All the w 's are in U_2, so this implies that $c_1w_1+\cdots+c_kw_k \in U_1\cap U_2$. Because $u_1,\ldots,u_m$ is a basis of $U_1\cap U_2$, we can write

$$c_1w_1+\cdots+c_kw_k=d_1u_1+\cdots+d_mu_m$$

for some choice of scalars $d_1,\ldots,d_m$. But $u_1,\ldots,u_m,w_1,\ldots,w_k$ is linearly independent, so the last equation implies that all the c 's (and d 's) equal 0. Thus our original equation involving the a 's, b 's, and c 's becomes

$$a_1u_1+\cdots+a_mu_m+b_1v_1+\cdots+b_jv_j=0$$

Because the list $u_1,\cdots,u_m,v_1,\cdots,v_j$ islinearly independent, this equation implies that all the a's and b's are 0. We now know that all the a's, b's, and c's equal 0, as desired.

■

Definition 1.1.13 — Linear map. A *linearmap* from V to W is a function $T:V\mapsto W$ with the following properties:

1. additivity:

$$T(u+v)=Tu+Tv \text{ for all } u,v\in V$$

2. homogeneity: $T(\lambda v)=\lambda(Tv)$ for all $\lambda\in\mathbb{F}$ and all $v\in V$

The set of all linear maps from V to W is denoted $\mathscr{L}(V,W)$.

$$(S+T)(v)=Sv+Tv \text{ and } (\lambda T)(v)=\lambda(Tv)$$

With the operations of addition and scalar multiplication as defined above, $\mathscr{L}(V,W)$ is a vector space.

Theorem 1.1.18 Suppose $v_1,\cdots,v_n$ is a basis of V and $w_1,\cdots,w_n\in W$. Then there exists a unique linear map $T:V\mapsto W$ such that

$$Tv_j=w_j$$

for all $j=1,2,\cdots,n$.

Definition 1.1.14 — Product of Linear Maps. If $T\in\mathscr{L}(U,V)$ and $S;VW$, then the product $ST\in\mathscr{L}(U,W)$ is defined by for

$$(ST)(u)=S(Tu)$$

for all $u\in U$.

Proposition 1.1.19 Suppose T is a linear map from V to W. Then

$$T(0)=0$$

Definition 1.1.15 — null space, null T. For $T \in \mathscr{L}(V,W)$, the null space of T, denoted null T ,is the subset of V consisting of those vectors that T maps to 0 :

$$\text{null } T = \{v \in V : Tv = 0\}$$

Theorem 1.1.20 Suppose $T \in \mathscr{L}(V,W)$. Then null$T$ is a subspace of V.(null$T$ contains 0 and is closed under addition and scalar multiplication.)

Definition 1.1.16 — one-to-one. A function $T : V \mapsto W$ is called *injective* if $Tu = Tv$ implies $u = v$.$\Leftrightarrow$ $[u \neq v$ implies that $Tu \neq Tv]$

Theorem 1.1.21 — Injectivity is equivalent to null space equals $\{0\}$. Let $T \in \mathscr{L}(V,W)$. Then T is injective if and only if null $T = \{0\}$

Definition 1.1.17 — Range. For T a function from V to W, the range of T is the subset of W consisting of those vectors that are of the form Tv for some $v \in V$:

$$\text{range } T = \{Tv : v \in V\}$$

Theorem 1.1.22 — The range is a subspace. If $T \in \mathscr{L}(V,W)$, then range T is a subspace of W

Definition 1.1.18 — Surjective. A function $T : V \to W$ is called surjective(onto) if its range equals W

R ”Surjective” depends on the space ’W’.

Theorem 1.1.23 — Fundamental Theorem of Linear Maps. Suppose V is finite-dimensional and $T \in \mathscr{L}(V,W)$. Then range T is finite-dimensional and

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

Proof. Proof Let $u_1,\ldots,u_m$ be a basis of null T; thus dim null $T = m$. The linearly independent list $u_1,\ldots,u_m$ can be extended to a basis

$$u_1,\ldots,u_m,v_1,\ldots,v_n$$

of V (by 2.33). Thus $\dim V = m+n$. To complete the proof, we need only show that range T is finite-dimensional and dim range $T = n$. We will do this by proving that $Tv_1,\ldots,Tv_n$ is a basis of range T. Let $v \in V$. Because $u_1,\ldots,u_m,v_1,\ldots,v_n$ spans V, we can write

$$v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n$$

where the a ’s and b ’s are in $\mathbf{F}$. Applying T to both sides of this equation, we get

$$Tv = b_1Tv_1 + \cdots + b_nTv_n$$

where the terms of the form Tu_j disappeared because each u_j is in null T The last equation implies that $Tv_1,\ldots,Tv_n$ spans range T. In particular, range T is finite-dimensional. To show $Tv_1,\ldots,Tv_n$ is linearly independent, suppose $c_1,\ldots,c_n \in \mathbf{F}$ and

$$c_1Tv_1+\cdots+c_nTv_n=0$$

Then

$$T\left(c_1v_1+\cdots+c_nv_n\right)=0$$

Hence

$$c_1v_1+\cdots+c_nv_n \in \text{ null } T$$

Because $u_1,\ldots,u_m$ spans null T, we can write

$$c_1v_1+\cdots+c_nv_n=d_1u_1+\cdots+d_mu_m$$

where the d 's are in $\mathbf{F}$. This equation implies that all the c 's (and d 's) are 0 (because $u_1,\ldots,u_m,v_1,\ldots,v_n$ is linearly independent). Thus $Tv_1,\ldots,Tv_n$ is linearly independent and hence is a basis of range T, as desired. ■

Theorem 1.1.24 — A map to a smaller dimensional space is not injective. Suppose V and W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof. Let $T \in \mathscr{L}(V,W)$. Then

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T\\ &\geq \dim V - \dim W\\ &> 0\end{aligned}$$

■

Theorem 1.1.25 — A map to a larger dimensional space is not surjective. Suppose V and W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective. Proof Let $T \in \mathscr{L}(V,W)$. Then

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T\\ &\leq \dim V\\ &< \dim W\end{aligned}$$

Definition 1.1.19 — See these equation as linear map. **Homogeneous** means that the constant term on the right side of each equation below is 0.

$$\sum_{k=1}^{n} A_{1,k}x_k=0 \tag{1.2}$$

$$\vdots \tag{1.3}$$

$$\sum_{k=1}^{n} A_{m,k}x_k=0 \tag{1.4}$$

Define $T : \mathbf{F}^n \to \mathbf{F}^m$ by

$$T(x_1, \ldots, x_n) = \left(\sum_{k=1}^{n} A_{1,k} x_k, \ldots, \sum_{k=1}^{n} A_{m,k} x_k \right)$$

The equation $T(x_1, \ldots, x_n) = 0$ (the 0 here is the additive identity in $\mathbf{F}^m$, namely, the list of length m of all 0 's) is the same as the homogeneous system of linear equations above.

Theorem 1.1.26 A homogeneous system of linear equations with more variables than equations has nonzero solutions.

Proof. T is not injective if $n > m$. ■

Theorem 1.1.27 An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.

Definition 1.1.20 3.32 Definition matrix of a linear map, $\mathcal{M}(T)$ Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The matrix of T with respect to these bases is the m -by- n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$T v_k = A_{1,k} w_1 + \cdots + A_{m,k} w_m$$

If the bases are not clear from the context, then the notation $\mathcal{M}(T, (v_1, \ldots, v_n), (w_1, \ldots, w_m))$ is used.

$$\mathcal{M}(T) = \begin{array}{c c} & \begin{array}{ccccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \begin{array}{c} w_1 \\ \vdots \\ w_m \end{array} & \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{pmatrix} \end{array} \tag{1.5}$$

The k^{th} column of $\mathcal{M}(T)$ consists of the scalars needed to write $T v_k$ as a linear combination of $(w_1, \ldots, w_m)$ $T v_k = \sum_{i=1}^{m} A_{j,k} w_j$

R If T maps an n -dimensional vector space to an m -dimensional vector space, then $\mathcal{M}(T)$ is an $m \times n$ matrix.

Proposition 1.1.28 — The matrix of linear maps. 1. Suppose $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$

2. Suppose $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$
3. If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$

1.1.1 Invertible Linear Maps

Definition 1.1.21 A linear map $T \in \mathcal{L}(V, W)$ is called **invertible** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equals the identity map on W

A linear map $S \in \mathscr{L}(W,V)$ satisfying $ST = I$ and $TS = I$ is called an **inverse** of T (note that the first I is the identity map on V and the second I is the identity map on W).

Proposition 1.1.29 1. An invertible linear map has a unique inverse.
2. A linear map is invertible if and only if it is injective and surjective.

Definition 1.1.22 An **isomorphism** is an invertible linear map.
Two vector spaces are called **isomorphic** if there is an isomorphism from one vector space onto the other one.

Theorem 1.1.30 Two finite-dimensional vector spaces over $\mathbf{F}$ are isomorphic if and only if they have the same dimension.

Proof. Proof First suppose V and W are isomorphic finite-dimensional vector spaces. Thus there exists an isomorphism T from V onto W. Because T is invertible, we have null $T = \{0\}$ and range $T = W$. Thus dim null $T = 0$ and dim range $T = \dim W$. The formula

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

(the Fundamental Theorem of Linear Maps, which is 3.22) thus becomes the equation $\dim V = \dim W$, completing the proof in one direction.

To prove the other direction, suppose V and W are finite-dimensional vector spaces with the same dimension. Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_n$ be a basis of W. Let $T \in \mathscr{L}(V,W)$ be defined by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n$$

Then T is a well-defined linear map because $v_1, \ldots, v_n$ is a basis of V (see 3.5). Also, T is surjective because $w_1, \ldots, w_n$ spans W. Furthermore, null $T = \{0\}$ because $w_1, \ldots, w_n$ is linearly independent; thus T is injective. Because T is injective and surjective, it is an isomorphism (see 3.56). Hence V and W are isomorphic, as desired. ∎

Definition 1.1.23 For m and n positive integers, the set of all m -by- n matrices with entries in $\mathbf{F}$ is denoted by $\mathbf{F}^{m,n}$

Theorem 1.1.31 — $\dim \mathbf{F}^{m,n} = mn$. Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbf{F}^{m,n}$ is a vector space with dimension mn

Theorem 1.1.32 — $\mathscr{L}(V,W)$ and $\mathbf{F}^{m,n}$ are isomorphic. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W Then $\mathscr{M}$ is an isomorphism between $\mathscr{L}(V,W)$ and $\mathbf{F}^{m,n}$

Proof. We already noted that $\mathscr{M}$ is linear. We need to prove that $\mathscr{M}$ is injective and surjective. Both are easy. We begin with injectivity. If $T \in \mathscr{L}(V,W)$ and $\mathscr{M}(T) = 0$, then $Tv_k = 0$ for $k = 1, \ldots, n$. Because $v_1, \ldots, v_n$ is a basis of V, this implies $T = 0$. Thus $\mathscr{M}$ is injective (by 3.16).

To prove that $\mathscr{M}$ is surjective, suppose $A \in \mathbf{F}^{m,n}$. Let T be the linear map from V to W such that

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

for $k=1,\ldots,n$ (see 3.5). Obviously $\mathscr{M}(T)$ equals A, and thus the range of $\mathscr{M}$ equals $\mathbf{F}^{m,n}$, as desired.

■

Theorem 1.1.33 — $\dim \mathscr{L}(V,W) = (\dim V)(\dim W)$. Suppose V and W are finite-dimensional. Then $\mathscr{L}(V,W)$ is finite dimensional and

$$\dim \mathscr{L}(V,W) = (\dim V)(\dim W)$$

Definition 1.1.24 — operator. A linear map from a vector space to itself is called an **operator**.

Theorem 1.1.34 — Injectivity is equivalent to surjectivity in finite dimensions . Suppose V is finite-dimensional and $T \in \mathscr{L}(V)$. Then the following are equivalent:

1. T is invertible;
2. T is injective;
3. T is surjective.

Proof. $\dim \operatorname{range} T = \dim V - \dim \operatorname{null} T$ ■

Definition 1.1.25 — product of vector spaces. Suppose $V_1,\ldots,V_m$ are vector spaces over **F**

The product $V_1 \times \cdots \times V_m$ is defined by

$$V_1 \times \cdots \times V_m = \{(v_1,\ldots,v_m) : v_1 \in V_1,\ldots,v_m \in V_m\}$$

Addition on $V_1 \times \cdots \times V_m$ is defined by

$$(u_1,\ldots,u_m)+(v_1,\ldots,v_m) = (u_1+v_1,\ldots,u_m+v_m)$$

Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by

$$\lambda(v_1,\ldots,v_m) = (\lambda v_1,\ldots,\lambda v_m)$$

Theorem 1.1.35 — Product of vector spaces is a vector space. Suppose $V_1,\ldots,V_m$ are vector spaces over **F**. Then $V_1 \times \cdots \times V_m$ is a vector space over **F**

Theorem 1.1.36 — Dimension of a product is the sum of dimensions. Suppose $V_1,\ldots,V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m$$

Proof. consider the element of $V_1 \times \cdots \times V_m$ that equals the basis vector in the j^{th} slot and 0 in the other slots. ■

Theorem 1.1.37 Suppose that $U_1,\ldots,U_m$ are subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by

$$\Gamma(u_1,\ldots,u_m) = u_1 + \cdots + u_m$$

Then $U_1 + \cdots + U_m$ is a direct sum if and only if Γ is injective.

Theorem 1.1.38 — A sum is a direct sum if and only if dimensions add up. Suppose V is finite-dimensional and $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m$$

Proof. The map Γ in 3.77 is surjective. Thus by the Fundamental Theorem of Linear Maps (3.22), Γ is injective if and only if

$$\dim(U_1 + \cdots + U_m) = \dim(U_1 \times \cdots \times U_m)$$

Combining 3.77 and 3.76 now shows that $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m$$

as desired. ■

Definition 1.1.26 — $v + U$. Suppose $v \in V$ and U is a subspace of V. Then $v + U$ is the subset of V defined by

$$v + U = \{v + u : u \in U\}$$

Definition 1.1.27 An **affine** subset of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V

For $v \in V$ and U a subspace of V, the affine subset $v + U$ is said to be **parallel** to U

Definition 1.1.28 — Quotient Space, V/U. Suppose U is a subspace of V. Then the quotient space V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U = \{v + U : v \in V\}$$

Theorem 1.1.39 — Two affine subsets parallel to U are equal or disjoint. Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent:

1. $v - w \in U$
2. $v + U = w + U$
3. $(v + U) \cap (w + U) \neq \varnothing$

Proof. First suppose (a) holds, so $v - w \in U$. If $u \in U$, then

$$v + u = w + ((v - w) + u) \in w + U$$

Thus $v + U \subset w + U$. Similarly, $w + U \subset v + U$. Thus $v + U = w + U$ completing the proof that (a) implies (b). Obviously (b) implies (c). Now suppose (c) holds, so $(v + U) \cap (w + U) \neq \varnothing$. Thus there exist $u_1, u_2 \in U$ such that $v + u_1 = w + u_2$ Thus $v - w = u_2 - u_1$. Hence $v - w \in U$, showing that (c) implies (a) and completing the proof. ■

Definition 1.1.29 — addition and scalar multiplication on V/U. Suppose U is a subspace of V. Then addition and scalar multiplication are defined on V/U by

$$(v+U)+(w+U)=(v+w)+U$$
$$\lambda(v+U)=(\lambda v)+U$$

for $v,w\in V$ and $\lambda\in\mathbf{F}$

Theorem 1.1.40 — Quotient space is a vector space. Suppose U is a subspace of V. Then V/U, with the operations of addition and scalar multiplication as defined above, is a vector space. Note that the additive identity of V/U is $0+U$ (which equals U) and that the additive inverse of $v+U$ is $(-v)+U$

Definition 1.1.30 — quotient map, π. Suppose U is a subspace of V. The quotient map π is the linear map $\pi: V\to V/U$ defined by

$$\pi(v)=v+U$$

for $v\in V$

Theorem 1.1.41 — Dimension of a quotient space. Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim V/U=\dim V-\dim U$$

Proof. Let π be the quotient map from V to V/U. From 3.85, we see that null $\pi=U$.(Note that the zero element in V /U is 0+U = U) Clearly range $\pi=V/U$. The Fundamental Theorem of Linear Maps (3.22) thus tells us that

$$\dim V=\dim U+\dim V/U$$

■

Definition 1.1.31 — Definition $\tilde{T}$. Suppose $T\in\mathscr{L}(V,W)$. Define $\tilde{T}: V/(\text{ null } T)\to W$ by

$$\tilde{T}(v+\text{ null } T)=Tv$$

R To show that the definition of $\tilde{T}$ makes sense, suppose $u,v\in V$ are such that $u+$ null $T=v+$ null T. By 3.85, we have $u-v\in$ null T. Thus $T(u-v)=0$. Hence $Tu=Tv$. Thus the definition of $\tilde{T}$ indeed makes sense.

Theorem 1.1.42 — Null space and range of $\tilde{T}$. Suppose $T\in\mathscr{L}(V,W)$. Then
1. $\tilde{T}$ is a linear map from $V/$ (null T) to W
2. $\tilde{T}$ is injective;
3. range $\tilde{T}=$ range T
4. $V/($ null $T)$ is isomorphic to range T

Proof. 1. (b) Suppose $v\in V$ and $\tilde{T}(v+\text{ null } T)=0$. Then $Tv=0$. Thus $v\in$ null T Hence 3.85 implies that $v+$ null $T=0+$ null T. This implies that null $\tilde{T}=0$, and hence $\tilde{T}$ is injective, as desired.

2. The definition of $\tilde{T}$ shows that range $\tilde{T}$ = range T.
3. Parts (b) and (c) imply that if we think of $\tilde{T}$ as mapping into range T, then $\tilde{T}$ is an isomorphism from $V/$ (null T) onto range T.

■

1.1.2 Duality

Definition 1.1.32 — Definition linear functional. A linear functional on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathscr{L}(V,\mathbf{F})$

Definition 1.1.33 — Definition dual space V'. The dual space of V, denoted V', is the vector space of all linear functionals on V. In other words, $V' = \mathscr{L}(V,\mathbf{F})$

Theorem 1.1.43 — $\dim V' = \dim V$. Suppose V is finite-dimensional. Then V' is also finite-dimensional and $\dim V' = \dim V$

Definition 1.1.34 — dual basis. If $v_1,\ldots,v_n$ is a basis of V, then the dual basis of $v_1,\ldots,v_n$ is the list $\varphi_1,\ldots,\varphi_n$ of elements of V', where each φ_j is the linear functional on V such that

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

Theorem 1.1.44 — Dual basis is a basis of the dual space. Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'

Proof. Suppose $v_1,\ldots,v_n$ is a basis of V. Let $\varphi_1,\ldots,\varphi_n$ denote the dual basis. To show that $\varphi_1,\ldots,\varphi_n$ is a linearly independent list of elements of V' suppose $a_1,\ldots,a_n \in F$ are such that

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0$$

Now $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$ for $j = 1,\ldots,n$. The equation above thus shows that $a_1 = \cdots = a_n = 0$. Hence $\varphi_1,\ldots,\varphi_n$ is linearly independent. Now 2.39 and 3.95 imply that $\varphi_1,\ldots,\varphi_n$ is a basis of V'

■

Definition 1.1.35 — dual map. If $T \in \mathscr{L}(V,W)$, then the dual map of T is the linear map $T' \in \mathscr{L}(W',V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$

Theorem 1.1.45 — Algebraic properties of dual maps.
1. $(S+T)' = S' + T'$ for all $S,T \in \mathscr{L}(V,W)$
2. $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbf{F}$ and all $T \in \mathscr{L}(V,W)$
3. $(ST)' = T'S'$ for all $T \in \mathscr{L}(U,V)$ and all $S \in \mathscr{L}(V,W)$

Proof. (3)suppose $\varphi \in W'$. Then $(ST)'(\varphi) = \varphi \circ (ST) = (\varphi \circ S) \circ T = T'(\varphi \circ S) = T'(S'(\varphi)) = (T'S')(\varphi)$ ■

Definition 1.1.36 — annihilator. For $U \subset V$, the annihilator of U, denoted U^0, is defined by

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}$$

Theorem 1.1.46 — The annihilator is a subspace. Suppose $U \subset V$. Then U^0 is a subspace of V'

Theorem 1.1.47 — Dimension of the annihilator. Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V$$

Proof. Let $i \in \mathscr{L}(U,V)$ be the inclusion map defined by $i(u) = u$ for $u \in U$ Thus i' is a linear map from V' to U'. The Fundamental Theorem of Linear Maps (3.22) applied to i' shows that

$$\text{dim range } i' + \text{dim null } i' = \dim V'$$

However, null $i' = U^0$ (as can be seen by thinking about the definitions) and $\dim V' = \dim V($ by 3.95), so we can rewrite the equation above as

$$\text{dim range } i' + \dim U^0 = \dim V$$

If $\varphi \in U'$, then φ can be extended to a linear functional ψ on V (see, for example, Exercise 11 in Section 3.A). The definition of i' shows that $i'(\psi) = \varphi$. Thus $\varphi \in$ range i', which implies that range $i' = U'$. Hence dim range $i' = \dim U' = \dim U$, and the displayed equation above becomes the desired result.

■

Exercise 1.1 — 3.A 11. Suppose V is finite-dimensional. Prove that every linear map on a subspace of V can be extended to a linear map on V. In other words, show that if U is a subspace of V and $S \in \mathscr{L}(U,W)$, then there exists $T \in \mathscr{L}(V,W)$ such that $Tu = Su$ for all $u \in U$ ■

Theorem 1.1.48 — The null space of T'. Suppose V and W are finite-dimensional and $T \in \mathscr{L}(V,W)$. Then

1. null $T' = ($ range $T)^0$
2. $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$

Proof. 1. First suppose $\varphi \in$ null T'. Thus $0 = T'(\varphi) = \varphi \circ T$. Hence

$$0 = (\varphi \circ T)(v) = \varphi(Tv) \quad \text{for every } v \in V$$

Thus $\varphi \in (\text{range } T\ )^0$. This implies that null $T' \subset (\text{range } T\ )^0$. To prove the inclusion in the opposite direction, now suppose that $\varphi \in ($ range $T)^0$. Thus $\varphi(Tv) = 0$ for every vector $v \in V$. Hence $0 = \varphi \circ T = T'(\varphi)$. In other words, $\varphi \in$ null T', which shows that (range $T)^0 \subset$ null T', completing the proof of (a).

2. We have

$$
\begin{aligned}
\dim \operatorname{null} T' &= \dim(\operatorname{range} T)^0 \\
&= \dim W - \dim \operatorname{range} T \\
&= \dim W - (\dim V - \dim \operatorname{null} T) \\
&= \dim \operatorname{null} T + \dim W - \dim V
\end{aligned}
$$

where the first equality comes from (a), the second equality comes from 3.106, and the third equality comes from the Fundamental Theorem of Linear Maps (3.22). ∎

Theorem 1.1.49 — T surjective is equivalent to T' injective. Suppose V and W are finite-dimensional and $T \in \mathscr{L}(V,W)$. Then T is surjective if and only if T' is injective.

