

MATH 110 Notes
based on *Linear Algebra Done Right*, third edition
Cheng, Feng

Contents

1	Vector Spaces	3
2	Finite-Dimensional Vector Spaces	4
3	Linear Maps	8
3.1	The Vector Space of Linear Maps	8
3.2	Null Spaces and Ranges	8
3.3	Matrices	10
3.4	Invertibility and Isomorphic Vector Spaces	11
3.5	Products and Quotients of Vector Spaces	13
3.6	Duality	15
4	Polynomials	19
5	Eigenvalues, Eigenvectors, and Invariant Subspaces	22
5.1	Invariant Subspaces	22
5.2	Eigenvectors and Upper-Triangular Matrices	23
5.3	Eigenspaces and Diagonal Matrices	25
6	Inner Product Spaces	28
6.1	Inner Products and Norms	28
6.2	Orthonormal Bases	30
6.3	Orthogonal Complements and Minimization Problems	33
7	Operators on Inner Product Spaces	36
7.1	Self-Adjoint and Normal Operators	36
7.2	The Spectral Theorem	39
7.3	Positive Operators and Isometries	42
7.4	Polar Decomposition and Singular Value Decomposition	44
8	Operators on Complex Vector Spaces	47
8.1	Generalized Eigenvectors and Nilpotent Operators	47
8.2	Decomposition of an Operator	49
8.3	Characteristic and Minimal Polynomials	52
8.4	Jordan Form	54

9	Operators on Real Vector Spaces	57
9.1	Complexification	57
9.2	Operators on Real Inner Product Spaces	61
10	Trace and Determinant	66
	Change of Basis	66
	Trace and Determinant of an Operator	67
	Trace of a Matrix	68
	Determinant of a Matrix	69
	Geometric Interpretation of the Determinant	72

1 Vector Spaces

- A *vector space* V over a field \mathbf{F}
 - (1) is an abelian group under addition;
 - (2) has the multiplicative identity; and
 - (3) has the distributive and associative properties in scalar multiplication.

(Addition and scalar multiplication must be closed in V .)
- \mathbf{F}^S denotes the set of functions from S to \mathbf{F} . It is a vector space over \mathbf{F} under the sum and scalar multiplication of functions defined in the usual pointwise way.
- A vector space holds the cancellation law and has a unique additive identity and a unique additive inverse because it is a group. $0v = 0$, $a0 = 0$, and $(-1)v = -v$ hold by the distributive property.
- A subset U of the whole vector space V that is still a vector space is called a vector subspace. Usually we use the three criteria i) $0_V \in U$ (or U is nonempty), ii) closed under addition in U , and iii) closed under scalar multiplication in U to verify. To show the criteria are equivalent to the definition, in one direction we show that 0_V is the (unique) identity element of U and in the other direction we only need to check the additive inverse property (satisfied by $(-1)u = -u$).
- The sum $U_1 + U_2 + \cdots + U_m$, where U_i 's are subsets of V , is the set of all possible sums of the respective elements from each U_i .
- When the U_i 's are subspaces of V , then the sum is the smallest vector space containing all U_i 's.
- When the U_i 's are subspaces of V , and every element of the sum $U_1 + U_2 + \cdots + U_m$ can be written in only one way as $u_1 + u_2 + \cdots + u_m$, we call the sum a *direct sum*, denoted by \oplus replacing $+$.
- The equivalent criterion for the sum to be a direct sum is that the only way to express 0 is to write it as the sum of 0 's from each U_i . One direction is trivial, while the other direction employs the common trick of subtracting two different expressions of the same vector $v \in V$.
- If we are only discussing two subspaces U and W of V , then $U + W$ is a direct sum iff their intersection is $\{0\}$. The proof of this, in both directions, considers $0 = v + (-v)$. The fact W always contains an element and its inverse will give us the answer.
 - Based on this, one can further show that subspaces U_1, U_2, \dots, U_n of V have pairwise intersections are all $\{0\}$ iff $U_1 + U_2 + \cdots + U_n$ is a direct sum.
- Arbitrary intersection of subspaces of V is a subspace of V , and the union of two subspaces is still a subspace iff one is contained in the other.

2 Finite-Dimensional Vector Spaces

Remark. Note that this section in the book builds on the idea of a “list” (ordered set) of vectors, which is no different than using a “set” of vectors because vector addition is commutative. Finite dimension is defined in terms of a spanning list, and the concept of a basis and the dimension of a vector space is pushed backward. In our exposition, we rearrange the progression of these