Proof. The map $T \in \mathscr{L}(V,W)$ is surjective if and only if range $T = W$ which happens if and only if (range T) $^0 = \{0\}$, which happens if and only if null $T' = \{0\}$[by 3.107(a)], which happens if and only if T' is injective. ∎

Theorem 1.1.50 — The range of T'. Suppose V and W are finite-dimensional and $T \in \mathscr{L}(V,W)$. Then
1. dim range $T' =$ dim range T
2. range $T' = ($ null $T)^0$

Proof. 1. We have

$$
\begin{aligned}
\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' \\
&= \dim W - \dim(\operatorname{range} T)^0 \\
&= \dim \operatorname{range} T
\end{aligned}
$$

2. First suppose $\varphi \in$ range T'. Thus there exists $\psi \in W'$ such that $\varphi = T'(\psi)$. If $v \in$ null T, then $\varphi(v) = (T'(\psi))v = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0$ Hence $\varphi \in ($ null $T)^0$. This implies that range $T' \subset ($ null $T)^0$ We will complete the proof by showing that range T' and (null T) ^{0} have the same dimension. To do this, note that

$$
\begin{aligned}
\dim \operatorname{range} T' &= \dim \operatorname{range} T \\
&= \dim V - \dim \operatorname{null} T \\
&= \dim(\operatorname{null} T)^0
\end{aligned}
$$

∎

Theorem 1.1.51 — T injective is equivalent to T' surjective. Suppose V and W are finite-dimensional and $T \in \mathscr{L}(V,W)$. Then T is injective if and only if T' is surjective.

Proof. The map $T \in \mathscr{L}(V,W)$ is injective if and only if null $T = \{0\}$ which happens if and only if (null $T)^0 = V'$, which happens if and only if range $T' = V'$[by 3.109(b)], which happens if and only if T' is surjective. ∎

Proposition 1.1.52 — The transpose of a matrix A. 1. $(A+C)^{\mathrm{t}} = A^{\mathrm{t}} + C^{\mathrm{t}}$
2. $(\lambda A)^t = \lambda A^t$ for all $m-$ by- n matrices A, C and all $\lambda \in \mathbf{F}$

Theorem 1.1.53 — The transpose of the product of matrices. If A is an m -by- n matrix and C is an n -by- p matrix, then

$$(AC)^t = C^t A^t$$

Proof. Suppose $1 \leq k \leq p$ and $1 \leq j \leq m$. Then

$$\begin{aligned}
\left((AC)^t\right)_{k,j} &= (AC)_{j,k} \\
&= \sum_{r=1}^{n} A_{j,r} C_{r,k} \\
&= \sum_{r=1}^{n} \left(C^t\right)_{k,r} \left(A^t\right)_{r,j} \\
&= \left(C^t A^t\right)_{k,j}
\end{aligned}$$

Thus $(AC)^{\mathrm{t}} = C^{\mathrm{t}} A^{\mathrm{t}}$, as desired. ■

Theorem 1.1.54 — The matrix of T' is the transpose of the matrix of T. Suppose $T \in \mathscr{L}(V, W)$. Then $\mathscr{M}(T') = (\mathscr{M}(T))^t$

Proof. Let $A = \mathscr{M}(T)$ and $C = \mathscr{M}(T')$. Suppose $1 \leq j \leq m$ and $1 \leq k \leq n$ From the definition of $\mathscr{M}(T')$ we have

$$T'(\psi_j) = \sum_{r=1}^{n} C_{r,j} \varphi_r$$

The left side of the equation above equals $\psi_j \circ T$. Thus applying both sides of the equation above to v_k gives

$$\begin{aligned}
(\psi_j \circ T)(v_k) &= \sum_{r=1}^{n} C_{r,j} \varphi_r(v_k) \\
&= C_{k,j}
\end{aligned}$$

We also have

$$\begin{aligned}
(\psi_j \circ T)(v_k) &= \psi_j(T v_k) \\
&= \psi_j \left(\sum_{r=1}^{m} A_{r,k} w_r \right) \\
&= \sum_{r=1}^{m} A_{r,k} \psi_j(w_r) \\
&= A_{j,k}
\end{aligned}$$

Comparing the last line of the last two sets of equations, we have $C_{k,j} = A_{j,k}$ Thus $C = A^{\mathrm{t}}$. In other words, $\mathscr{M}(T') = (\mathscr{M}(T))^{\mathrm{t}}$, as desired.

■

Definition 1.1.37 — row rank, column rank. Suppose A is an m -by- n matrix with entries in $\mathbf{F}$.

The row rank of A is the dimension of the span of the rows of A in $\mathbf{F}^{1,n}$

The column rank of A is the dimension of the span of the columns of A in $\mathbf{F}^{m,1}$

Theorem 1.1.55 — Dimension of range T equals column rank of $\mathcal{M}(T)$. Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then dim range T equals the column rank of $\mathcal{M}(T)$

Proof. Suppose $v_1,\dots,v_n$ is a basis of V and $w_1,\dots,w_m$ is a basis of W. The function that takes $w \in \operatorname{span}(Tv_1,\dots,Tv_n)$ to $\mathcal{M}(w)$ is easily seen to be an isomorphism from $\operatorname{span}(Tv_1,\dots,Tv_n)$ onto $\operatorname{span}(\mathcal{M}(Tv_1),\dots,\mathcal{M}(Tv_n))$ Thus

$$\dim\operatorname{span}(Tv_1,\dots,Tv_n) = \dim\operatorname{span}(\mathcal{M}(Tv_1),\dots,\mathcal{M}(Tv_n))$$

where the last dimension equals the column rank of $\mathcal{M}(T)$

It is easy to see that range $T = \operatorname{span}(Tv_1,\dots,Tv_n)$. Thus we have $\dim\operatorname{range}T = \dim\operatorname{span}(Tv_1,\dots,Tv_n) =$ the column rank of $\mathcal{M}(T)$, as desired.

■

Theorem 1.1.56 — Row rank equals column rank. Suppose $A \in \mathbf{F}^{m,n}$. Then the row rank of A equals the column rank of A

Proof. Define $T : \mathbf{F}^{n,1} \to \mathbf{F}^{m,1}$ by $Tx = Ax$. Thus $\mathcal{M}(T) = A$, where $\mathcal{M}(T)$ is computed with respect to the standard bases of $\mathbf{F}^{n,1}$ and $\mathbf{F}^{m,1}$.

Now

$$\begin{aligned}
columnrankofA &= columnrankof\mathcal{M}(T)\\
&= \dim\operatorname{range}T\\
&= \dim\operatorname{range}T'\\
&= \text{column rank of } \mathcal{M}(T')\\
&= \text{column rank of } A^{t}\\
&= \operatorname{row}\operatorname{rank}\ \text{of } A
\end{aligned}$$

■

Definition 1.1.38 — rank. The rank of a matrix $A \in \mathbf{F}^{m,n}$ is the column rank of A

2. Polynomials

Definition 2.0.1 — polynomial. A function $p : \mathbf{F} \to \mathbf{F}$ is called a polynomial with coefficients in $\mathbf{F}$ if there exist $a_0, \dots, a_m \in \mathbf{F}$ such that
$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$ for all $z \in \mathbf{F}$

Theorem 2.0.1 — If a polynomial is the zero function, then all coefficients are 0. Suppose $a_0, \dots, a_m \in \mathbf{F}$. If

$$a_0 + a_1 z + \cdots + a_m z^m = 0$$

for every $z \in \mathbf{F}$, then $a_0 = \cdots = a_m = 0$

Proof. We will prove the contrapositive. If not all the coefficients are 0, then by changing m we can assume $a_m \neq 0$. Let

$$z = \frac{|a_0| + |a_1| + \cdots + |a_{m-1}|}{|a_m|} + 1, \quad |a_0| + |a_1| + \cdots + |a_{m-1}| + |a_m| = |a_m| z$$

Note that $z \geq 1$, and thus $z^j \leq z^{m-1}$ for $j = 0, 1, \dots, m-1$. Using the Triangle Inequality, we have

$$\begin{aligned} \left|a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}\right| &\leq (|a_0| + |a_1| + \cdots + |a_{m-1}|) z^{m-1} \\ &< |a_m z^m| \end{aligned}$$

Thus $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} \neq -a_m z^m$. Hence we conclude that $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + a_m z^m \neq 0$ ■

Definition 2.0.2 If p and s are nonnegative integers, with $s \neq 0$, then there exist nonnegative integers q and r such that

$$p = sq + r$$

and $r < s$. Think of dividing p by s, getting quotient q with remainder r.

Definition 2.0.3 Recall that $\mathscr{P}(F)$ denotes the vector space of all polynomials with coefficients in $\mathbf{F}$ and that $\mathscr{P}_m(\mathbf{F})$ is the subspace of $\mathscr{P}(\mathbf{F})$ consisting of the polynomials with coefficients in $\mathbf{F}$ and degree at most m.

Theorem 2.0.2 — Division Algorithm for Polynomials. Suppose that $p, s \in \mathscr{P}(\mathbf{F})$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathscr{P}(\mathbf{F})$ such that

$$p = sq + r$$

and $\deg r < \deg s$

Proof. Let $n = \deg p$ and $m = \deg s$. If $n < m$, then take $q = 0$ and $r = p$ to get the desired result. Thus we can assume that $n \geq m$ Define $T : \mathscr{P}_{n-m}(\mathbf{F}) \times \mathscr{P}_{m-1}(\mathbf{F}) \to \mathscr{P}_n(\mathbf{F})$ by $T(q, r) = sq + r$ The reader can easily verify that T is a linear map. If $(q, r) \in$ null T, then $sq + r = 0$, which implies that $q = 0$ and $r = 0$ [because otherwise $\deg sq \geq m$ and thus sq cannot equal $-r$]. Thus dim null $T = 0$ (proving the "unique" part of the result). From 3.76 we have $\dim(P_{n-m}(\mathbf{F}) \times \mathscr{P}_{m-1}(\mathbf{F})) = (n - m + 1) + (m - 1 + 1) = n + 1$ The Fundamental Theorem of Linear Maps (3.22) and the equation displayed above now imply that dim range $T = n + 1$, which equals $\dim \mathscr{P}_n(\mathbf{F})$. Thus range $T = \mathscr{P}_n(\mathbf{F})$, and hence there exist $q \in \mathscr{P}_{n-m}(\mathbf{F})$ and $r \in \mathscr{P}_{m-1}(\mathbf{F})$ such that $p = T(q, r) = sq + r$ ■

Definition 2.0.4 — Definition zero of a polynomial. A number $\lambda \in \mathbf{F}$ is called a zero (or root) of a polynomial $p \in \mathscr{P}(\mathbf{F})$ if

$$p(\lambda) = 0$$

Definition 2.0.5 — factor. A polynomial $s \in \mathscr{P}(\mathbf{F})$ is called a factor of $p \in \mathscr{P}(\mathbf{F})$ if there exists a polynomial $q \in \mathscr{P}(\mathbf{F})$ such that $p = sq$

Theorem 2.0.3 — Each zero of a polynomial corresponds to a degree-1 factor. Suppose $p \in \mathscr{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in \mathscr{P}(\mathbf{F})$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbf{F}$

Proof. BY Division Algorithm for Polynomials. ■

Theorem 2.0.4 — A polynomial has at most as many zeros as its degree. Suppose $p \in \mathscr{P}(\mathbf{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbf{F}$.

Proof. By $p(z) = (z - \lambda)q(z)$ and induction. ■

Theorem 2.0.5 — Fundamental Theorem of Algebra. Every nonconstant polynomial with complex coefficients has a zero.

Proof. Let p be a nonconstant polynomial with complex coefficients. Suppose p has no zeros. Then $1/p$ is an analytic function on $\mathbf{C}$. Furthermore, $|p(z)| \to \infty$ as $|z| \to \infty$, which implies that $1/p \to 0$ as $|z| \to \infty$. Thus $1/p$ is a bounded analytic function on $\mathbf{C}$. By Liouville's theorem, every such function is constant. But if $1/p$ is constant, then p is constant, contradicting our assumption that p is nonconstant. ∎

Theorem 2.0.6 — Liouville's theorem. Every bounded entire function must be constant. That is, every holomorphic function f for which there exists a positive number M such that $|f(z)| \leq M$ for all z in $\mathbb{C}$ is constant. Equivalently, non-constant holomorphic functions on $\mathbb{C}$ have unbounded images.

Theorem 2.0.7 — Factorization of a polynomial over C . If $p \in \mathscr{P}(\mathbf{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form

$$p(z) = c\,(z - \lambda_1) \cdots (z - \lambda_m)$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbf{C}$

(R) A polynomial with real coefficients may have no real zeros.

Theorem 2.0.8 — Polynomials with real coefficients have zeros in pairs. Suppose $p \in \mathscr{P}(\mathbf{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbf{C}$ is a zero of p, then so is $\bar{\lambda}$

(R) If $(x - \lambda)$ with λ a nonreal complex number, then $(x - \bar{\lambda})$ is also a term in the factorization. Multiplying together these two terms, we get

$$(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2)$$

Theorem 2.0.9 — Factorization of a polynomial over R. Suppose $p \in \mathscr{P}(\mathbf{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of the factors) of the form

$$p(x) = c\,(x - \lambda_1) \cdots (x - \lambda_m)\,(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$, with $b_j^2 < 4c_j$ for each j

Proof. Think of p as an element of $\mathscr{P}(\mathbf{C})$. If all the (complex) zeros of p are real, then we are done by 4.14. Thus suppose p has a zero $\lambda \in \mathbf{C}$ with $\lambda \notin \mathbf{R}$ By 4.15, $\bar{\lambda}$ is a zero of p. Thus we can write

$$\begin{aligned} p(x) &= (x - \lambda)(x - \bar{\lambda})q(x) \\ &= \left(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\right) q(x) \end{aligned}$$

for some polynomial $q \in \mathscr{P}(\mathbf{C})$ with degree two less than the degree of p. If we can prove that q has real coefficients, then by using induction on the degree of p, we can conclude that $(x-\lambda)$ appears in the factorization of p exactly as many times as $(x-\bar{\lambda})$. To prove that q has real coefficients, we solve the equation above for q getting

$$q(x) = \frac{p(x)}{x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2}$$

for all $x \in \mathbf{R}$. The equation above implies that $q(x) \in \mathbf{R}$ for all $x \in \mathbf{R}$ Writing

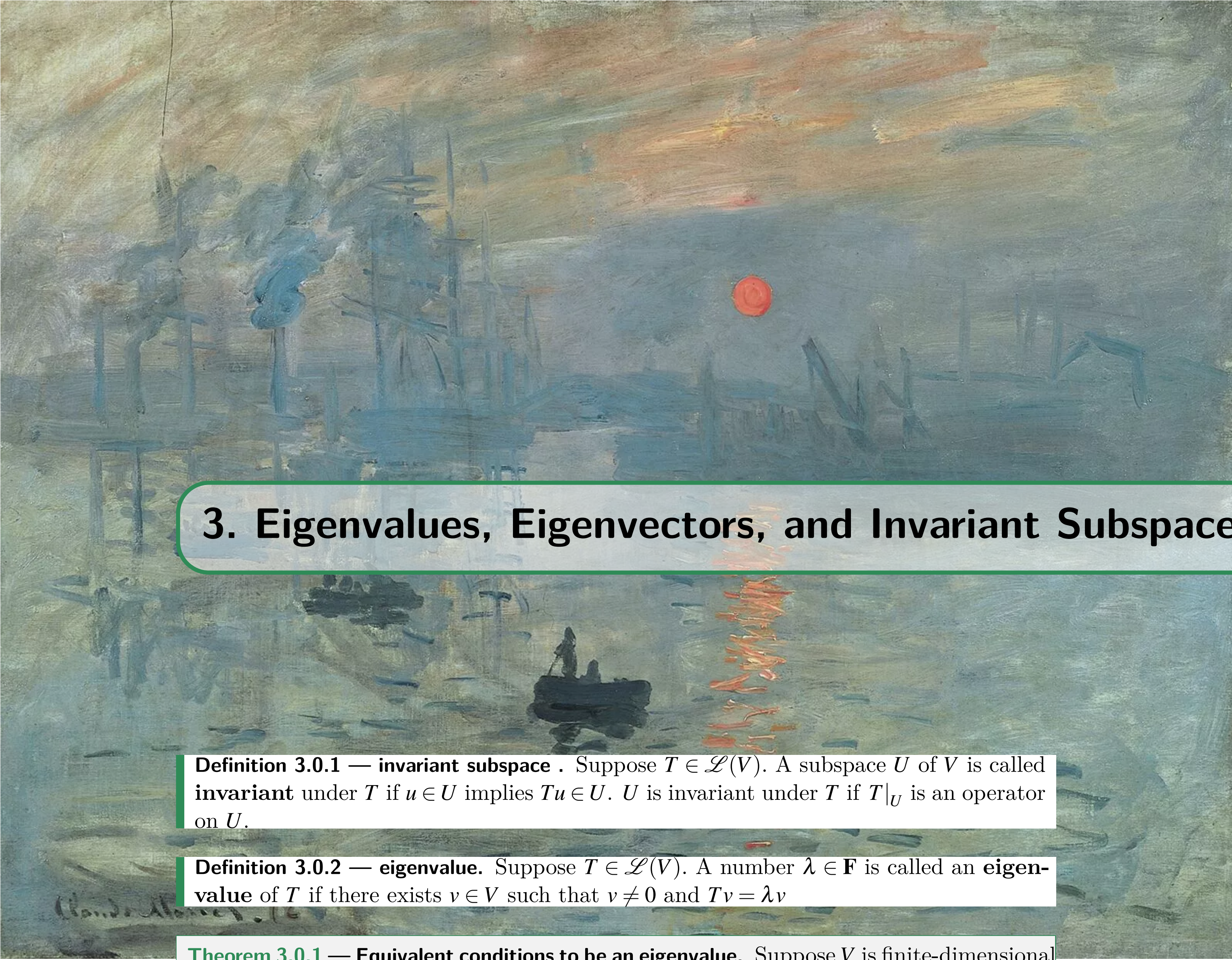
$$q(x) = a_0 + a_1 x + \cdots + a_{n-2} x^{n-2}$$

where $n = \deg p$ and $a_0, \ldots, a_{n-2} \in \mathbf{C}$, we thus have

$$0 = \operatorname{Im} q(x) = (\operatorname{Im} a_0) + (\operatorname{Im} a_1) x + \cdots + (\operatorname{Im} a_{n-2}) x^{n-2}$$

for all $x \in \mathbf{R}$. This implies that $\operatorname{Im} a_0, \ldots, \operatorname{Im} a_{n-2}$ all equal 0 (by 4.7). Thus all the coefficients of q are real, as desired. Hence the desired factorization exists.

Now we turn to the question of uniqueness of our factorization. A factor of p of the form $x^2 + b_j x + c_j$ with $b_j^2 < 4c_j$ can be uniquely written as $(x-\lambda_j)(x-\lambda_j)$ with $\lambda_j \in \mathbf{C}$. A moment's thought shows that two different factorizations of p as an element of $\mathscr{P}(\mathbf{R})$ would lead to two different factorizations of p as an element of $\mathscr{P}(\mathbf{C})$, contradicting 4.14. ∎



3. Eigenvalues, Eigenvectors, and Invariant Subspace

Definition 3.0.1 — invariant subspace . Suppose $T \in \mathscr{L}(V)$. A subspace U of V is called **invariant** under T if $u \in U$ implies $Tu \in U$. U is invariant under T if $T|_U$ is an operator on U.

Definition 3.0.2 — eigenvalue. Suppose $T \in \mathscr{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an **eigenvalue** of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$

Theorem 3.0.1 — Equivalent conditions to be an eigenvalue. Suppose V is finite-dimensional, $T \in \mathscr{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

1. λ is an eigenvalue of T;
2. $T - \lambda I$ is not injective;
3. $T - \lambda I$ is not surjective;
4. $T - \lambda I$ is not invertible.

Definition 3.0.3 — eigenvector . Suppose $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an **eigenvector** of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$

R Because $Tv = \lambda v$ if and only if $(T - \lambda I)v = 0$, a vector $v \in V$ with $v \neq 0$ is an eigenvector of T corresponding to λ if and only if $v \in \text{null}(T - \lambda I)$

Theorem 3.0.2 — Linearly independent eigenvectors . Let $T \in \mathscr{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding eigenvectors. Then $v_1, \ldots, v_m$ is linearly independent.

Proof. Suppose $v_1, \ldots, v_m$ is linearly dependent. Let k be the smallest positive integer such that

$$v_k \in \text{span}(v_1, \ldots, v_{k-1})$$

the existence of k with this property follows from the Linear Dependence Lemma. Thus there exist $a_1,\ldots,a_{k-1} \in \mathbf{F}$ such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$$

Apply T to both sides of this equation, getting

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}$$

Multiply both sides of by λ_k and then subtract the equation above, getting

$$0 = a_1 (\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1} (\lambda_k - \lambda_{k-1}) v_{k-1}$$

Because we chose k to be the smallest positive integer satisfying $v_1,\ldots,v_{k-1}$ is linearly independent. Thus the equation above implies that all the a 's are 0 (recall that λ_k is not equal to any of $\lambda_1,\ldots,\lambda_{k-1}$). However, this means that v_k equals 0, contradicting our hypothesis that v_k is an eigenvector. Therefore our assumption that $v_1,\ldots,v_m$ is linearly dependent was false. ■

Theorem 3.0.3 — Number of eigenvalues. Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Definition 3.0.4 — $T|_U$ and T/U. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T

The restriction operator $T|_U \in \mathcal{L}(U)$ is defined by

$$T|_U (u) = Tu$$

for $u \in U$

The quotient operator $T/U \in \mathcal{L}(V/U)$ is defined by

$$(T/U)(v+U) = Tv + U$$

for $v \in V$

Theorem 3.0.4 — Operators on complex vector spaces have an eigenvalue. Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Proof. Suppose V is a complex vector space with dimension $n > 0$ and $T \in \mathcal{L}(V)$. Choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2 v, \ldots, T^n v$$

is not linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist complex numbers $a_0,\ldots,a_n$, not all 0, such that

$$0 = a_0 v + a_1 Tv + \cdots + a_n T^n v$$

Note that $a_1,\ldots,a_n$ cannot all be 0, because otherwise the equation above would become $0 = a_0 v$, which would force a_0 also to be 0

Make the a 's the coefficients of a polynomial, which by the Fundamental Theorem of Algebra (4.14) has a factorization

$$a_0 + a_1 z + \cdots + a_n z^n = c(z - \lambda_1)\cdots(z - \lambda_m)$$

where c is a nonzero complex number, each λ_j is in $\mathbf{C}$, and the equation holds for all $z \in \mathbf{C}$ (here m is not necessarily equal to n, because a_n may equal 0). We then have

$$\begin{aligned} 0 &= a_0 v + a_1 T v + \cdots + a_n T^n v \\ &= (a_0 I + a_1 T + \cdots + a_n T^n) v \\ &= c (T - \lambda_1 I) \cdots (T - \lambda_m I) v \end{aligned}$$

Thus $T - \lambda_j I$ is not injective for at least one j. In other words, T has an eigenvalue. ■

Definition 3.0.5 — matrix of an operator. Suppose $T \in \mathscr{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. The matrix of T with respect to this basis is the n -by- n matrix

$$\mathscr{M}(T) = \begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \ldots & A_{n,n} \end{pmatrix}$$

whose entries $A_{j,k}$ are defined by

$$T v_k = A_{1,k} v_1 + \cdots + A_{n,k} v_n$$

If the basis is not clear from the context, then the notation $\mathscr{M}(T, (v_1, \ldots, v_n))$ is used.

R The k^{th} column of the matrix $\mathscr{M}(T)$ is formed from the coefficients used to write $T v_k$ as a linear combination of $v_1, \ldots, v_n$

Theorem 3.0.5 — Conditions for upper-triangular matrix. Suppose $T \in \mathscr{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent:

1. the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular;
2. $T v_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$
3. $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$

Proof. The equivalence of (a) and (b) follows easily from the definitions and a moment's thought. Obviously (c) implies (b). Hence to complete the proof, we need only prove that (b) implies (c). Thus suppose (b) holds. Fix $j \in \{1, \ldots, n\}$. From (b), we know that

$$\begin{aligned} & T v_1 \in \operatorname{span}(v_1) \subset \operatorname{span}(v_1, \ldots, v_j) \\ & T v_2 \in \operatorname{span}(v_1, v_2) \subset \operatorname{span}(v_1, \ldots, v_j) \\ & \quad \vdots \\ & T v_j \in \operatorname{span}(v_1, \ldots, v_j) \end{aligned}$$

Thus if v is a linear combination of $v_1, \ldots, v_j$, then

$$T v \in \operatorname{span}(v_1, \ldots, v_j)$$

In other words, $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T, completing the proof. ■

Theorem 3.0.6 — Over C, every operator has an upper-triangular matrix. Suppose V is a finite-dimensional complex vector space and $T \in \mathscr{L}(V)$ Then T has an upper-triangular matrix with respect to some basis of V.

Proof. We will use induction on the dimension of V. Clearly the desired result holds if $\dim V = 1$

Suppose now that $\dim V > 1$ and the desired result holds for all complex vector spaces whose dimension is less than the dimension of V. Let λ be any eigenvalue of T (3.0.4 guarantee T has an eigenvalue). Let

$$U = \text{range}(T - \lambda I)$$

Because $T - \lambda I$ is not surjective,$\dim U < \dim V$. Furthermore, U is invariant under T. To prove this, suppose $u \in U$. Then

$$Tu = (T - \lambda I)u + \lambda u$$

Obviously $(T - \lambda I)u \in U$ (because U equals the range of $T - \lambda I$) and $\lambda u \in U$. Thus the equation above shows that $Tu \in U$. Hence U is invariant under T, as claimed.

Thus $T|_U$ is an operator on U. By our induction hypothesis, there is a basis $u_1, \ldots \ldots u_m$ of U with respect to which $T|_U$ has an upper-triangular matrix. Thus for each j we have

$$Tu_j = (T|_U)(u_j) \in \text{span}(u_1, \ldots, u_j)$$

Extend $u_1, \ldots, u_m$ to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ of V. For each k, we have

$$Tv_k = (T - \lambda I)v_k + \lambda v_k$$

The definition of U shows that $(T - \lambda I)v_k \in U = \text{span}(u_1, \ldots, u_m)$. Thus the equation above shows that $Tv_k \in \text{span}(u_1 \ldots, u_m, v_1, \ldots, v_k)$. We conclude that T has an upper triangular matrix with respect to the basis $u_1 \ldots \ldots u_m, v_1, \ldots, v_n$ of V, as desired. ■

Theorem 3.0.7 — Determination of invertibility from upper-triangular matrix. Suppose $T \in \mathscr{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Proof. By using determinant. ■

Theorem 3.0.8 — Determination of eigenvalues from upper-triangular matrix. Suppose $T \in \mathscr{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

Proof. Suppose $v_1, \ldots, v_n$ is a basis of V with respect to which T has an upper-triangular matrix

$$\mathscr{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

Let $\lambda \in \mathbf{F}$. Then

$$\mathscr{M}(T-\lambda I)=\begin{pmatrix} \lambda_1-\lambda & & & * \\ & \lambda_2-\lambda & & \\ & & \ddots & \\ 0 & & & \lambda_n-\lambda \end{pmatrix}$$

Hence $T-\lambda I$ is not invertible if and only if λ equals one of the numbers $\lambda_1,\ldots,\lambda_n$. Thus λ is an eigenvalue of T if and only if λ equals one of the numbers $\lambda_1,\ldots,\lambda_n$ ■

R Be familiar with the language of 'Linear Map'

Definition 3.0.6 — diagonal matrix . A diagonal matrix is a square matrix that is 0 everywhere except possibly along the diagonal.

Definition 3.0.7 — eigenspace $E(\lambda,T)$. Suppose $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$. The eigenspace of T corresponding to λ denoted $E(\lambda,T)$, is defined by

$$E(\lambda,T)=\operatorname{null}(T-\lambda I)$$

In other words, $E(\lambda,T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

For $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$, the eigenspace $E(\lambda,T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V). The definitions imply that λ is an eigenvalue of T if and only if $E(\lambda,T)\neq\{0\}$

If λ is an eigenvalue of an operator $T \in \mathscr{L}(V)$, then T restricted to $E(\lambda,T)$ is just the operator of multiplication by λ:

$$Tv^k=\lambda v^k$$

Theorem 3.0.9 — Sum of eigenspaces is a direct sum. Suppose V is finite-dimensional and $T \in \mathscr{L}(V)$. Suppose also that $\lambda_1,\ldots,\lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1,T)+\cdots+E(\lambda_m,T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1,T)+\cdots+\dim E(\lambda_m,T)\leq \dim V$$

Proof. Because eigenvectors corresponding to distinct eigenvalues are linearly independent and

$$\dim E(\lambda_1,T)+\cdots+\dim E(\lambda_m,T)=\dim(E(\lambda_1,T)\oplus\cdots\oplus E(\lambda_m,T))\leq \dim V$$

■

Definition 3.0.8 — diagonalizable. An operator $T \in \mathscr{L}(V)$ is called diagonalizable if the operator has a diagonal matrix with respect to some basis of V

Theorem 3.0.10 — Conditions equivalent to diagonalizability. Suppose V is finite-dimensional and $T \in \mathscr{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

1. T is diagonalizable;
2. V has a basis consisting of eigenvectors of T
3. there exist 1 -dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \cdots \oplus U_n$
4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$
5. $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$

Proof. An operator $T \in \mathscr{L}(V)$ has a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with respect to a basis $v_1, \ldots, v_n$ of V if and only if $T v_j = \lambda_j v_j$ for each j Thus (a) and (b) are equivalent.

Suppose (b) holds; thus V has a basis $v_1, \ldots, v_n$ consisting of eigenvectors of T. For each j, let $U_j = \operatorname{span}(v_j)$. Obviously each U_j is a 1 -dimensional subspace of V that is invariant under T. Because $v_1, \ldots, v_n$ is a basis of V each vector in V can be written uniquely as a linear combination of $v_1, \ldots, v_n$ In other words, each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$ where each u_j is in U_j. Thus $V = U_1 \oplus \cdots \oplus U_n$. Hence (b) implies (c). Suppose now that (c) holds; thus there are 1 -dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \cdots \oplus U_n$ For each j, let v_j be a nonzero vector in U_j. Then each v_j is an eigenvector of T. Because each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$ where each u_j is in U_j (so each u_j is a scalar multiple of v_j), we see that $v_1, \ldots, v_n$ is a basis of V. Thus (c) implies (b).

At this stage of the proof we know that (a), (b), and (c) are all equivalent. We will finish the proof by showing that (b) implies (d), that (d) implies (e), and that (e) implies (b).

Suppose (b) holds; thus V has a basis consisting of eigenvectors of T Hence every vector in V is a linear combination of eigenvectors of T, which implies that

$$V = E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

Now 5.38 shows that (d) holds. That (d) implies (e) follows immediately from Exercise 16 in Section 2.C. Finally, suppose (e) holds; thus 5.42

$$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$$

Choose a basis of each $E(\lambda_j, T)$; put all these bases together to form a list $v_1, \ldots, v_n$ of eigenvectors of T, where $n = \dim V$ (by 5.42). To show that this list is linearly independent, suppose

$$a_1 v_1 + \cdots + a_n v_n = 0$$

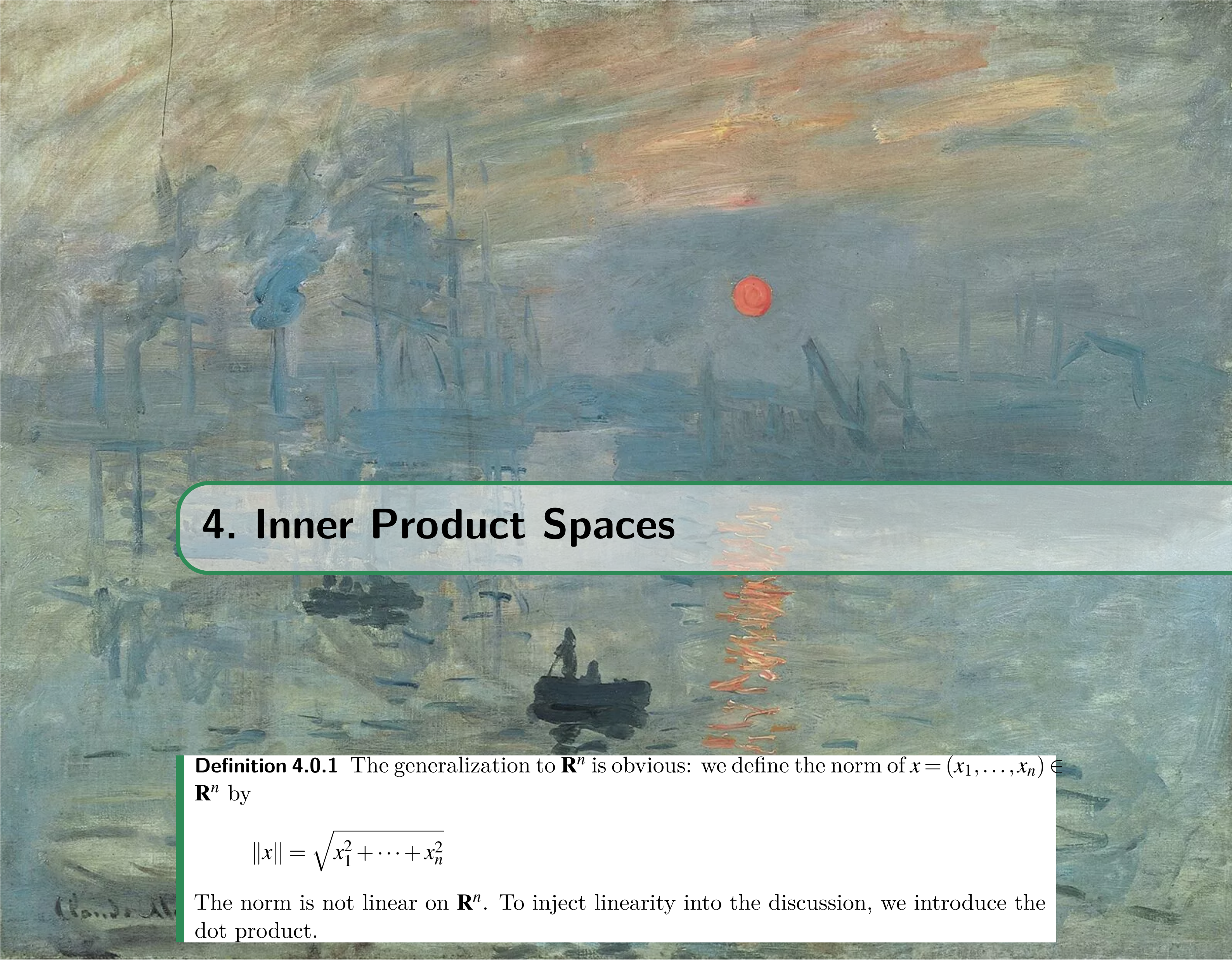
where $a_1, \ldots, a_n \in \mathbf{F}$. For each $j = 1, \ldots, m$, let u_j denote the sum of all the terms $a_k v_k$ such that $v_k \in E(\lambda_j, T)$. Thus each u_j is in $E(\lambda_j, T)$, and

$$u_1 + \cdots + u_m = 0$$

Because eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that each u_j equals 0. Because each u_j is a sum of terms $a_k v_k$, where the v_k 's were chosen to be a basis of $E(\lambda_j, T)$, this implies that all the a_k 's equal 0. Thus $v_1, \ldots, v_n$ is linearly independent and hence is a basis of V. Thus (e) implies (b), completing the proof. ■

Theorem 3.0.11 — Enough eigenvalues implies diagonalizability. If $T \in \mathscr{L}(V)$ has dim V distinct eigenvalues, then T is diagonalizable.

Proof. By eigenvectors corresponding to distinct eigenvalues are linearly independent. ■



4. Inner Product Spaces

Definition 4.0.1 The generalization to $\mathbf{R}^n$ is obvious: we define the norm of $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ by

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

The norm is not linear on $\mathbf{R}^n$. To inject linearity into the discussion, we introduce the dot product.

Definition 4.0.2 — Definition dot product. For $x, y \in \mathbf{R}^n$, the dot product of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Note that the dot product of two vectors in $\mathbb{R}^n$ is a number, not a vector.

Proposition 4.0.1 An inner product on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ and has the following properties:

1. positivity

$$\langle v, v\rangle \geq 0 \text{ for all } v \in V$$

2. definiteness

$$\langle v, v\rangle = 0 \text{ if and only if } v = 0$$

3. additivity in first slot

$$\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle \text{ for all } u, v, w \in V$$

4. homogeneity in first

$$\langle \lambda u, v\rangle = \lambda \langle u, v\rangle \text{for all} \lambda \in \textbf{Fand all} u, v \in V$$

5. conjugate symmetry

$$\langle u,v\rangle = \overline{\langle v,u\rangle} \text{ for all } u,v\in V$$

■ **Example 4.1** The Euclidean inner product on $\mathbf{F}^n$ is defined by

$$\langle (w_1,\dots,w_n),(z_1,\dots,z_n)\rangle = w_1\overline{z_1}+\cdots+w_n\overline{z_n}$$

■

Definition 4.0.3 — Definition inner product space. An inner product space is a vector space V along with an inner product on V

Proposition 4.0.2 Basic properties of an inner product

1. For each fixed $u\in V$, the function that takes v to $\langle v,u\rangle$ is a linear map from V to $\mathbf{F}$.
2. $\langle 0,u\rangle = 0$ for every $u\in V$
3. $\langle u,0\rangle = 0$ for every $u\in V$
4. $\langle u,v+w\rangle = \langle u,v\rangle + \langle u,w\rangle$ for all $u,v,w\in V$
5. $\langle u,\lambda v\rangle = \bar{\lambda}\langle u,v\rangle$ for all $\lambda\in\mathbf{F}$ and $u,v\in V$

Definition 4.0.4 — each inner product determines a norm. For $v\in V$, the norm of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v,v\rangle}$$

Proposition 4.0.3 — Basic properties of the norm: . Suppose $v\in V$

1. $\|v\| = 0$ if and only if $v=0$
2. $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda\in\mathbf{F}$

Definition 4.0.5 — Definition orthogonal. Two vectors $u,v\in V$ are called orthogonal if $\langle u,v\rangle = 0$

Theorem 4.0.4 Orthogonality and 0

1. 0 is orthogonal to every vector in V
2. 0 is the only vector in V that is orthogonal to itself.

Theorem 4.0.5 — Pythagorean Theorem. Suppose u and v are orthogonal vectors in V. Then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

Theorem 4.0.6 — An orthogonal decomposition. Suppose $u,v\in V$, with $v\neq 0$. Set

$$c = \frac{\langle u,v\rangle}{\|v\|^2}$$

and

$$w = u - \frac{\langle u,v\rangle}{\|v\|^2}v.$$

Then

$$\langle w, v\rangle = 0 \quad \text{and} \quad u = cv + w$$

Theorem 4.0.7 — Cauchy-Schwarz Inequality . Suppose $u, v \in V$. Then

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.

Proof. If $v = 0$, then both sides of the desired inequality equal 0. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + w$$

where w is orthogonal to v. By the Pythagorean Theorem,

$$\begin{aligned}
\|u\|^2 &= \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 + \|w\|^2 \\
&= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\
&\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}
\end{aligned}$$

Multiplying both sides of this inequality by $\|v\|^2$ and then taking square roots gives the desired inequality.

Looking at the proof in the paragraph above, note that the Cauchy-Schwarz Inequality is an equality if and only if 6.16 is an equality. Obviously this happens if and only if $w = 0$. But $w = 0$ if and only if u is a multiple of v. Thus the Cauchy-Schwarz Inequality is an equality if and only if u is a scalar multiple of v or v is a scalar multiple of u (or both; the phrasing has been chosen to cover cases in which either u or v equals 0).

■

Theorem 4.0.8 — Triangle Inequality. Suppose $u, v \in V$. Then

$$\|u + v\| \leq \|u\| + \|v\|$$

This inequality is an equality if and only if one of u, v is a nonnegative multiple of the other.

Proof. We have

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle u, v\rangle \\
&= \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u, v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| \\
&= (\|u\| + \|v\|)^2
\end{aligned}$$

■

Theorem 4.0.9 — Parallelogram Equality. Suppose $u,v \in V$. Then

$$\|u+v\|^2+\|u-v\|^2=2\left(\|u\|^2+\|v\|^2\right)$$

Proof. We have

$$\begin{aligned}\|u+v\|^2+\|u-v\|^2 =&\langle u+v,u+v\rangle+\langle u-v,u-v\rangle\\ =&\|u\|^2+\|v\|^2+\langle u,v\rangle+\langle v,u\rangle\\ &+\|u\|^2+\|v\|^2-\langle u,v\rangle-\langle v,u\rangle\\ =&2\left(\|u\|^2+\|v\|^2\right)\end{aligned}$$

as desired. ■

Definition 4.0.6 — Orthonormal . A list of vectors is called orthonormal if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list. In other words, a list $e_1,\dots,e_m$ of vectors in V is orthonormal if

$$\langle e_j,e_k\rangle=\begin{cases}1 & \text{if } j=k\\ 0 & \text{if } j\neq k\end{cases}$$

Theorem 4.0.10 — The norm of an orthonormal linear combination. If $e_1,\dots,e_m$ is an orthonormal list of vectors in V, then

$$\|a_1e_1+\cdots+a_me_m\|^2=|a_1|^2+\cdots+|a_m|^2$$

for all $a_1,\dots,a_m \in \mathbf{F}$

Theorem 4.0.11 — An orthonormal list is linearly independent. Every orthonormal list of vectors is linearly independent.

Definition 4.0.7 — Orthonormal basis. An orthonormal basis of V is an orthonormal list of vectors in V that is also a basis of V

Theorem 4.0.12 — An orthonormal list of the right length is an orthonormal basis. Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

Theorem 4.0.13 — Writing a vector as linear combination of orthonormal basis. Suppose $e_1,\dots,e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v=\langle v,e_1\rangle e_1+\cdots+\langle v,e_n\rangle e_n$$

and

$$\|v\|^2=|\langle v,e_1\rangle|^2+\cdots+|\langle v,e_n\rangle|^2$$

Proof. Because $e_1,\dots,e_n$ is a basis of V, there exist scalars $a_1,\dots,a_n$ such that

$$v=a_1e_1+\cdots+a_ne_n$$

Because $e_1,\ldots,e_n$ is orthonormal, taking the inner product of both sides of this equation with e_j gives $\langle v,e_j\rangle = a_j$. Thus the first equation in 6.30 holds. ■

Theorem 4.0.14 — Gram-Schmidt Procedure. Suppose $v_1,\ldots,v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2,\ldots,m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j,e_1\rangle e_1 - \cdots - \langle v_j,e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j,e_1\rangle e_1 - \cdots - \langle v_j,e_{j-1}\rangle e_{j-1}\|}$$

Then $e_1,\ldots,e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1,\ldots,v_j) = \operatorname{span}(e_1,\ldots,e_j)$$

for $j = 1,\ldots,m$

Proof. We will show by induction on j that the desired conclusion holds. To get started with $j = 1$, note that $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ because v_1 is a positive multiple of e_1. Suppose $1 < j < m$ and we have verified that

$$\operatorname{span}(v_1,\ldots,v_{j-1}) = \operatorname{span}(e_1,\ldots,e_{j-1}) \tag{4.1}$$

Note that $v_j \notin \operatorname{span}(v_1,\ldots,v_{j-1})$ (because $v_1,\ldots,v_m$ is linearly independent). Thus $v_j \notin \operatorname{span}(e_1,\ldots,e_{j-1})$. Hence we are not dividing by 0 in the definition of e_j. Dividing a vector by its norm produces a new vector with norm 1; thus $\|e_j\| = 1$. Let $1 \leq k < j$. Then

$$\begin{aligned}
\langle e_j,e_k\rangle &= \left\langle \frac{v_j - \langle v_j,e_1\rangle e_1 - \cdots - \langle v_j,e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j,e_1\rangle e_1 - \cdots - \langle v_j,e_{j-1}\rangle e_{j-1}\|}, e_k \right\rangle \\
&= \frac{\langle v_j,e_k\rangle - \langle v_j,e_k\rangle}{\|v_j - \langle v_j,e_1\rangle e_1 - \cdots - \langle v_j,e_{j-1}\rangle e_{j-1}\|} \\
&= 0
\end{aligned}$$

Thus $e_1,\ldots,e_j$ is an orthonormal list. From the definition of e_j, we see that $v_j \in \operatorname{span}(e_1,\ldots,e_j)$ Combining this information4.1 shows that

$$\operatorname{span}(v_1,\ldots,v_j) \subset \operatorname{span}(e_1,\ldots,e_j)$$

Both lists above are linearly independent (the v 's by hypothesis, the e 's by orthonormality). Thus both subspaces above have dimension j, and hence they are equal, completing the proof. ■

Theorem 4.0.15 — Existence of orthonormal basis. Every finite-dimensional inner product space has an orthonormal basis.

Theorem 4.0.16 — Orthonormal list extends to orthonormal basis. Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Theorem 4.0.17 — Upper-triangular matrix with respect to orthonormal basis. Suppose $T \in \mathscr{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V

Proof. Suppose T has an upper-triangular matrix with respect to some basis $v_1,\ldots,v_n$ of V. Thus $\operatorname{span}(v_1,\ldots,v_j)$ is invariant under T for each $j = 1,\ldots,n$(see 3.0.5) Apply the Gram-Schmidt Procedure to $v_1,\ldots,v_n$, producing an orthonormal basis $e_1,\ldots,e_n$ of V. Because

$$\operatorname{span}(e_1,\ldots,e_j) = \operatorname{span}(v_1,\ldots,v_j)$$

for each j , we conclude that $\operatorname{span}(e_1,\ldots,e_j)$ is invariant under T for each $j = 1,\ldots,n$. Thus, by 3.0.5,T has an upper-triangular matrix with respect to the orthonormal basis $e_1,\ldots,e_n$ ■

Theorem 4.0.18 — Schur's Theorem. Suppose V is a finite-dimensional complex vector space and $T \in \mathscr{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V

Proof. Over $\mathbf{C}$, every operator has an upper-triangular matrix ■

Theorem 4.0.19 — Riesz Representation Theorem. Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v,u\rangle$$

for every $v \in V$

Proof. First we show there exists a vector $u \in V$ such that $\varphi(v) = \langle v,u\rangle$ for every $v \in V$. Let $e_1,\ldots,e_n$ be an orthonormal basis of V. Then

$$\begin{aligned}\varphi(v) &= \varphi\left(\langle v,e_1\rangle e_1 + \cdots + \langle v,e_n\rangle e_n\right)\\ &= \langle v,e_1\rangle \varphi(e_1) + \cdots + \langle v,e_n\rangle \varphi(e_n)\\ &= \left\langle v, \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n\right\rangle\end{aligned}$$

for every $v \in V$, where the first equality comes from 6.30. Thus setting 6.43

$$u = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$$

we have $\varphi(v) = \langle v,u\rangle$ for every $v \in V$, as desired. Now we prove that only one vector $u \in V$ has the desired behavior. Suppose $u_1,u_2 \in V$ are such that

$$\varphi(v) = \langle v,u_1\rangle = \langle v,u_2\rangle$$

for every $v \in V$. Then

$$0 = \langle v,u_1\rangle - \langle v,u_2\rangle = \langle v,u_1-u_2\rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$. In other words, $u_1 = u_2$, completing the proof of the uniqueness part of the result.

■

Definition 4.0.8 — Definition orthogonal complement, $U^{\perp}$. If U is a subset of V, then the orthogonal complement of U, denoted $U^{\perp}$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^{\perp} = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}$$

Proposition 4.0.20 Basic properties of orthogonal complement

1. If U is a subset of V, then $U^{\perp}$ is a subspace of V
2. $\{0\}^{\perp} = V$
3. $V^{\perp} = \{0\}$
4. If U is a subset of V, then $U \cap U^{\perp} \subset \{0\}$.
5. If U and W are subsets of V and $U \subset W$, then $W^{\perp} \subset U^{\perp}$

Theorem 4.0.21 Recall that if U, W are subspaces of V, then V is the direct sum of U and W (written $V = U \oplus W$) if each element of V can be written in exactly one way as a vector in U plus a vector in W .

Theorem 4.0.22 — Direct sum of a subspace and its orthogonal complement. Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^{\perp}$$

Proof. First we will show that 6.48

$$V = U + U^{\perp}$$

To do this, suppose $v \in V$. Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Obviously

$$v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m}_{w}$$

Let u and w be defined as in the equation above. Clearly $u \in U$. Because $e_1, \ldots, e_m$ is an orthonormal list, for each $j = 1, \ldots, m$ we have

$$\langle w, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$. In other words, $w \in U^{\perp}$. Thus we have written $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. From previous theorem, we know that $U \cap U^{\perp} = \{0\}$. Along with 6.48, this implies that $V = U \oplus U^{\perp}$.

■

Theorem 4.0.23 — Dimension of the orthogonal complement. Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U$$

Theorem 4.0.24 — The orthogonal complement of the orthogonal complement. Suppose U is a finite-dimensional subspace of V. Then $U = \left(U^{\perp}\right)^{\perp}$

Proof. First we will show that $U \subset \left(U^{\perp}\right)^{\perp}$ To do this, suppose $u \in U$. Then $\langle u, v\rangle = 0$ for every $v \in U^{\perp}$ (by the definition of $U^{\perp}$). Because u is orthogonal to every vector in $U^{\perp}$, we have $u \in \left(U^{\perp}\right)^{\perp}$.

To prove the inclusion in the other direction, suppose $v \in \left(U^{\perp}\right)^{\perp}$. We can write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. We have $v - u = w \in U^{\perp}$. Because $v \in \left(U^{\perp}\right)^{\perp}$ and $u \in \left(U^{\perp}\right)^{\perp}$(from 6.52), we have $v - u \in \left(U^{\perp}\right)^{\perp}$. Thus $v - u \in U^{\perp} \cap \left(U^{\perp}\right)^{\perp}$, which implies that $v - u$ is orthogonal to itself, which implies that $v - u = 0$, which implies that $v = u$, which implies that $v \in U$. Thus $\left(U^{\perp}\right)^{\perp} \subset U$.

■

Definition 4.0.9 — Orthogonal projection, P_U . Suppose U is a finite-dimensional subspace of V. The orthogonal projection of V onto U is the operator $P_U \in \mathscr{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then $P_U v = u$

Proposition 4.0.25 — Properties of the orthogonal projection P_U. Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

1. $P_U \in \mathscr{L}(V)$
2. $P_U u = u$ for every $u \in U$
3. $P_U w = 0$ for every $w \in U^{\perp}$
4. range $P_U = U$
5. null $P_U = U^{\perp}$
6. $v - P_U v \in U^{\perp}$
7. $P_U^2 = P_U$
8. $\|P_U v\| \leq \|v\|$
9. for every orthonormal basis $e_1, \ldots, e_m$ of $U, P_U v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$

Theorem 4.0.26 — Minimizing the distance to a subspace. Suppose U is a finite-dimensional subspace of $V, v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \leq \|v - u\|$$

Furthermore, the inequality above is an equality if and only if $u = P_U v$

Proof. We have

$$\begin{aligned} \|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2 \end{aligned}$$

The second line above comes from the Pythagorean Theorem $\left[\text{which applies because } v - P_U v \in U^{\perp}, \text{ and } P_U v - u \in U\right]$,

Our inequality above is an equality if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$

■

■ **Example 4.2** Example Find a polynomial u with real coefficients and degree at most 5 that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 dx$$

is as small as possible. Compare this result to the Taylor series approximation. Solution Let $C_{\mathbf{R}}[-\pi, \pi]$ denote the real inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

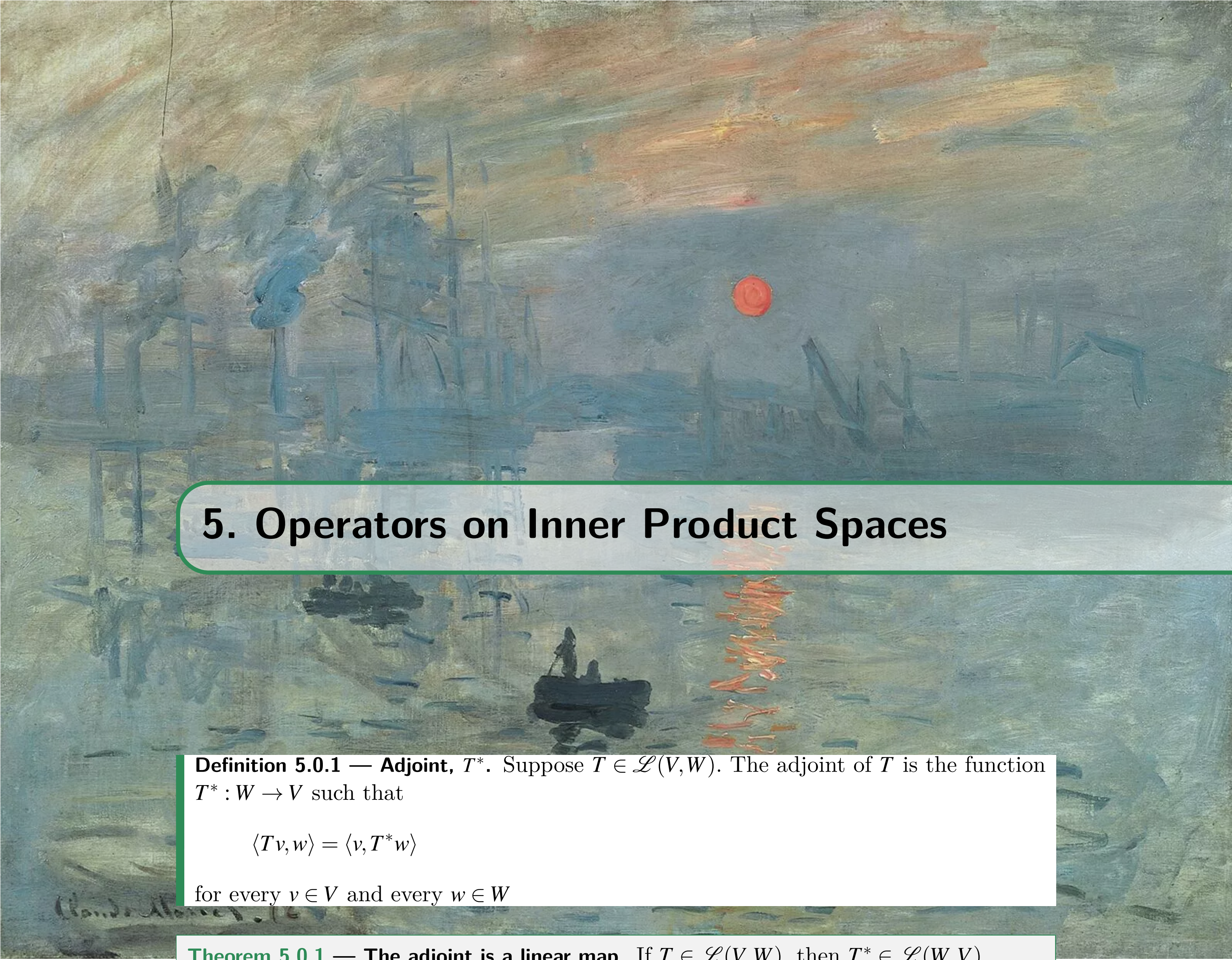
$$\langle f, g\rangle=\int_{-\pi}^{\pi} f(x) g(x) d x$$

Let $v \in C_{\mathbf{R}}[-\pi, \pi]$ be the function defined by $v(x)=\sin x$. Let U denote the subspace of $C_{\mathbf{R}}[-\pi, \pi]$ consisting of the polynomials with real coefficients and degree at most 5. Our problem can now be reformulated as follows: Find $u \in U$ such that $\|v-u\|$ is as small as possible.

To compute the solution to our approximation problem, first apply the grations is useful here. Gram-Schmidt Procedure (using the in- ner product given by 6.59) to the basis $1, x, x^{2}, x^{3}, x^{4}, x^{5}$ of U, producing an orthonormal basis $e_{1}, e_{2}, e_{3}, e_{4}, e_{5}, e_{6}$ of U. Then, again using the inner product, compute $P_{U} v$ (with $m=6$). Doing this computation shows that $P_{U} v$ is the function u defined by

$$u(x)=0.987862 x-0.155271 x^{3}+0.00564312 x^{5}$$

where the π 's that appear in the exact answer have been replaced with a good decimal approximation. ■



5. Operators on Inner Product Spaces

Definition 5.0.1 — Adjoint, T^*. Suppose $T \in \mathscr{L}(V,W)$. The adjoint of T is the function $T^* : W \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for every $v \in V$ and every $w \in W$

Theorem 5.0.1 — The adjoint is a linear map. If $T \in \mathscr{L}(V,W)$, then $T^* \in \mathscr{L}(W,V)$

Proposition 5.0.2 — Properties of the adjoint. Properties of the adjoint

1. $(S+T)^* = S^* + T^*$ for all $S,T \in \mathscr{L}(V,W)$
2. $(\lambda T)^* = \bar{\lambda} T^*$ for all $\lambda \in \mathbf{F}$ and $T \in \mathscr{L}(V,W)$
3. $(T^*)^* = T$ for all $T \in \mathscr{L}(V,W)$
4. $I^* = I$, where I is the identity operator on V
5. $(ST)^* = T^*S^*$ for all $T \in \mathscr{L}(V,W)$ and $S \in \mathscr{L}(W,U)$ (here U is an inner product space over $\mathbf{F}$).

Theorem 5.0.3 — Null space and range of T^*. Suppose $T \in \mathscr{L}(V,W)$. Then

1. null $T^* = (\text{ range } T)^\perp$
2. range $T^* = (\text{ null } T)^\perp$
3. null $T = (\text{ range } T^*)^\perp$
4. range $T = (\text{ null } T^*)^\perp$

Proof. We begin by proving (a). Let $w \in W$. Then

$$\begin{aligned}
w \in \text{ null } T^* &\iff T^*w = 0 \\
&\iff \langle v, T^*w\rangle = 0 \text{ for all } v \in V \\
&\iff \langle Tv, w\rangle = 0 \text{ for all } v \in V \\
&\iff w \in (\text{ range } T)^\perp
\end{aligned}$$

■

Definition 5.0.2 — conjugate transpose. The conjugate transpose of an m -by- n matrix is the n -by- m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry.

Definition 5.0.3 — The matrix of T^*. Let $T \in \mathscr{L}(V,W)$. Suppose $e_1,\ldots,e_n$ is an orthonormal basis of V and $f_1,\ldots.f_m$ is an orthonormal basis of W. Then

$$\mathscr{M}\left(T^*,(f_1,\ldots,f_m),(e_1,\ldots,e_n)\right)$$

is the conjugate transpose of

$$\mathscr{M}\left(T,(e_1,\ldots,e_n),(f_1,\ldots,f_m)\right)$$

Definition 5.0.4 — self-adjoint. An operator $T \in \mathscr{L}(V)$ is called self-adjoint if $T = T^*$. In other words, $T \in \mathscr{L}(V)$ is self-adjoint if and only if

$$\langle Tv,w\rangle = \langle v,Tw\rangle$$

for all $v,w \in V$

Theorem 5.0.4 — Eigenvalues of self-adjoint operators are real. Every eigenvalue of a self-adjoint operator is real.

Proof.

$$\lambda\|v\|^2 = \langle\lambda v,v\rangle = \langle Tv,v\rangle = \langle v,Tv\rangle = \langle v,\lambda v\rangle = \bar{\lambda}\|v\|^2$$

■

Theorem 5.0.5 — Over C, Tv is orthogonal to v for all v only for the 0 operator. Suppose V is a complex inner product space and $T \in \mathscr{L}(V)$. Suppose

$$\langle Tv,v\rangle = 0$$

for all $v \in V$. Then $T = 0$

Proof. We have

$$\begin{aligned}\langle Tu,w\rangle =& \frac{\langle T(u+w),u+w\rangle - \langle T(u-w),u-w\rangle}{4}\\ &+ \frac{\langle T(u+iw),u+iw\rangle - \langle T(u-iw),u-iw\rangle}{4}i\end{aligned}$$

■

Theorem 5.0.6 — Over $\mathbf{C}$, $\langle Tv,v\rangle$ is real for all v only for self-adjoint operators. Suppose V is a complex inner product space and $T \in \mathscr{L}(V)$. Then T is self-adjoint if and only if

$$\langle Tv,v\rangle \in \mathbf{R}$$

for every $v \in V$

Proof. Let $v \in V$. Then $\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \langle T^* v, v\rangle = \langle (T - T^*)v, v\rangle$ ■

Theorem 5.0.7 — If $T = T^*$ and $\langle Tv, v\rangle = 0$ for all v, then $T = 0$. Suppose T is a self-adjoint operator on V such that

$$\langle Tv, v\rangle = 0$$

for all $v \in V$. Then $T = 0$

Proof. We have already proved this (without the hypothesis that T is self-adjoint) when V is a complex inner product space . Thus we can assume that V is a real inner product space. If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4}$$

this is proved by computing the right side using the equation

$$\langle Tw, u\rangle = \langle w, Tu\rangle = \langle Tu, w\rangle$$

■

Definition 5.0.5 — normal . An operator on an inner product space is called **normal** if it commutes with its adjoint.
In other words, $T \in \mathcal{L}(V)$ is normal if

$$TT^* = T^*T$$

R Obviously every self-adjoint operator is normal.

Theorem 5.0.8 null T = null T^* for every normal operator T

Theorem 5.0.9 — T is normal if and only if $\|Tv\| = \|T^*v\|$ for all v . An operator $T \in \mathcal{L}(V)$ is normal if and only if

$$\|Tv\| = \|T^*v\|$$

for all $v \in V$

Proof. Let $T \in \mathcal{L}(V)$. We will prove both directions of this result at the same time. Note that

$$\begin{aligned} T\text{is normal} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \quad \text{for all } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \ \text{ for all } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \quad \text{for all } v \in V \end{aligned}$$

■

Theorem 5.0.10 — For T normal, T and T^* have the same eigenvectors. Suppose $T \in \mathscr{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$

Proof. Because T is normal, so is $T - \lambda I$, $[(T-\lambda I)\cdot\left(T^* - \bar{\lambda} I\right)]$. Using 7.20 we have

$$0 = \|(T-\lambda I)v\| = \|(T-\lambda I)^* v\| = \left\|\left(T^* - \bar{\lambda} I\right) v\right\|$$

■

Theorem 5.0.11 — Orthogonal eigenvectors for normal operators. Suppose $T \in \mathscr{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof. Suppose α, β are distinct eigenvalues of T, with corresponding eigenvectors u, v. Thus $Tu = \alpha u$ and $Tv = \beta v$. From 7.21 we have $T^* v = \bar{\beta} v$ Thus

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \bar{\beta} v\rangle \\ &= \langle Tu, v\rangle - \langle u, T^* v\rangle \\ &= 0\end{aligned}$$

Because $\alpha \neq \beta$, the equation above implies that $\langle u, v\rangle = 0$. Thus u and v are orthogonal, as desired. ■

Theorem 5.0.12 — Complex Spectral Theorem. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathscr{L}(V)$. Then the following are equivalent:

1. T is normal.
2. V has an orthonormal basis consisting of eigenvectors of T
3. T has a diagonal matrix with respect to some orthonormal basis of V

Proof. First suppose (c) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. The matrix of T^* (with respect to the same basis) is obtained by taking the conjugate transpose of the matrix of T; henceT^* also has a diagonal matrix. Any two diagonal matrices commute; thus T commutes with T^*, which means that T is normal. In other words, (a) holds. Now suppose (a) holds, so T is normal. By Schur's Theorem (6.38), there is an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which T has an upper-triangular matrix. Thus we can write

$$\mathscr{M}\left(T, (e_1, \ldots, e_n)\right) = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}$$

We will show that this matrix is actually a diagonal matrix.

We see from the matrix above that

$$\|Te_1\|^2 = |a_{1,1}|^2$$

and

$$\|T^* e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2$$

Because T is normal, $\|Te_1\| = \|T^*e_1\|$. Thus the two equations above imply that all entries in the first row of the matrix in 7.25, except possibly the first entry $a_{1,1}$, equal 0 Now we see that

$$\|Te_2\|^2 = |a_{2,2}|^2$$

(because $a_{1,2} = 0$, as we showed in the paragraph above) and

$$\|T^*e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + \cdots + |a_{2,n}|^2$$

Because T is normal, $\|Te_2\| = \|T^*e_2\|$. Thus the two equations above imply that all entries in the second row of the matrix in 7.25, except possibly the diagonal entry $a_{2,2}$, equal 0

Continuing in this fashion, we see that all the non-diagonal entries in the matrix equal 0. Thus (c) holds. (B)by: an operator on V has a diagonal matrix with respect to a basis if and only if the basis consists of eigenvectors of the operator . ■

R

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0$$

Theorem 5.0.13 — Invertible quadratic expressions. Suppose $T \in \mathscr{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$ Then

$$T^2 + bT + cI$$

is invertible.

Proof. Let v be a nonzero vector in V. Then

$$\begin{aligned}
\left\langle \left(T^2 + bT + cI\right) v, v \right\rangle &= \left\langle T^2 v, v \right\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\
&= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 \\
&= \left(\|Tv\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\
&> 0
\end{aligned}$$

where the third line above holds by the Cauchy-Schwarz Inequality 4.0.7 The last inequality implies that $\left(T^2 + bT + cI\right) v \neq 0$. Thus $T^2 + bT + cI$ is injective, which implies that it is invertible. ■

Theorem 5.0.14 — Self-adjoint operators have eigenvalues. Suppose $V \neq \{0\}$ and $T \in \mathscr{L}(V)$ is a self-adjoint operator. Then T has an eigenvalue.

Proof. We can assume that V is a real inner product space, as we have already noted. Let $n = \dim V$ and choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2 v, \ldots, T^n v$$

cannot be linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist real numbers $a_0, \ldots, a_n$, not all 0, such that

$$0 = a_0 v + a_1 T v + \cdots + a_n T^n v$$

Make the a 's the coefficients of a polynomial, which can be written in factored form as

$$\begin{aligned} &a_0 + a_1 x + \cdots + a_n x^n \\ &\quad = c\left(x^2 + b_1 x + c_1\right) \cdots \left(x^2 + b_M x + c_M\right)(x - \lambda_1) \cdots (x - \lambda_m) \end{aligned}$$

where c is a nonzero real number, each b_j, c_j, and λ_j is real, each b_j^2 is less than $4c_j, m+M \geq 1$, and the equation holds for all real x. We then have

$$\begin{aligned} 0 &= a_0 v + a_1 T v + \cdots + a_n T^n v \\ &= (a_0 I + a_1 T + \cdots + a_n T^n) v \\ &= c\left(T^2 + b_1 T + c_1 I\right) \cdots \left(T^2 + b_M T + c_M I\right)(T - \lambda_1 I) \cdots (T - \lambda_m I) v \end{aligned}$$

Each $T^2 + b_j T + c_j I$ is invertible. Recall also that $c \neq 0$. Thus the equation above implies that $m > 0$ and

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_m I) v$$

Hence $T - \lambda_j I$ is not injective for at least one j. In other words, T has an eigenvalue. ■

Theorem 5.0.15 — Self-adjoint operators and invariant subspaces. Suppose $T \in \mathscr{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T. Then

1. $U^{\perp}$ is invariant under T
2. $T|_U \in \mathscr{L}(U)$ is self-adjoint;
3. $T|_{U^{\perp}} \in \mathscr{L}\left(U^{\perp}\right)$ is self-adjoint.

Proof. To prove (a), suppose $v \in U^{\perp}$. Let $u \in U$. Then

$$\langle T v, u \rangle = \langle v, T u \rangle = 0$$

where the first equality above holds because T is self-adjoint and the second equality above holds because U is invariant under T (and hence $Tu \in U$) and because $v \in U^{\perp}$. Because the equation above holds for each $u \in U$, we conclude that $Tv \in U^{\perp}$. Thus $U^{\perp}$ is invariant under T, completing the proof of (a). To prove (b), note that if $u, v \in U$, then

$$\langle (T|_U) u, v \rangle = \langle T u, v \rangle = \langle u, T v \rangle = \langle u, (T|_U) v \rangle$$

Thus $T|_U$ is self-adjoint. Now (c) follows from replacing U with $U^{\perp}$ in (b), which makes sense by (a). ■

Theorem 5.0.16 — Real Spectral Theorem. Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathscr{L}(V)$. Then the following are equivalent:

1. T is self-adjoint.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix with respect to some orthonormal basis of V.

Proof. First suppose (c) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. A diagonal matrix equals its transpose. Hence $T = T^*$, and thus T is self-adjoint. In other words, (a) holds.

We will prove that (a) implies (b) by induction on dim V. To get started, note that if $\dim V = 1$, then (a) implies (b). Now assume that $\dim V > 1$ and that (a) implies (b) for all real inner product spaces of smaller dimension. Suppose (a) holds, so $T \in \mathscr{L}(V)$ is self-adjoint. Let u be an eigenvector of T with $\|u\| = 1$ (5.0.14 guarantees that T has an eigenvector, which can then be divided by its norm to produce an eigenvector with norm 1). Let $U = \operatorname{span}(u)$. Then U is a 1 -dimensional subspace of V that is invariant under T. (By eigenvector's def) . The operator $T|_{U\perp} \in \mathscr{L}\left(U^{\perp}\right)$ is self-adjoint. By our induction hypothesis, there is an orthonormal basis of $U^{\perp}$ consisting of eigenvectors of $T|_{U\perp}$. Adjoining u to this orthonormal basis of $U^{\perp}$ gives an orthonormal basis of V consisting of eigenvectors of T, completing the proof that (a) implies (b).

We have proved that (c) implies (a) and that (a) implies (b). Clearly (b) implies (c), completing the proof. ■

Definition 5.0.6 — positive operator. An operator $T \in \mathscr{L}(V)$ is called positive if T is self-adjoint and

$$\langle Tv, v\rangle \geq 0$$

for all $v \in V$

■ **Example 5.1** If U is a subspace of V, then the orthogonal projection P_U is a positive operator, as you should verify. ■

Definition 5.0.7 — square root . An operator R is called a square root of an operator T if $R^2 = T$

Theorem 5.0.17 — Characterization of positive operators. Let $T \in \mathscr{L}(V)$. Then the following are equivalent:
(a) T is positive;
(b) T is self-adjoint and all the eigenvalues of T are nonnegative;
(c) T has a positive square root;
(d) T has a self-adjoint square root
(e) there exists an operator $R \in \mathscr{L}(V)$ such that $T = R^*R$

Proof. We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a)

First suppose (a) holds, so that T is positive. Obviously T is self-adjoint (by the definition of a positive operator). To prove the other condition in (b), suppose λ is an eigenvalue of T. Let v be an eigenvector of T corresponding to λ. Then

$$0 \leq \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle$$

Thus λ is a nonnegative number. Hence (b) holds.

Now suppose (b) holds, so that T is self-adjoint and all the eigenvalues of T are nonnegative. By the Spectral Theorem 5.0.12,5.0.16, there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T corresponding to $e_1, \ldots, e_n$,; thus each λ_j is a nonnegative number. Let R be the linear map from V to V such that

$$Re_j = \sqrt{\lambda_j} e_j$$

for $j=1,\ldots,n$. Then R is a positive operator, as you should verify. Furthermore, $R^2e_j=\lambda_je_j=Te_j$ for each j, which implies that $R^2=T$ Thus R is a positive square root of T. Hence (c) holds.

Clearly (c) implies (d) (because, by definition, every positive operator is self-adjoint).

Now suppose (d) holds, meaning that there exists a self-adjoint operator R on V such that $T=R^2$. Then $T=R^*R$ (because $R^*=R$). Hence (e) holds.

Finally, suppose (e) holds. Let $R\in\mathcal{L}(V)$ be such that $T=R^*R$. Then $T^*=(R^*R)^*=R^*(R^*)^*=R^*R=T$. Hence T is self-adjoint. To complete the proof that (a) holds, note that

$$\langle Tv,v\rangle=\langle R^*Rv,v\rangle=\langle Rv,Rv\rangle\geq 0$$

for every $v\in V$. Thus T is positive.

■

Theorem 5.0.18 — Each positive operator has only one positive square root. Every positive operator on V has a unique positive square root.

Proof. Suppose $T\in\mathcal{L}(V)$ is positive. Suppose $v\in V$ is an eigenvector of T.(Why does this operator have a eigenvector?5.0.14) Thus there exists $\lambda\geq 0$ such that $Tv=\lambda v$

Let R be a positive square root of T. We will prove that $Rv=\sqrt{\lambda}v$. This will imply that the behavior of R on the eigenvectors of T is uniquely determined. Because there is a basis of V consisting of eigenvectors of T (by the Spectral Theorem), this will imply that R is uniquely determined. To prove that $Rv=\sqrt{\lambda}v$, note that the Spectral Theorem asserts that there is an orthonormal basis $e_1,\ldots,e_n$ of V consisting of eigenvectors of R. Because R is a positive operator, all its eigenvalues are nonnegative. Thus there exist nonnegative numbers $\lambda_1,\ldots,\lambda_n$ such that $Re_j=\sqrt{\lambda_j}e_j$ for $j=1,\ldots,n$

Because $e_1,\ldots,e_n$ is a basis of V, we can write

$$v=a_1e_1+\cdots+a_ne_n$$

for some numbers $a_1,\ldots,a_n\in\mathbf{F}$. Thus

$$Rv=a_1\sqrt{\lambda_1}e_1+\cdots+a_n\sqrt{\lambda_n}e_n$$

and hence

$$R^2v=a_1\lambda_1e_1+\cdots+a_n\lambda_ne_n$$

But $R^2=T$, and $Tv=\lambda v$. Thus the equation above implies

$$a_1\lambda e_1+\cdots+a_n\lambda e_n=a_1\lambda_1e_1+\cdots+a_n\lambda_ne_n$$

The equation above implies that $a_j(\lambda-\lambda_j)=0$ for $j=1,\ldots,n$. Hence

$$v=\sum_{\{j:\lambda_j=\lambda\}}a_je_j$$

and thus

$$Rv=\sum_{\{j:\lambda_j=\lambda\}}a_j\sqrt{\lambda}e_j=\sqrt{\lambda}v$$

as desired.

■

R A positive operator can have infinitely many square roots (although only one of them can be positive). For example, the identity operator on V has infinitely many square roots if $\dim V > 1$

Definition 5.0.8 — isometry. An operator $S \in \mathscr{L}(V)$ is called an isometry if $\|Sv\| = \|v\|$ for all $v \in V$. In other words, an operator is an isometry if it preserves norms.

Theorem 5.0.19 Suppose $\lambda_1 \ldots \ldots \lambda_n$ are scalars with absolute value 1 and $S \in \mathscr{L}(V)$ satisfies $Se_j = \lambda_j e_j$ for some orthonormal basis $e_1, \ldots, e_n$ of V . Show that S is an isometry.

Proof. Suppose $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \tag{5.1}$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 \tag{5.2}$$

Applying S to both sides of 5.1 gives

$$\begin{aligned} Sv &= \langle v, e_1 \rangle Se_1 + \cdots + \langle v, e_n \rangle Se_n \\ &= \lambda_1 \langle v, e_1 \rangle e_1 + \cdots + \lambda_n \langle v, e_n \rangle e_n \end{aligned}$$

The last equation, along with the equation $|\lambda_j| = 1$, shows that

$$\|Sv\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

Comparing 5.2 shows that $\|v\| = \|Sv\|$. In other words, S is an isometry. ■

Theorem 5.0.20 — Every isometry is normal,Characterization of isometries . Suppose $S \in \mathscr{L}(V)$. Then the following are equivalent:

1. S is an isometry;
2. $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$
3. $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V
4. there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal
5. $S^*S = I$
6. $SS^* = I$
7. S^* is an isometry;
8. S is invertible and $S^{-1} = S^*$

Proof. First suppose (a) holds, so S is an isometry. Exercises 19 and 20 in Section 6.A show that inner products can be computed from norms. Because S preserves norms, this implies that S preserves inner products, and hence (b) holds. More precisely, if V is a real inner product space, then for every $u, v \in V$ we have

$$\begin{aligned} \langle Su, Sv \rangle &= \left(\|Su + Sv\|^2 - \|Su - Sv\|^2 \right) / 4 \\ &= \left(\|S(u+v)\|^2 - \|S(u-v)\|^2 \right) / 4 \\ &= \left(\|u+v\|^2 - \|u-v\|^2 \right) / 4 \\ &= \langle u, v \rangle \end{aligned}$$

where the first equality comes from Exercise 19 in Section 6. A, the second equality comes from the linearity of S, the third equality holds because S is an isometry, and the last equality again comes from Exercise 19 in Section 6. A. If V is a complex inner product space, then use Exercise 20 in Section 6.A instead of Exercise 19 to obtain the same conclusion. In either case, we see that (b) holds.

Now suppose (b) holds, so S preserves inner products. Suppose that $e_1,\ldots,e_n$ is an orthonormal list of vectors in V. Then we see that the list $Se_1,\ldots,Se_n$ is orthonormal because $\left\langle Se_j,Se_k\right\rangle = \left\langle e_j,e_k\right\rangle$. Thus (c) holds. Clearly (c) implies (d). Now suppose (d) holds. Let $e_1,\ldots,e_n$ be an orthonormal basis of V such that $Se_1,\ldots,Se_n$ is orthonormal. Thus

$$\left\langle S^*Se_j,e_k\right\rangle = \left\langle e_j,e_k\right\rangle$$

for $j,k=1,\ldots,n$ [because the term on the left equals $\left\langle Se_j,Se_k\right\rangle$ and $(Se_1,\ldots,Se_n)$ is orthonormal]. **All vectors $u,v\in V$ can be written as linear combinations of $e_1,\ldots,e_n$, and thus the equation above implies that** $\langle S^*Su,v\rangle = \langle u,v\rangle$. Hence $S^*S=I$; in other words, (e) holds.

Now suppose (e) holds, so that $S^*S=I$. In general, an operator S need not commute with S^*. However, $S^*S=I$ if and only if $SS^*=I$; this is a special case of Exercise 10 in Section 3. D. Thus $SS^*=I$, showing that (f) holds. Now suppose (f) holds, so $SS^*=I$. If $v\in V$, then

$$\|S^*v\|^2 = \langle S^*v,S^*v\rangle = \langle SS^*v,v\rangle = \langle v,v\rangle = \|v\|^2$$

Thus S^* is an isometry, showing that (g) holds. Now suppose (g) holds, so S^* is an isometry. We know that (a) $\Rightarrow$ (e) and (a) $\Rightarrow$ (f) because we have shown (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f). Using the implications (a) $\Rightarrow$ (e) and (a) $\Rightarrow$ (f) but with S replaced with S^*[and using the equation $(S^*)^* = S$], we conclude that $SS^*=I$ and $S^*S=I$ Thus S is invertible and $S^{-1}=S^*$; in other words, (h) holds.

Now suppose (h) holds, so S is invertible and $S^{-1}=S^*$. Thus $S^*S=I$. If $v\in V$, then

$$\|Sv\|^2 = \langle Sv,Sv\rangle = \langle S^*Sv,v\rangle = \langle v,v\rangle = \|v\|^2$$

Thus S is an isometry, showing that (a) holds. We have shown (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (g) $\Rightarrow$ (h) $\Rightarrow$ (a) completing the proof. ■

Exercise 5.1 — 6.A 19 . Suppose V is a real inner product space. Prove that

$$\langle u,v\rangle = \frac{\|u+v\|^2-\|u-v\|^2}{4}$$

for all $u,v\in V$ ■

Exercise 5.2 — 6.A 20 . Suppose V is a complex inner product space. Prove that

$$\langle u,v\rangle = \frac{\|u+v\|^2-\|u-v\|^2+\|u+iv\|^2 i-\|u-iv\|^2 i}{4}$$

for all $u,v\in V$ ■

Exercise 5.3 — 3.D 10. Suppose V is finite-dimensional and $S,T \in \mathcal{L}(V)$. Prove that $ST = I$ if and only if $TS = I$.

Theorem 5.0.21 — Description of isometries when $\mathbf{F} = \mathbf{C}$. Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

1. S is an isometry.
2. There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1

Proof. We have already shown that (b) implies (a). To prove the other direction, suppose (a) holds, so S is an isometry. By the Complex Spectral Theorem (7.24), there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of S. For $j \in \{1, \ldots, n\}$, let λ_j be the eigenvalue corresponding to e_j. Then

$$|\lambda_j| = \|\lambda_j e_j\| = \|Se_j\| = \|e_j\| = 1$$

Thus each eigenvalue of S has absolute value 1, completing the proof. ∎

Definition 5.0.9 — $\sqrt{T}$. If T is a positive operator, then $\sqrt{T}$ denotes the unique positive square root of T

Theorem 5.0.22 — Polar Decomposition. Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}$$

Proof. If $v \in V$, then

$$\begin{aligned}
\|Tv\|^2 = \langle Tv, Tv \rangle = \langle T^*Tv, v \rangle \\
&= \langle \sqrt{T^*T}\sqrt{T^*T}v, v \rangle \\
&= \langle \sqrt{T^*T}v, \sqrt{T^*T}v \rangle \\
&= \|\sqrt{T^*T}v\|^2
\end{aligned}$$

$\sqrt{T^*T}$ is positive for 5.0.17. Thus

$$\|Tv\| = \|\sqrt{T^*T}v\|$$

for all $v \in V$. Define a linear map $S_1 : \text{range}\,\sqrt{T^*T} \to \text{range}\,T$

$$S_1(\sqrt{T^*T}v) = Tv$$

The idea of the proof is to extend S_1 to an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Now for the details.

First we must check that S_1 is well defined. To do this, suppose $v_1, v_2 \in V$ are such that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$. For the definition to make sense, we must show that $Tv_1 = Tv_2$. Note that

$$\begin{aligned}
\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| \\
&= \left\|\sqrt{T^*T}(v_1 - v_2)\right\| \\
&= \left\|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\right\| \\
&= 0
\end{aligned}$$

The equation above shows that $Tv_1 = Tv_2$, so S_1 is indeed well defined. You should verify that S_1 is a linear map.

We see that S_1 maps range $\sqrt{T^*T}$ onto range T, and

$$\|S_1 u\| = \|u\|$$

for all $u \in$ range $\sqrt{T^*T}$ The rest of the proof extends S_1 to an isometry S on all of V. In particular, S_1 is injective. Thus from the Fundamental Theorem of Linear Maps (3.22), applied to S_1, we have

$$\dim \operatorname{range} \sqrt{T^*T} = \dim \operatorname{range} T$$

This implies that $\dim(\text{ range } \sqrt{T^*T})^\perp = \dim(\text{ range } T)^\perp$ (By $\dim U^\perp = \dim V - \dim U$). Thus orthonormal bases $e_1, \ldots, e_m$ of $(\text{range } \sqrt{T^*T})^\perp$ and $f_1, \ldots, f_m$ of $(\text{range } T)^\perp$ can be chosen; the key point here is that these two orthonormal bases have the same length (denoted m). Now define a linear map $S_2 : (\text{ range } \sqrt{T^*T})^\perp \to (\text{ range } T)^\perp$ by

$$S_2(a_1 e_1 + \cdots + a_m e_m) = a_1 f_1 + \cdots + a_m f_m$$

For all $w \in (\text{range } \sqrt{T^*T})^\perp$, we have $\|S_2 w\| = \|w\|$. Now let S be the operator on V that equals S_1 on range $\sqrt{T^*T}$ and equals S_2 on $(\text{range } \sqrt{T^*T})^\perp$. More precisely, recall that each $v \in V$ can be written uniquely in the form $v = u + w$ where $u \in$ range $\sqrt{T^*T}$ and $w \in (\text{ range } \sqrt{T^*T})^\perp$. For $v \in V$ with decomposition as above, define Sv by

$$Sv = S_1 u + S_2 w$$

For each $v \in V$ we have

$$S(\sqrt{T^*T}v) = S_1(\sqrt{T^*T}v) = Tv$$

so $T = S\sqrt{T^*T}$, as desired. All that remains is to show that S is an isometry. However, this follows easily from two uses of the Pythagorean Theorem: if $v \in V$ has decomposition $v = u + w$, then

$$\|Sv\|^2 = \|S_1 u + S_2 w\|^2 = \|S_1 u\|^2 + \|S_2 w\|^2 = \|u\|^2 + \|w\|^2 = \|v\|^2$$

the second equality holds because $S_1 u \in$ range T and $S_2 w \in (\text{ range } T)^\perp$. ■

R The Polar Decomposition states that each operator on V is the product of an isometry and a positive operator. Specifically, consider the case $\mathbf{F} = \mathbf{C}$, and suppose $T = S\sqrt{T^*T}$ is a Polar Decomposition of an operator $T \in \mathscr{L}(V)$, where S is an isometry. Then there is an orthonormal basis of V with respect to which S has a diagonal matrix, and there is an orthonormal basis of V with respect to which $*T$ has a diagonal matrix. (Note: two bases.)

Definition 5.0.10 — singular value. Suppose $T \in \mathscr{L}(V)$. The singular values of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

Proposition 5.0.23
1. The singular values of T are all nonnegative, because they are the eigenvalues of the positive operator $\sqrt{T^*T}$
2. Each $T \in L(V)$ has dim V singular values
3. Every operator on V has a clean description in terms of its singular values and two orthonormal bases of V.

Theorem 5.0.24 — Singular Value Decomposition. Suppose $T \in \mathscr{L}(V)$ has singular values $s_1,\ldots,s_n$. Then there exist orthonormal bases $e_1,\ldots,e_n$ and $f_1,\ldots,f_n$ of V such that

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n$$

for every $v \in V$

Proof. By the Spectral Theorem applied to $\sqrt{T^*T}$, there is an orthonormal basis $e_1,\ldots,e_n$ of V such that $\sqrt{T^*T}e_j = s_j e_j$ for $j = 1,\ldots,n$. We have

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

for every $v \in V$. Apply $\sqrt{T^*T}$ to both sides of this equation, getting

$$\sqrt{T^*T}v = s_1 \langle v, e_1 \rangle e_1 + \cdots + s_n \langle v, e_n \rangle e_n$$

for every $v \in V$. By the Polar Decomposition, there is an isometry $S \in \mathscr{L}(V)$ such that $T = S\sqrt{T^*T}$. Apply S to both sides of the equation above, getting

$$Tv = s_1 \langle v, e_1 \rangle Se_1 + \cdots + s_n \langle v, e_n \rangle Se_n$$

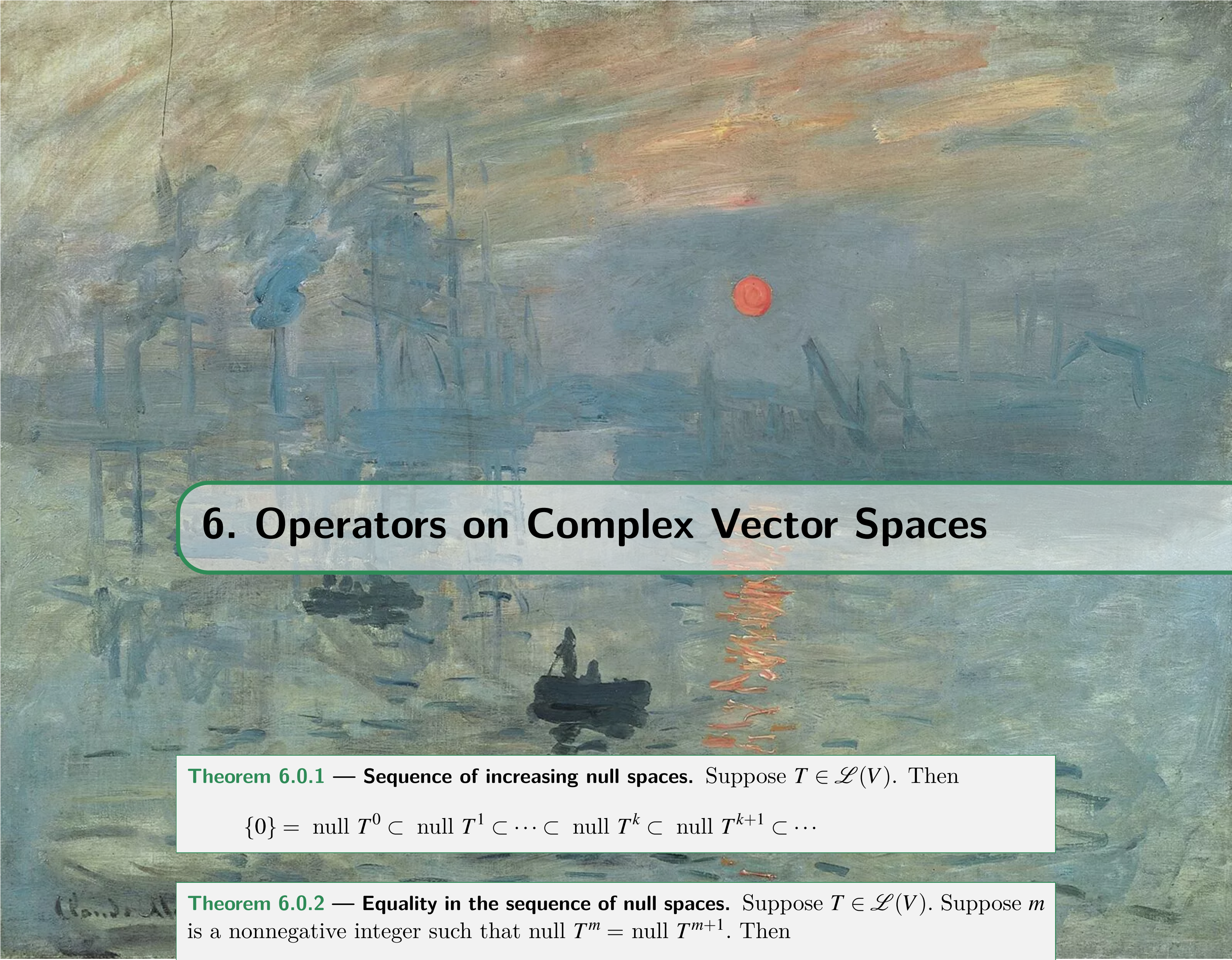
for every $v \in V$. For each j, let $f_j = Se_j$. Because S is an isometry, $f_1, \ldots, f_n$ is an orthonormal basis of V. The equation above now becomes

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n$$

for every $v \in V$, completing the proof. ■

Theorem 5.0.25 — Singular values without taking square root of an operator. Suppose $T \in \mathscr{L}(V)$. Then the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, with each eigenvalue λ repeated $\dim E(\lambda, T^*T)$ times.

Proof. The Spectral Theorem implies that there are an orthonormal basis $e_1,\ldots,e_n$ and nonnegative numbers $\lambda_1,\ldots,\lambda_n$ such that $T^*Te_j = \lambda_j e_j$ for $j = 1,\ldots,n$. It is easy to see that $\sqrt{T^*T}e_j = \sqrt{\lambda_j}e_j$ for $j = 1,\ldots,n$ which implies the desired result. ■



6. Operators on Complex Vector Spaces

Theorem 6.0.1 — Sequence of increasing null spaces. Suppose $T \in \mathscr{L}(V)$. Then

$$\{0\} = \text{ null } T^0 \subset \text{ null } T^1 \subset \cdots \subset \text{ null } T^k \subset \text{ null } T^{k+1} \subset \cdots$$

Theorem 6.0.2 — Equality in the sequence of null spaces. Suppose $T \in \mathscr{L}(V)$. Suppose m is a nonnegative integer such that null $T^m =$ null T^{m+1}. Then

$$\text{null } T^m = \text{ null } T^{m+1} = \text{ null } T^{m+2} = \text{ null } T^{m+3} = \cdots$$

Proof. To prove $T^{m+k} \supset$ null T^{m+k+1}, suppose $v \in$ null T^{m+k+1}. Then

$$T^{m+1}\left(T^k v\right) = T^{m+k+1} v = 0$$

Hence

$$T^k v \in \text{ null } T^{m+1} = \text{ null } T^m$$

Thus $T^{m+k} v = T^m \left(T^k v\right) = 0$, which means that $v \in$ null T^{m+k}. This implies that null $T^{m+k+1} \subset$ null T^{m+k}, completing the proof. ■

Theorem 6.0.3 — Null spaces stop growing. Suppose $T \in \mathscr{L}(V)$. Let $n = \dim V$. Then

$$\text{null } T^n = \text{ null } T^{n+1} = \text{ null } T^{n+2} = \cdots$$

Proof. We need only prove that null $T^n =$ null T^{n+1} . Suppose this is not true. Then, we have

$$\{0\} = \text{ null } T^0 \subsetneq \text{ null } T^1 \subsetneq \cdots \subsetneq \text{ null } T^n \subsetneq \text{ null } T^{n+1}$$

where the symbol $\subsetneq$ means "contained in but not equal to". At each of the strict inclusions in the chain above, the dimension increases by at least 1 Thus dim null $T^{n+1} \geq n+1$, a contradiction because a subspace of V cannot have a larger dimension than n ■

Theorem 6.0.4 — V is the direct sum of null $T^{\dim V}$ and range $T^{\dim V}$. Suppose $T \in \mathscr{L}(V)$. Let $n = \dim V$. Then

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n$$

Proof. First we show that

$$(\operatorname{null} T^n) \cap (\text{ range } T^n) = \{0\}$$

Suppose $v \in$ (null T^n)$\cap$ (range T^n). Then $T^n v = 0$, and there exists $u \in V$ such that $v = T^n u$. Applying T^n to both sides of the last equation shows that $T^n v = T^{2n} u$. Hence $T^{2n} u = 0$, which implies that $T^n u = 0$. Thus $v = T^n u = 0$, completing the proof of 8.6 Now null T^n+ range T^n is a direct sum . Also, dim(null $T^n \oplus$ range T^n) = dim null T^n+ dim range $T^n = \dim V$ ■

Definition 6.0.1 — generalized eigenvector. Suppose $T \in \mathscr{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a generalized eigenvector of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j

Definition 6.0.2 — generalized eigenspace, $G(\lambda, T)$. Suppose $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$. The generalized eigenspace of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

R If $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$, then

$$E(\lambda, T) \subset G(\lambda, T)$$

Theorem 6.0.5 — Description of generalized eigenspaces. Suppose $T \in \mathscr{L}(V)$ and $\lambda \in \mathbf{F}$. Then $G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}$

Proof. Suppose $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. The definitions imply $v \in G(\lambda, T)$ Thus $G(\lambda, T) \supset$ null $(T - \lambda I)^{\dim V}$ Conversely, suppose $v \in G(\lambda, T)$. Thus there is a positive integer j such that $v \in \operatorname{null}(T - \lambda I)^j$ From 6.0.1 and 6.0.3 (with $T - \lambda I$ replacing T), we get $v \in \operatorname{null}(T - \lambda I)^{\dim V}$ Thus $G(\lambda, T) \subset \operatorname{null}(T - \lambda I)^{\dim V}$, completing the proof. ■

Theorem 6.0.6 — Linearly independent generalized eigenvectors. Let $T \in \mathscr{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \ldots, v_m$ is linearly independent.

Proof. Suppose $a_1,\dots,a_m$ are complex numbers such that 8.14 $0 = a_1v_1 + \cdots + a_mv_m$ Let k be the largest nonnegative integer such that $(T-\lambda_1 I)^k v_1 \neq 0$. Let $w = (T-\lambda_1 I)^k v_1$ Thus

$$(T-\lambda_1 I)w = (T-\lambda_1 I)^{k+1} v_1 = 0 \tag{6.1}$$

and hence $Tw = \lambda_1 w$. Thus $(T-\lambda I)w = (\lambda_1 - \lambda)w$ for every $\lambda \in \mathbf{F}$ and hence

$$(T-\lambda I)^n w = (\lambda_1 - \lambda)^n w \tag{6.2}$$

for every $\lambda \in \mathbf{F}$, where $n = \dim V$. Apply the operator

$$(T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n$$

to both sides of 6.1 ,getting

$$\begin{aligned} 0 &= a_1 (T-\lambda_1 I)^k (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n v_1 \\ &= a_1 (T-\lambda_2 I)^n \cdots (T-\lambda_m I)^n w \\ &= a_1 (\lambda_1-\lambda_2)^n \cdots (\lambda_1-\lambda_m)^n w \end{aligned}$$

where we have used 6.0.5 to get the first equation above and 6.2 to get the last equation above.

The equation above implies that $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, which implies that $v_1,\dots,v_m$ is linearly independent.

■

Definition 6.0.3 — nilpotent. An operator is called nilpotent if some power of it equals 0 .

Definition 6.0.4 — Nilpotent operator raised to dimension of domain is 0. Suppose $N \in \mathscr{L}(V)$ is nilpotent. Then $N^{\dim V} = 0$

Proof. Because N is nilpotent, $G(0,N) = V$. Thus 6.0.5 implies that null $N^{\dim V} = V$, as desired. ■

Theorem 6.0.7 — Matrix of a nilpotent operator. Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}$$

here all entries on and below the diagonal are 0 ’s.

Proof. First choose a basis of null N. Then extend this to a basis of null N^2 Then extend to a basis of null N^3. Continue in this fashion, eventually getting a basis of V (because 8.18 states that null $N^{\dim V} = V$).

Now let’s think about the matrix of N with respect to this basis. The first column, and perhaps additional columns at the beginning, consists of all 0’s, because the corresponding basis vectors are in null N. The next set of columns comes from basis vectors in null N^2. Applying N to any such vector, we get a vector in null N; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus all nonzero entries in these columns lie above the diagonal. The next set of columns comes from basis vectors in null N^3. Applying N to any such vector, we get a vector in null N^2; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus once again, all nonzero entries in these columns lie above the diagonal. Continue in this fashion to complete the proof. ∎

R The k^{th} column of $\mathscr{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $(w_1,\ldots,w_m)$ $Tv_k = \sum_{i=1}^{m} A_{j,k}w_j$

Theorem 6.0.8 — The null space and range of $p(T)$ are invariant under T. Suppose $T \in \mathscr{L}(V)$ and $p \in \mathscr{P}(\mathbf{F})$. Then null $p(T)$ and range $p(T)$ are invariant under T

Theorem 6.0.9 — Description of operators on complex vector spaces. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Let $\lambda_1,\ldots,\lambda_m$ be the distinct eigenvalues of T. Then

1. $V = G(\lambda_1,T)\oplus\cdots\oplus G(\lambda_m,T)$
2. each $G(\lambda_j,T)$ is invariant under T
3. each $(T-\lambda_j I)|_{G(\lambda_j,T)}$ is nilpotent.

Proof. Let $n = \dim V$. Recall that $G(\lambda_j,T) = \text{null}\,(T-\lambda_j I)^n$ for each j. From 6.0.8 [with $p(z) = (z-\lambda_j)^n$], we get (b). Obviously (c) follows from the definitions.

We will prove (a) by induction on n. To get started, note that the desired result holds if $n = 1$. Thus we can assume that $n > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because V is a complex vector space, T has an eigenvalue ; thus $m \geq 1$. Applying 6.0.4 to $T-\lambda_1 I$ shows that

$$V = G(\lambda_1,T)\oplus U \tag{6.3}$$

where $U = \text{range}\,(T-\lambda_1 I)^n$. Using 6.0.8[with $p(z) = (z-\lambda_1)^n$], we see that U is invariant under T. Because $G(\lambda_1,T) \neq \{0\}$, we have $\dim U < n$ Thus we can apply our induction hypothesis to $T|_U$

None of the generalized eigenvectors of $T|_U$ correspond to the eigenvalue λ_1, because all generalized eigenvectors of T corresponding to λ_1 are in $G(\lambda_1,T)$. Thus each eigenvalue of $T|_U$ is in $\{\lambda_2,\ldots,\lambda_m\}$ By our induction hypothesis, $U = G(\lambda_2, T|_U)\oplus\cdots\oplus G(\lambda_m, T|_U)$. Combining this information with 6.3 will complete the proof if we can show that $G(\lambda_k, T|_U) = G(\lambda_k,T)$ for $k = 2,\ldots,m$

Thus fix $k \in \{2,\ldots,m\}$. The inclusion $G(\lambda_k, T|_U) \subset G(\lambda_k,T)$ is clear. To prove the inclusion in the other direction, suppose $v \in G(\lambda_k,T)$. By 6.3, we can write $v = v_1 + u$, where $v_1 \in G(\lambda_1,T)$ and $u \in U$. Our induction hypothesis implies that

$$u = v_2 + \cdots + v_m$$

where each v_j is in $G(\lambda_j, T|_U)$, which is a subset of $G(\lambda_j,T)$. Thus

$$v = v_1 + v_2 + \cdots + v_m$$

Because generalized eigenvectors corresponding to distinct eigenvalues are linearly independent ($v \in G(\lambda_k,T)$), the equation above implies that each v_j equals 0 except possibly when $j = k$. In particular, $v_1 = 0$ and thus $v = u \in U$ Because $v \in U$, we can conclude that $v \in G(\lambda_k, T|_U)$, completing the proof. ∎

Theorem 6.0.10 — A basis of generalized eigenvectors. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T

Definition 6.0.5 — multiplicity. Suppose $T \in \mathscr{L}(V)$. The multiplicity of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$ In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$

Theorem 6.0.11 — Sum of the multiplicities equals dim V . Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals dim V.

Definition 6.0.6 The terms algebraic multiplicity and geometric multiplicity are used in some books. In case you encounter this terminology, be aware that the algebraic multiplicity is the same as the multiplicity defined here and the geometric multiplicity is the dimension of the corresponding eigenspace. In other words, if $T \in \mathscr{L}(V)$ and λ is an eigenvalue of T, then algebraic multiplicity of $\lambda = \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T)$ geometric multiplicity of $\lambda = \dim \operatorname{null}(T - \lambda I) = \dim E(\lambda, T)$

Definition 6.0.7 — block diagonal matrix. A block diagonal matrix is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

where $A_1, \ldots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0

Theorem 6.0.12 — Block diagonal matrix with upper-triangular blocks. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

where each A_j is a d_j -by- d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}$$

Proof. Each $(T - \lambda_j I)|_{G(\lambda_i, T)}$ is nilpotent . For each j, choose a basis of $G(\lambda_j, T)$, which is a vector space with dimension d_j, such that the matrix of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ with respect to this basis is as in 6.0.7 . Thus the matrix of $T|_{G(\lambda_j, T)}$, which equals $(T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$, with respect to this basis will look like the desired form shown above for A_j

Putting the bases of the $G(\lambda_j,T)$'s together gives a basis of V. The matrix of T with respect to this basis has the desired form. ■

Theorem 6.0.13 — Identity plus nilpotent has a square root. Suppose $N \in \mathscr{L}(V)$ is nilpotent. Then $I+N$ has a square root.

Proof. Consider the Taylor series for the function $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots \tag{6.4}$$

Because $a_1 = 1/2$, the formula above shows that $1+x/2$ is a good estimate for $\sqrt{1+x}$ when x is small. We will not find an explicit formula for the coefficients or worry about whether the infinite sum converges because we will use this equation only as motivation.

Because N is nilpotent, $N^m = 0$ for some positive integer m. In 6.4 suppose we replace x with N and 1 with I. Then the infinite sum on the right side becomes a finite sum (because $N^j = 0$ for all $j \geq m$). In other words, we guess that there is a square root of $I+N$ of the form

$$I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}$$

Having made this guess, we can try to choose $a_1, a_2, \ldots, a_{m-1}$ such that the operator above has its square equal to $I+N$. Now

$$\begin{aligned}&\left(I + a_1 N + a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1}\right)^2 \\ &\quad = I + 2a_1 N + \left(2a_2 + a_1^2\right) N^2 + \left(2a_3 + 2a_1 a_2\right) N^3 + \cdots \\ &\quad + \left(2a_{m-1} + \text{ terms involving } a_1, \ldots, a_{m-2}\right) N^{m-1}\end{aligned}$$

We want the right side of the equation above to equal $I+N$. Hence choose a_1 such that $2a_1 = 1$ (thus $a_1 = 1/2$). Next, choose a_2 such that $2a_2 + a_1^2 = 0$ (thus $a_2 = -1/8$). Then choose a_3 such that the coefficient of N^3 on the right side of the equation above equals 0 (thus $a_3 = 1/16$). Continue in this fashion for $j = 4, \ldots, m-1$, at each step solving for a_j so that the coefficient of N^j on the right side of the equation above equals 0. Actually we do not care about the explicit formula for the a_j 's. We need only know that some choice of the a_j 's gives a square root of $I+N$. ■

Theorem 6.0.14 — Over C, invertible operators have square roots. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$ is invertible. Then T has a square root.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. For each j, there exists a nilpotent operator $N_j \in \mathscr{L}\left(G(\lambda_j, T)\right)$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$ [see 8.21(c)]. Because T is invertible, none of the λ_j 's equals 0, so we can write

$$T|_{G(\lambda_j,T)} = \lambda_j \left(I + \frac{N_j}{\lambda_j}\right)$$

for each j. Clearly N_j/λ_j is nilpotent, and so $I + N_j/\lambda_j$ has a square root (by 8.31). Multiplying a square root of the complex number λ_j by a square root of $I + N_j/\lambda_j$, we obtain a square root R_j of $T|_{G(\lambda_j,T)}$ A typical vector $v \in V$ can be written uniquely in the form

$$v = u_1 + \cdots + u_m$$

where each u_j is in $G(\lambda_j, T)$ (see 8.21). Using this decomposition, define an operator $R \in \mathscr{L}(V)$ by

$$Rv = R_1 u_1 + \cdots + R_m u_m$$

You should verify that this operator R is a square root of T, completing the proof. ■

Definition 6.0.8 — characteristic polynomial. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the characteristic polynomial of T.

Theorem 6.0.15 — Degree and zeros of characteristic polynomial. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Then

1. the characteristic polynomial of T has degree $\dim V$
2. the zeros of the characteristic polynomial of T are the eigenvalues of T

Theorem 6.0.16 — Cayley-Hamilton Theorem. Suppose V is a complex vector space and $T \in \mathscr{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of the operator T, and let $d_1, \ldots, d_m$ be the dimensions of the corresponding generalized eigenspaces $G(\lambda_1, T), \ldots, G(\lambda_m, T)$. For each $j \in \{1, \ldots, m\}$, we know that $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus we have $(T - \lambda_j I)^{d_j}\Big|_{G(\lambda_j, T)} = 0$ Every vector in V is a sum of vectors in $G(\lambda_1, T), \ldots, G(\lambda_m, T)$ (by 8.21). Thus to prove that $q(T) = 0$, we need only show that $q(T)|_{G(\lambda_j, T)} = 0$ for each j Thus fix $j \in \{1, \ldots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}$$

The operators on the right side of the equation above all commute, so we can move the factor $(T - \lambda_j I)^{d_j}$ to be the last term in the expression on the right. Because $(T - \lambda_j I)^{d_j}\Big|_{G(\lambda_j, T)} = 0$, we conclude that $q(T)|_{G(\lambda_j, T)} = 0$, as desired. ■

Definition 6.0.9 — monic polynomial. A monic polynomial is a polynomial whose highest-degree coefficient equals 1.

Theorem 6.0.17 — Minimal polynomial. Suppose $T \in \mathscr{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$

Proof. Let $n = \dim V$. Then the list

$$I, T, T^2, \ldots, T^{n^2}$$

is not linearly independent in $\mathscr{L}(V)$, because the vector space $\mathscr{L}(V)$ has dimension n^2 and we have a list of length $n^2 + 1$. Let m be the smallest positive integer such that the list

$$I, T, T^2, \ldots, T^m$$

is linearly dependent. The Linear Dependence Lemma implies that one of the operators in the list above is a linear combination of the previous ones. Because m was chosen to be the smallest positive integer such that the list above is linearly dependent, we conclude that T^m is a linear combination of $I, T, T^2, \ldots, T^{m-1}$. Thus there exist scalars $a_0, a_1, a_2, \ldots, a_{m-1} \in \mathbf{F}$ such that

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0 \tag{6.5}$$

Define a monic polynomial $p \in \mathscr{P}(\mathbf{F})$ by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

Then 6.5 implies that $p(T) = 0$ To prove the uniqueness part of the result, note that the choice of m implies that no monic polynomial $q \in \mathscr{P}(\mathbf{F})$ with degree smaller than m can satisfy $q(T) = 0$. Suppose $q \in \mathscr{P}(\mathbf{F})$ is a monic polynomial with degree m and $q(T) = 0$. Then $(p-q)(T) = 0$ and $\deg(p-q) < m$. The choice of m now implies that $q = p$, completing the proof. ■

Definition 6.0.10 — minimal polynomial. Suppose $T \in \mathscr{L}(V)$. Then the minimal polynomial of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

R The proof of the last result shows that the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$. The Cayley-Hamilton Theorem (8.37) tells us that if V is a complex vector space, then the minimal polynomial of each operator on V has degree at most $\dim V$.

Theorem 6.0.18 — $q(T) = 0$ implies q is a multiple of the minimal polynomial. Suppose $T \in \mathscr{L}(V)$ and $q \in \mathscr{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T

Proof. Let p denote the minimal polynomial of T To prove the difficult direction, now suppose $q(T) = 0$.

$$0 = q(T) = p(T)s(T) + r(T) = r(T)$$

$\deg r < \deg p$. The equation above implies that $r = 0$ (otherwise, dividing r by its highest degree coefficient would produce a monic polynomial that when applied to T gives 0; this polynomial would have a smaller degree than the minimal polynomial, which would be a contradiction). Thus **??**becomes the equation $q = ps$. Hence q is a polynomial multiple of p, as desired.

■

Theorem 6.0.19 — Characteristic polynomial is a multiple of minimal polynomial. Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathscr{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T

Theorem 6.0.20 — Eigenvalues are the zeros of the minimal polynomial. Let $T \in \mathscr{L}(V)$. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T

Proof. Let

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

be the minimal polynomial of T First suppose $\lambda \in \mathbf{F}$ is a zero of p. Then p can be written in the form

$$p(z) = (z - \lambda) q(z)$$

where q is a monic polynomial with coefficients in $\mathbf{F}$ (see 4.11). Because $p(T) = 0$, we have

$$0 = (T - \lambda I)(q(T)v)$$

for all $v \in V$. Because the degree of q is less than the degree of the minimal polynomial p, there exists at least one vector $v \in V$ such that $q(T)v \neq 0$ .The equation above thus implies that λ is an eigenvalue of T, as desired. To prove the other direction, now suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T Thus there exists $v \in V$ with $v \neq 0$ such that $Tv = \lambda v$. Repeated applications of T to both sides of this equation show that $T^j v = \lambda^j v$ for every nonnegative integer j. Thus

$$\begin{aligned} 0 = p(T)v &= \left(a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m\right) v \\ &= \left(a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m\right) v \\ &= p(\lambda) v \end{aligned}$$

Because $v \neq 0$, the equation above implies that $p(\lambda) = 0$, as desired. ■

6.1 Jordan Form

Theorem 6.1.1 — Basis corresponding to a nilpotent operator. Suppose $N \in \mathscr{L}(V)$ is nilpotent. Then there exist vectors $v_1, \ldots, v_n \in V$ and nonnegative integers $m_1, \ldots, m_n$ such that

1. $N^{m_1} v_1, \ldots, N v_1, v_1, \ldots, N^{m_n} v_n, \ldots, N v_n, v_n$ is a basis of V
2. $N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0$

Proof. We will prove this result by induction on dim V. To get started, note that the desired result obviously holds if $\dim V = 1$ (in that case, the only nilpotent operator is the 0 operator, so take v_1 to be any nonzero vector and $m_1 = 0$). Now assume that $\dim V > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because N is nilpotent, N is not injective.(Why?) Thus N is not surjective and hence range N is a subspace of V that has a smaller dimension than V. Thus we can apply our induction hypothesis to the restriction operator $N|_{\text{range } N} \in \mathscr{L}(\text{ range } N)$. [We can ignore the trivial case range $N = \{0\}$ because in that case N is the 0 operator and we can choose $v_1, \ldots, v_n$ to be any basis of V and $m_1 = \cdots = m_n = 0$ to get the desired result.] By our induction hypothesis applied to $N|_{\text{range } N}$, there exist vectors $v_1, \ldots, v_n \in$ range N and nonnegative integers $m_1, \ldots, m_n$ such that

$$\ldots, N v_1, v_1, \ldots, N^{m_n} v_n, \ldots, N v_n, v_n \tag{6.6}$$

is a basis of range N and

$$N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$$

Because each v_j is in range N, for each j there exists $u_j \in V$ such that $v_j = Nu_j$. Thus $N^{k+1}u_j = N^k v_j$ for each j and each nonnegative integer k. We now claim that

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n \tag{6.7}$$

is a linearly independent list of vectors in V. To verify this claim, suppose that some linear combination of 6.7 equals 0. Applying N to that linear combination, we get a linear combination of 6.6 equal to 0. However, the list 6.6 is linearly independent, and hence all the coefficients in our original linear combination of 6.7 equal 0 except possibly the coefficients of the vectors

$$N^{m_1+1}u_1, \ldots, N^{m_n+1}u_n$$

which equal the vectors

$$N^{m_1}v_1, \ldots, N^{m_n}v_n$$

Again using the linear independence of the list 6.6, we conclude that those coefficients also equal 0, completing our proof that the list 6.7 is linearly independent. Now extend 6.7 to a basis

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n, w_1, \ldots, w_p \tag{6.8}$$

of V. Each Nw_j is in range N and hence is in the span of 6.6(basis) . Each vector in the list 6.6 equals N applied to one vector in the list 6.7. Thus there exists x_j in the span of 6.7 such that $Nw_j = Nx_j$ Now let

$$u_{n+j} = w_j - x_j$$

Then $Nu_{n+j} = 0$. Furthermore,

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n, u_{n+1}, \ldots, u_{n+p}$$

spans V because its span contains each x_j(in the span of 6.7) and each u_{n+j} and hence each w_j (and because 6.8 spans V).

Thus the spanning list above is a basis of V because it has the same length as the basis 6.8 . This basis has the required form, completing the proof. ■

Definition 6.1.1 — Jordan basis. Suppose $T \in \mathscr{L}(V)$. A basis of V is called a Jordan basis for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix}$$

where each A_j is an upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}$$

Theorem 6.1.2 — Jordan Form. Suppose V is a complex vector space. If $T \in \mathscr{L}(V)$, then there is a basis of V that is a Jordan basis for T.

Proof. First consider a nilpotent operator $N \in \mathscr{L}(V)$ and the vectors $v_1, \dots, v_n \in V$ given by 6.1.1. For each j, note that N sends the first vector in the list $N^{mj}v_j, \dots, Nv_j, v_j$ to 0 and that N sends each vector in this list other than the first vector to the previous vector(note this order). In other words, 6.1.1 gives a basis of V with respect to which N has a block diagonal matrix, where each matrix on the diagonal has the form

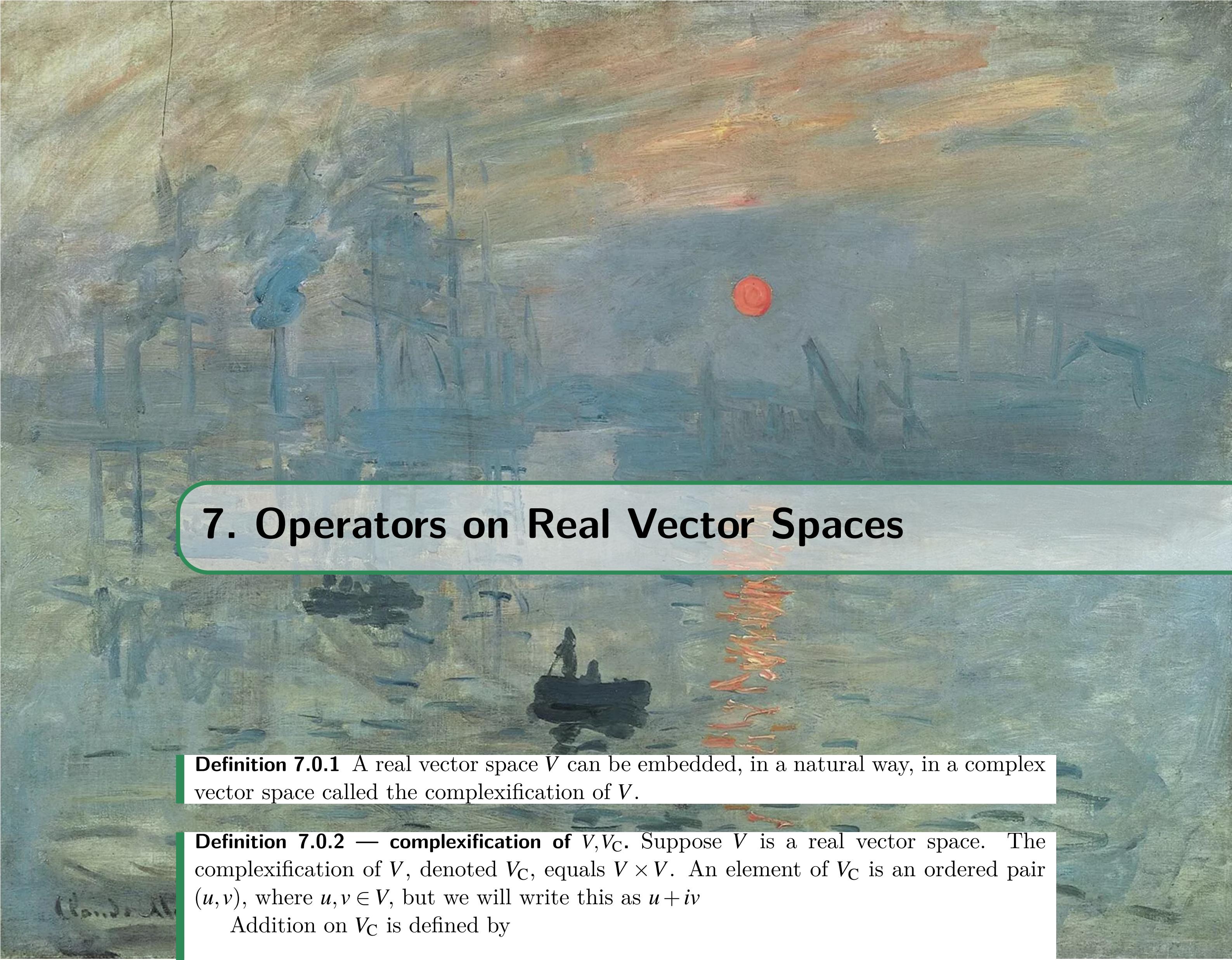
$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

Thus the desired result holds for nilpotent operators. Now suppose $T \in \mathscr{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T We have the generalized eigenspace decomposition

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

where each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent (see 8.21). Thus some basis of each $G(\lambda_j, T)$ is a Jordan basis for $(T - \lambda_j I)|_{G(\lambda_m, T)}$ (see previous paragraph). Put these bases together to get a basis of V that is a Jordan basis for T.

■



7. Operators on Real Vector Spaces

Definition 7.0.1 A real vector space V can be embedded, in a natural way, in a complex vector space called the complexification of V.

Definition 7.0.2 — complexification of $V, V_{\mathbf{C}}$. Suppose V is a real vector space. The complexification of V, denoted $V_{\mathbf{C}}$, equals $V \times V$. An element of $V_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we will write this as $u + iv$

Addition on $V_{\mathbf{C}}$ is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

for $u_1, v_1, u_2, v_2 \in V$

Complex scalar multiplication on $V_{\mathbf{C}}$ is defined by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbf{R}$ and $u, v \in V$

Theorem 7.0.1 — $V_{\mathbf{C}}$ is a complex vector space.. Suppose V is a real vector space. Then with the definitions of addition and scalar multiplication as above, $V_{\mathbf{C}}$ is a complex vector space.

Theorem 7.0.2 — Basis of V is basis of $V_{\mathbf{C}}$. Suppose V is a real vector space.

1. If $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_{\mathbf{C}}$ (as a complex vector space).
2. The dimension of $V_{\mathbf{C}}$ (as a complex vector space) equals the dimension of V (as a real vector space).

Proof. To prove (a), suppose $v_1, \ldots, v_n$ is a basis of the real vector space V Then $\operatorname{span}(v_1, \ldots, v_n)$ in the complex vector space $V_{\mathbf{C}}$ contains all the vectors $v_1, \ldots, v_n, iv_1, \ldots, iv_n$. Thus $v_1, \ldots, v_n$

spans the complex vector space V_{C} To show that $v_1, \ldots, v_n$ is linearly independent in the complex vector space $V_{\mathbf{C}}$, suppose $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$ and

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$$

Then the equation above and our definitions imply that $(\operatorname{Re} \lambda_1) v_1 + \cdots + (\operatorname{Re} \lambda_n) v_n = 0$ and $(\operatorname{Im} \lambda_1) v_1 + \cdots + (\operatorname{Im} \lambda_n) v_n = 0$ Because $v_1, \ldots, v_n$ is linearly independent in V, the equations above imply $\operatorname{Re} \lambda_1 = \cdots = \operatorname{Re} \lambda_n = 0$ and $\operatorname{Im} \lambda_1 = \cdots = \operatorname{Im}_n = 0$. Thus we have $\lambda_1 = \cdots = \lambda_n = 0$. Hence $v_1, \ldots, v_n$ is linearly independent in V_{C} completing the proof of (a). Clearly (b) follows immediately from (a). ■

Definition 7.0.3 — complexification of T, T_{C}. Suppose V is a real vector space and $T \in \mathscr{L}(V)$. The complexification of T, denoted $T_{\mathbf{C}}$, is the operator $T_{\mathbf{C}} \in \mathscr{L}(V_{\mathbf{C}})$ defined by

$$T_{\mathrm{C}}(u + iv) = Tu + iTv$$

for $u, v \in V$

R $T_{\mathrm{C}}(\lambda(u + iv)) = \lambda T_{\mathrm{C}}(u + iv)$ for all $u, v \in V$ and all complex numbers λ.

Theorem 7.0.3 — Matrix of T_{C} equals matrix of T. Suppose V is a real vector space with basis $v_1, \ldots, v_n$ and $T \in \mathscr{L}(V)$ Then $\mathscr{M}(T) = \mathscr{M}(T_{\mathrm{C}})$, where both matrices are with respect to the basis $v_1, \ldots, v_n$

Theorem 7.0.4 — Every operator has an invariant subspace of dimension 1 or 2. Every operator on a nonzero finite-dimensional vector space has an invariant subspace of dimension 1 or 2 .

Proof. Every operator on a nonzero finite-dimensional complex vector space has an eigenvalue (5.21) and thus has a 1 -dimensional invariant subspace. Hence assume V is a real vector space and $T \in \mathscr{L}(V)$. The complexification T_{C} has an eigenvalue $a + bi$ (by 5.21), where $a, b \in \mathbf{R}$. Thus there exist $u, v \in V$, not both 0, such that $T_{\mathrm{C}}(u + iv) = (a + bi)(u + iv)$. Using the definition of T_{C}, the last equation can be rewritten as

$$Tu + iTv = (au - bv) + (av + bu)i$$

Thus $Tu = au - bv$ and $Tv = av + bu$ Let U equal the span in V of the list u, v. Then U is a subspace of V with dimension 1 or 2. The equations above show that U is invariant under T completing the proof. ■

R Suppose V is a real vector space and $T \in \mathscr{L}(V)$. Repeated application of the definition of T_{C} shows that

$$(T_{\mathrm{C}})^n (u + iv) = T^n u + iT^n v \tag{7.1}$$

for every positive integer n and all $u, v \in V$

Theorem 7.0.5 — Minimal polynomial of $T_{\mathbf{C}}$ equals minimal polynomial of T. Suppose V is a real vector space and $T \in \mathscr{L}(V)$. Then the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T

Proof. Let $p \in \mathscr{P}(\mathbf{R})$ denote the minimal polynomial of T. From 9.9 it is easy to see that $p(T_{\mathbf{C}}) = (p(T))_{\mathbf{C}}$, and thus $p(T_{\mathbf{C}}) = 0$ Suppose $q \in \mathscr{P}(\mathbf{C})$ is a monic polynomial such that $q(T_{\mathbf{C}}) = 0$. Then $(q(T_{\mathbf{C}}))(u) = 0$ for every $u \in V$. Letting r denote the polynomial whose j^{th} coefficient is the real part of the j^{th} coefficient of q, we see that r is a monic polynomial and $r(T) = 0$. Thus $\deg q = \deg r \geq \deg p$ The conclusions of the two previous paragraphs imply that p is the minimal polynomial of $T_{\mathbf{C}}$, as desired.

■

Theorem 7.0.6 — Real eigenvalues of $T_{\mathbf{C}}$. Suppose V is a real vector space, $T \in \mathscr{L}(V)$, and $\lambda \in \mathbf{R}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if λ is an eigenvalue of T

Proof.

$$T_{\mathbf{C}}(u+iv) = \lambda(u+iv), Tu = \lambda u \text{ and } Tv = \lambda v$$

■

Theorem 7.0.7 — $T_{\mathbf{C}} - \lambda I$ and $T_{\mathbf{C}} - \bar{\lambda} I$. Suppose V is a real vector space, $T \in \mathscr{L}(V), \lambda \in \mathbf{C}, j$ is a nonnegative integer, and $u, v \in V$. Then $(T_{\mathbf{C}} - \lambda I)^j (u+iv) = 0$ if and only if $\left(T_{\mathbf{C}} - \bar{\lambda} I\right)^j (u-iv) = 0$

Proof. We will prove this result by induction on j. To get started, note that if $j = 0$ then (because an operator raised to the power 0 equals the identity operator) the result claims that $u + iv = 0$ if and only if $u - iv = 0$, which is clearly true.

Thus assume by induction that $j \geq 1$ and the desired result holds for $j-1$ Suppose $(T_{\mathbf{C}} - \lambda I)^j (u+iv) = 0$. Then

$$(T_{\mathbf{C}} - \lambda I)^{j-1} ((T_{\mathbf{C}} - \lambda I)(u+iv)) = 0 \tag{7.2}$$

Writing $\lambda = a + bi$, where $a, b \in \mathbf{R}$, we have

$$(T_{\mathbf{C}} - \lambda I)(u+iv) = (Tu - au + bv) + i(Tv - av - bu) \tag{7.3}$$

and

$$\left(T_{\mathbf{C}} - \bar{\lambda} I\right)(u-iv) = (Tu - au + bv) - i(Tv - av - bu) \tag{7.4}$$

Our induction hypothesis, 7.2 and 7.3 imply that

$$\left(T_{\mathbf{C}} - \bar{\lambda} I\right)^{j-1} ((Tu - au + bv) - i(Tv - av - bu)) = 0$$

Now the equation above and 7.4 imply that $\left(T_{\mathbf{C}} - \bar{\lambda} I\right)^j (u-iv) = 0$ completing the proof in one direction.

The other direction is proved by replacing λ with $\bar{\lambda}$, replacing v with $-v$ and then using the first direction.

■

Theorem 7.0.8 — Nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs. Suppose V is a real vector space, $T \in \mathscr{L}(V)$, and $\lambda \in \mathbf{C}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$

Theorem 7.0.9 — Multiplicity of λequals multiplicity of $\bar{\lambda}$. Suppose V is a real vector space, $T \in \mathscr{L}(V)$, and $\lambda \in \mathbf{C}$ is an eigenvalue of $T_{\mathbf{C}}$. Then the multiplicity of λ as an eigenvalue of $T_{\mathbf{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbf{C}}$

Proof. Suppose $u_1 + iv_1, \ldots, u_m + iv_m$ is a basis of the generalized eigenspace $G(\lambda, T_{\mathbf{C}})$, where $u_1, \ldots, u_m, v_1, \ldots, v_m \in V$. Then using7.0.7 routine arguments show that $u_1 - iv_1, \ldots, u_m - iv_m$ is a basis of the generalized eigenspace $G(\lambda, T_{\mathbf{C}})$. Thus both λ and λ have multiplicity m as eigenvalues of $T_{\mathbf{C}}$ ■

Theorem 7.0.10 — Operator on odd-dimensional vector space has eigenvalue. Every operator on an odd-dimensional real vector space has an eigenvalue.

Proof. Proof Suppose V is a real vector space with odd dimension and $T \in \mathscr{L}(V)$ Because the nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs with equal multiplicity , the sum of the multiplicities of all the nonreal eigenvalues of $T_{\mathbf{C}}$ is an even number.

Because the sum of the multiplicities of all the eigenvalues of $T_{\mathbf{C}}$ equals the (complex) dimension of $V_{\mathbf{C}}$ (by Theorem 8.26), the conclusion of the paragraph above implies that $T_{\mathbf{C}}$ has a real eigenvalue. Every real eigenvalue of $T_{\mathbf{C}}$ is also an eigenvalue of T (by 9.11), giving the desired result.

■

Theorem 7.0.11 — Characteristic polynomial of $T_{\mathbf{C}}$. Suppose V is a real vector space and $T \in \mathscr{L}(V)$. Then the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real.

Proof. Suppose λ is a non-real eigenvalue of $T_{\mathbf{C}}$ with multiplicity m. Then $\bar{\lambda}$ is also an eigenvalue of $T_{\mathbf{C}}$ with multiplicity m . Thus the characteristic polynomial of $T_{\mathbf{C}}$ includes factors of $(z-\lambda)^m$ and $(z-\bar{\lambda})^m$. Multiplying together these two factors, we have

$$(z-\lambda)^m (z-\bar{\lambda})^m = \left(z^2 - 2(\operatorname{Re}\lambda)z + |\lambda|^2\right)^m$$

The polynomial above on the right has real coefficients. The characteristic polynomial of $T_{\mathbf{C}}$ is the product of terms of the form above and terms of the form $(z-t)^d$, where t is a real eigenvalue of $T_{\mathbf{C}}$ with multiplicity d . Thus the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real. ■

Definition 7.0.4 — Characteristic polynomial. Suppose V is a real vector space and $T \in \mathscr{L}(V)$. Then the characteristic polynomial of T is defined to be the characteristic polynomial of $T_{\mathbf{C}}$

Theorem 7.0.12 — Degree and zeros of characteristic polynomial. Suppose V is a real vector space and $T \in \mathscr{L}(V)$. Then

1. the coefficients of the characteristic polynomial of T are all real;
2. the characteristic polynomial of T has degree $\dim V$

3. the eigenvalues of T are precisely the real zeros of the characteristic polynomial of T

Theorem 7.0.13 — Cayley-Hamilton Theorem. Suppose $T \in \mathscr{L}(V)$. Let q denote the characteristic polynomial of T Then $q(T)=0$ The complex case of the Cayley-Hamilton Theorem (8.37) implies that $q\left(T_{\mathrm{C}}\right)=0$. Thus we also have $q(T)=0$, as desired.

Theorem 7.0.14 — Characteristic polynomial is a multiple of minimal polynomial. Suppose $T \in \mathscr{L}(V)$. Then
1. the degree of the minimal polynomial of T is at most $\operatorname{dim} V$
2. the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T

7.1 Operators on Real Inner Product Spaces

Theorem 7.1.1 — Normal but not self-adjoint operators. Suppose V is a 2 -dimensional real inner product space and $T \in \mathscr{L}(V)$ Then the following are equivalent:
1. T is normal but not self-adjoint.
2. The matrix of T with respect to every orthonormal basis of V has the form
$$\left(\begin{array}{cc} a & -b \\ b & a \end{array}\right)$$
with $b \neq 0$
3. The matrix of T with respect to some orthonormal basis of V has the form
$$\left(\begin{array}{cc} a & -b \\ b & a \end{array}\right)$$
with $b>0$

Proof. Recall normal and self-adjoint's defs and props. First suppose (a) holds, so that T is normal but not self-adjoint. Let e_1, e_2 be an orthonormal basis of V. Suppose

$$\mathscr{M}\left(T,\left(e_1, e_2\right)\right)=\left(\begin{array}{ll} a & c \\ b & d \end{array}\right) \tag{7.5}$$

Then $\left\|T e_1\right\|^2=\left\|a e_1+b e_2\right\|^2=a^2+b^2$ and $\left\|T^* e_1\right\|^2=a^2+c^2$. Because T is normal, $\left\|T e_1\right\|= \left\|T^* e_1\right\|$; thus these equations imply that $b^2=c^2$. Thus $c=b$ or $c=-b$. But $c \neq b$, because otherwise T would be self-adjoint, as can be seen from the matrix in 9.28. Hence $c=-b$, so 9.29 $\quad \mathscr{M}\left(T,\left(e_1, e_2\right)\right)=\left(\begin{array}{cc} a & -b \\ b & d \end{array}\right)$ The matrix of T^* is the transpose of the matrix above. Use matrix multiplication to compute the matrices of $T T^*$ and $T^* T$ (do it now). Because T is normal, these two matrices are equal. Equating the entries in the upper-right corner of the two matrices you computed, you will discover that $b d=a b$ Now $b \neq 0$, because otherwise T would be self-adjoint, as can be seen from the matrix in 9.29. Thus $d=a$, completing the proof that (a) implies (b). Now suppose (b) holds. We want to prove that (c) holds. Choose an orthonormal basis e_1, e_2 of V. We know that the matrix of T with respect to this basis has the form given by (b), with $b \neq 0$. If $b>0$, then (c) holds and we have proved that (b) implies (c). If $b<0$, then, as you should verify, the matrix of

T with respect to the orthonormal basis $e_1, -e_2$ equals $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ where $-b > 0$; thus in this case we also see that (b) implies (c).

Now suppose (c) holds, so that the matrix of T with respect to some orthonormal basis has the form given in (c) with $b > 0$. Clearly the matrix of T is not equal to its transpose (because $b \neq 0$). Hence T is not self-adjoint. Now use matrix multiplication to verify that the matrices of TT^* and T^*T are equal. We conclude that $TT^* = T^*T$. Hence T is normal. Thus (c) implies (a), completing the proof.

■

Theorem 7.1.2 — Normal operators and invariant subspaces. Suppose V is an inner product space, $T \in \mathscr{L}(V)$ is normal, and U is a subspace of V that is invariant under T. Then

1. $U^{\perp}$ is invariant under T
2. U is invariant under T^*
3. $(T|_U)^* = (T^*)|_U$
4. $T|_U \in \mathscr{L}(U)$ and $T|_{U^\perp} \in \mathscr{L}\left(U^{\perp}\right)$ are normal operators.

Proof. Proof First we will prove (a). Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Extend to an orthonormal basis $e_1, \ldots, e_m, f_1, \ldots, f_n$ of V. Because U is invariant under T, each Te_j is a linear combination of $e_1, \ldots, e_m$. Thus the matrix of T with respect to the basis $e_1, \ldots, e_m, f_1, \ldots, f_n$ is of the form

$$\mathscr{M}(T) = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

here A denotes an m -by- m matrix, 0 denotes the n -by- m matrix of all 0 's, B denotes an $m-$ by- n matrix, C denotes an n -by- n matrix, and for convenience the basis has been listed along the top and left sides of the matrix.

For each $j \in \{1, \ldots, m\}$, $\left\|Te_j\right\|^2$ equals the sum of the squares of the absolute values of the entries in the j^{th} column of A. Hence

$$\sum_{j=1}^{m} \left\|Te_j\right\|^2 = \begin{array}{l}\text{the sum of the squares of the absolute} \\ \text{values of the entries of } A\end{array} \tag{7.6}$$

For each $j \in \{1, \ldots, m\}$, $\left\|T^*e_j\right\|^2$ equals the sum of the squares of the absolute values of the entries in the j^{th} rows of A and B. Hence

$$\sum_{j=1}^{m} \left\|T^*e_j\right\|^2 = \begin{array}{l}\text{the sum of the squares of the absolute} \\ \text{values of the entries of } A \text{ and } B.\end{array} \tag{7.7}$$

Because T is normal, $\left\|Te_j\right\| = \left\|T^*e_j\right\|$ for each j; thus

$$\sum_{j=1}^{m} \left\|Te_j\right\|^2 = \sum_{j=1}^{m} \left\|T^*e_j\right\|^2$$

This equation, along with 7.6 and 7.7 , implies that the sum of the squares of the absolute values of the entries of B equals 0. In other words, B is the matrix of all 0's. Thus

$$\mathscr{M}(T) = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}$$

This representation shows that Tf_k is in the span of $f_1,\ldots,f_n$ for each k Because $f_1,\ldots,f_n$ is a basis of $U^\perp$, this implies that $Tv \in U^\perp$ whenever $v \in U^\perp$. In other words, $U^\perp$ is invariant under T, completing the proof of (a). To prove (b), note that $\mathscr{M}(T^*)$, which is the conjugate transpose of $\mathscr{M}(T)$, has a block of 0 's in the lower left corner (because $\mathscr{M}(T)$, as given above, has a block of 0 's in the upper right corner). In other words, each T^*e_j can be written as a linear combination of $e_1,\ldots,e_m$. Thus U is invariant under T^* completing the proof of (b). To prove (c), let $S = T|_U \in \mathscr{L}(U)$. Fix $v \in U$. Then

$$\begin{aligned}\langle Su, v\rangle &= \langle Tu, v\rangle \\ &= \langle u, T^*v\rangle\end{aligned}$$

for all $u \in U$. Because $T^*v \in U$[by (b)], the equation above shows that $S^*v = T^*v$. In other words, $(T|_U)^* = (T^*)|_U$, completing the proof of (c). To prove (d), note that T commutes with T^* (because T is normal) and that $(T|_U)^* = (T^*)|_U$ [by (c)]. Thus $T|_U$ commutes with its adjoint and hence is normal. Interchanging the roles of U and $U^\perp$, which is justified by (a), shows that $T|_U \perp$ is also normal, completing the proof of (d). ■

R Note that if an operator T has a block diagonal matrix with respect to some basis, then the entry in each $1-by-1$ block on the diagonal of this matrix is an eigenvalue of T

R a normal operator restricted to an invariant subspace is normal

Theorem 7.1.3 — Characterization of normal operators when $\mathbf{F} = \mathbf{R}$. Suppose V is a real inner product space and $T \in \mathscr{L}(V)$. Then the following are equivalent:
1. T is normal.
2. There is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is a 1 -by- 1 matrix or a 2-by-2 matrix of the form with $b > 0$

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Proof. First suppose (b) holds. With respect to the basis given by (b), the matrix of T commutes with the matrix of T^* (which is the transpose of the matrix of T), as you should verify (use Exercise 9 in Section 8. B for the product of two block diagonal matrices). Thus T commutes with T^*, which means that T is normal, completing the proof that (b) implies (a).

Now suppose (a) holds, so T is normal. We will prove that (b) holds by induction on $\dim V$. To get started, note that our desired result holds if $\dim V = 1$ (trivially) or if $\dim V = 2$ [if T is self-adjoint, use the Real Spectral Theorem 5.0.16 ; if T is not self-adjoint, use 7.1.1].

Now assume that $\dim V > 2$ and that the desired result holds on vector spaces of smaller dimension. Let U be a subspace of V of dimension 1 that is invariant under T if such a subspace exists (in other words, if T has an eigenvector, let U be the span of this eigenvector). If no such subspace exists, let U be a subspace of V of dimension 2 that is invariant under T (an invariant subspace of dimension 1 or 2 always exists).

If $\dim U = 1$, choose a vector in U with norm 1; this vector will be an orthonormal basis of U, and of course the matrix of $T|_U \in \mathscr{L}(U)$ is a 1-by-1 matrix. If $\dim U = 2$, then $T|_U \in \mathscr{L}(U)$ is normal 7.1.2 but not self-adjoint (otherwise $T|_U$, and hence T, would have an eigenvector by 5.0.14). Thus we can choose an orthonormal basis of U with respect to which the matrix of $T|_U \in \mathscr{L}(U)$ has the required form.

Now $U^\perp$ is invariant under T and $T|_{U\perp}$ is a normal operator on $U^\perp$ 7.1.2. Thus by our induction hypothesis, there is an orthonormal basis of $U^\perp$ with respect to which the matrix of $T|_{U\perp}$ has the desired form. Adjoining this basis to the basis of U gives an orthonormal basis of V with respect to which the matrix of T has the desired form. Thus (b) holds. ■

Theorem 7.1.4 — Description of isometries when $\mathbf{F} = \mathbf{R}$. Suppose V is a real inner product space and $S \in \mathscr{L}(V)$. Then the following are equivalent:

1. S is an isometry.
2. There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or is a 2-by-2 matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $\theta \in (0, \pi)$

Proof. First suppose (a) holds, so S is an isometry. Because S is normal, there is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block is a 1 -by- 1 matrix or a 2 -by- 2 matrix of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{7.8}$$

with $b > 0$. If λ is an entry in a 1 -by- 1 matrix along the diagonal of the matrix of S (with respect to the basis mentioned above), then there is a basis vector e_j such that $Se_j = \lambda e_j$. Because S is an isometry, this implies that $|\lambda| = 1$. Thus $\lambda = 1$ or $\lambda = -1$, because these are the only real numbers with absolute value 1.

Now consider a 2 -by- 2 matrix of the form 9.37 along the diagonal of the matrix of S. There are basis vectors e_j, e_{j+1} such that

$$Se_j = ae_j + be_{j+1}$$

Thus

$$1 = \|e_j\|^2 = \|Se_j\|^2 = a^2 + b^2$$

The equation above, along with the condition $b > 0$, implies that there exists a number $\theta \in (0, \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$. Thus the matrix 9.37 has the required form, completing the proof in this direction.

Conversely, now suppose (b) holds, so there is an orthonomal basis of V with respect to which the matrix of S has the form required by the theorem. Thus there is a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m$$

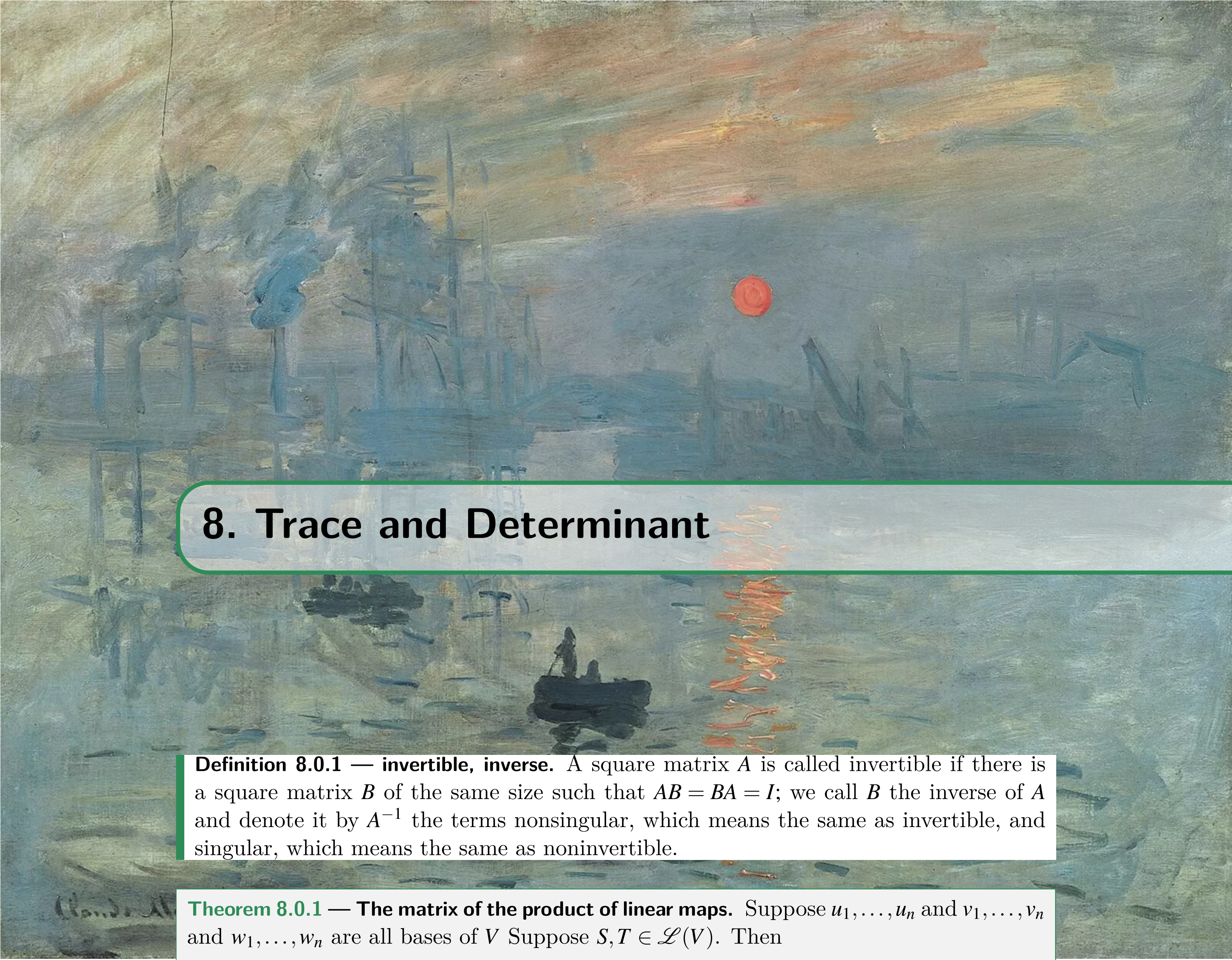
where each U_j is a subspace of V of dimension 1 or 2 . Furthermore, any two vectors belonging to distinct U 's are orthogonal, and each $S|v_j$ is an isometry mapping U_j into U_j. If $v \in V$, we can write

$$v = u_1 + \cdots + u_m$$

where each u_j is in U_j. Applying S to the equation above and then taking norms gives

$$\begin{aligned}
\|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\
&= \|Su_1\|^2 + \cdots + \|Su_m\|^2 \\
&= \|u_1\|^2 + \cdots + \|u_m\|^2 \\
&= \|v\|^2
\end{aligned}$$

Thus S is an isometry, and hence (a) holds. ■



8. Trace and Determinant

Definition 8.0.1 — invertible, inverse. A square matrix A is called invertible if there is a square matrix B of the same size such that $AB = BA = I$; we call B the inverse of A and denote it by A^{-1} the terms nonsingular, which means the same as invertible, and singular, which means the same as noninvertible.

Theorem 8.0.1 — The matrix of the product of linear maps. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are all bases of V Suppose $S, T \in \mathscr{L}(V)$. Then

$$\mathscr{M}(ST, (u_1, \ldots, u_n), (w_1, \ldots, w_n)) =$$
$$\mathscr{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_n))\, \mathscr{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$$

Definition 8.0.2 $\mathscr{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$ consists of the scalars needed to write u_k as a linear combination of $v_1, \ldots v_n$

Theorem 8.0.2 — Matrix of the identity with respect to two bases. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then the matrices $\mathscr{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$ and $\mathscr{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n))$ are invertible, and each is the inverse of the other.

Theorem 8.0.3 — Change of basis formula. Suppose $T \in \mathscr{L}(V)$. Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be bases of V. Let $A = \mathscr{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Then

$$\mathscr{M}(T, (u_1, \ldots, u_n)) = A^{-1} \mathscr{M}(T, (v_1, \ldots, v_n)) A$$

Definition 8.0.3 — trace of an operator. Suppose $T \in \mathscr{L}(V)$ If $\mathbf{F} = \mathbf{C}$, then the trace of T is the sum of the eigenvalues of T with each eigenvalue repeated according to its multiplicity. If $\mathbf{F} = \mathbf{R}$, then the trace of T is the sum of the eigenvalues of $T_{\mathbf{C}}$ with each eigenvalue repeated according to its multiplicity. The trace of T is denoted by trace T.

Theorem 8.0.4 — Trace and characteristic polynomial. Suppose $T \in \mathscr{L}(V)$. Let $n = \dim V$. Then trace T equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T

Definition 8.0.4 — trace of a matrix . The trace of a square matrix A, denoted trace A, is defined to be the sum of the diagonal entries of A

R It is indeed true that trace $T = \operatorname{trace} \mathscr{M}(T, (v_1, \ldots, v_n))$ where $v_1, \ldots, v_n$ is an arbitrary basis of V.

Theorem 8.0.5 — Trace of AB equals trace of BA. If A and B are square matrices of the same size, then

$$\operatorname{trace}(AB) = \operatorname{trace}(BA)$$

R This theorem can be extended to non-square matrix.

Theorem 8.0.6 — Trace of matrix of operator does not depend on basis. Let $T \in \mathscr{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\operatorname{trace} \mathscr{M}(T, (u_1, \ldots, u_n)) = \operatorname{trace} \mathscr{M}(T, (v_1, \ldots, v_n))$$

Proof Let $A = \mathscr{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Then

$$\begin{aligned}
\operatorname{trace} \mathscr{M}(T, (u_1, \ldots, u_n)) &= \operatorname{trace}\left(A^{-1}(\mathscr{M}(T, (v_1, \ldots, v_n))A)\right) \\
&= \operatorname{trace}\left((\mathscr{M}(T, (v_1, \ldots, v_n))A)A^{-1}\right) \\
&= \operatorname{trace} \mathscr{M}(T, (v_1, \ldots, v_n))
\end{aligned}$$

Theorem 8.0.7 Trace of an operator equals trace of its matrix Suppose $T \in \mathscr{L}(V)$. Then trace $T = \operatorname{trace} \mathscr{M}(T)$

Theorem 8.0.8 — Trace is additive. Suppose $S, T \in \mathscr{L}(V)$. Then trace $(S+T) = \operatorname{trace} S +$ trace T

Theorem 8.0.9 — The identity is not the difference of ST and TS. There do not exist operators $S, T \in \mathscr{L}(V)$ such that $ST - TS = I$

Proof. Suppose $S, T \in \mathscr{L}(V)$. Choose a basis of V. Then

$$\begin{aligned}
\operatorname{trace}(ST - TS) &= \operatorname{trace}(ST) - \operatorname{trace}(TS) \\
&= \operatorname{trace} \mathscr{M}(ST) - \operatorname{trace} \mathscr{M}(TS) \\
&= \operatorname{trace}(\mathscr{M}(S)\mathscr{M}(T)) - \operatorname{trace}(\mathscr{M}(T)\mathscr{M}(S)) \\
&= 0
\end{aligned}$$

Clearly the trace of I equals $\dim V$, which is not 0. Because $ST - TS$ and I have different traces, they cannot be equal.

■

Definition 8.0.5 — determinant of an operator, det T. Suppose $T \in \mathscr{L}(V)$ If $\mathbf{F} = \mathbf{C}$, then the determinant of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity. If $\mathbf{F} = \mathbf{R}$, then the determinant of T is the product of the eigenvalues of T_{C}, with each eigenvalue repeated according to its multiplicity. The determinant of T is denoted by det T.

If $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or of T_{C} if V is a real vector space) with multiplicities $d_1, \dots, d_m$, then the definition above implies

$$\det T = \lambda_1^{d_1} \cdots \lambda_m^{d_m}$$

Theorem 8.0.10 — Determinant and characteristic polynomial. Suppose $T \in \mathscr{L}(V)$. Let $n = \dim V$. Then det T equals $(-1)^n$ times the constant term of the characteristic polynomial of T

Theorem 8.0.11 — Characteristic polynomial, trace, and determinant. Suppose $T \in \mathscr{L}(V)$. Then the characteristic polynomial of T can be Written as

$$z^n - (\operatorname{trace} T) z^{n-1} + \cdots + (-1)^n (\det T)$$

Theorem 8.0.12 — Invertible is equivalent to nonzero determinant. An operator on V is invertible if and only if its determinant is nonzero.

Theorem 8.0.13 — Characteristic polynomial of T equals $\det(zI - T)$. Suppose $T \in \mathscr{L}(V)$. Then the characteristic polynomial of T equals $\det(zI - T)$

Proof. First suppose V is a complex vector space. If $\lambda, z \in \mathbf{C}$, then λ is an eigenvalue of T if and only if $z - \lambda$ is an eigenvalue of $zI - T$, as can be seen from the equation

$$-(T - \lambda I) = (zI - T) - (z - \lambda) I$$

Raising both sides of this equation to the dim V power and then taking null spaces of both sides shows that the multiplicity of λ as an eigenvalue of T equals the multiplicity of $z - \lambda$ as an eigenvalue of $zI - T$

Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of T, repeated according to multiplicity. Thus for $z \in \mathbf{C}$, the paragraph above shows that the eigenvalues of $zI - T$ are $z - \lambda_1, \dots, z - \lambda_n$, repeated according to multiplicity. The determinant of $zI - T$ is the product of these eigenvalues. In other words,

$$\det(zI - T) = (z - \lambda_1) \cdots (z - \lambda_n)$$

The right side of the equation above is, by definition, the characteristic polynomial of T, completing the proof when V is a complex vector space.

Now suppose V is a real vector space. Applying the complex case to T_{C} gives the desired result.

■

Definition 8.0.6 — permutation, perm n . A permutation of $(1,\ldots,n)$ is a list $(m_1,\ldots,m_n)$ that contains each of the numbers $1,\ldots,n$ exactly once.

The set of all permutations of $(1,\ldots,n)$ is denoted perm n

Definition 8.0.7 — sign of a permutation. The sign of a permutation $(m_1,\ldots.,m_n)$ is defined to be 1 if the number of pairs of integers (j,k) with $1\leq j<k\leq n$ such that j appears after k in the list $(m_1,\ldots,m_n)$ is even and -1 if the number of such pairs is odd.

In other words, the sign of a permutation equals 1 if the natural order has been changed an even number of times and equals -1 if the natural order has been changed an odd number of times.

Suppose A is an n -by- n matrix

$$A=\begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \ldots & A_{n,n} \end{pmatrix}$$

The determinant of A, denoted det A, is defined by

$$\det A=\sum_{(m_1,\ldots,m_n)\in \text{ perm } n}(\operatorname{sign}(m_1,\ldots,m_n))A_{m_1,1}\cdots A_{m_n,n}$$

Theorem 8.0.14 — Isometries have determinant with absolute value 1. Suppose V is an inner product space and $S\in\mathscr{L}(V)$ is an isometry. Then $|\det S|=1$

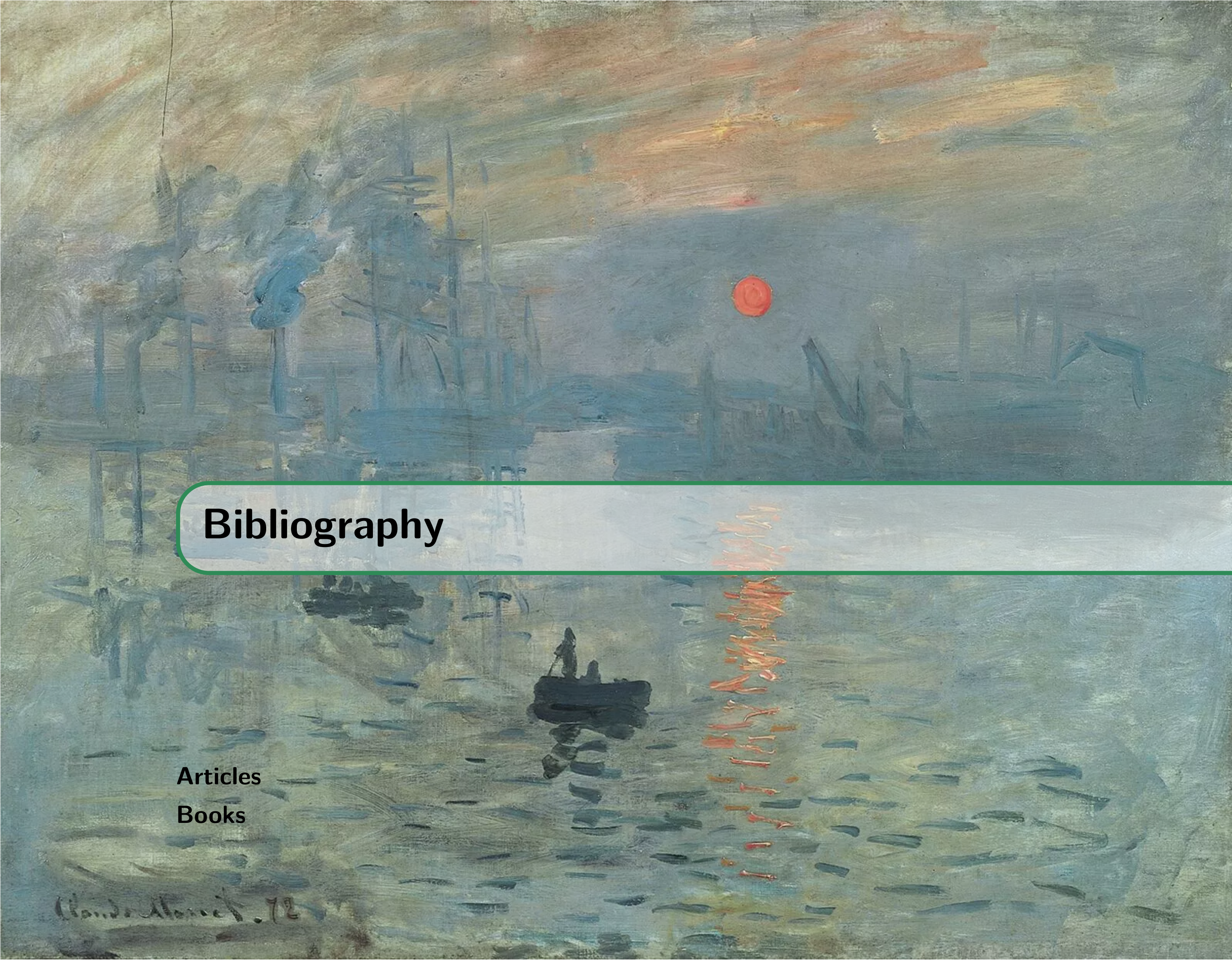
Proof. First consider the case where V is a complex inner product space. Then all the eigenvalues of S have absolute value 1 (see the proof of 7.43). Thus the product of the eigenvalues of S, counting multiplicity, has absolute value one. In other words, $|\det S|=1$, as desired.

Now suppose V is a real inner product space. We present two different proofs in this case.

By 9.36, there is an orthonormal basis of V with respect to which $\mathscr{M}(S)$ is a block diagonal matrix, where each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or a 2-by-2 matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $\theta\in(0,\pi)$. Note that the determinant of each 2 -by- 2 matrix of the form above equals 1 (because $\cos^2\theta+\sin^2\theta=1$). Thus the determinant of S which is the product of the determinants of the blocks, is the product of 1 's and -1 's. Hence, $|\det S|=1$, as desired. ■



Bibliography

Articles
Books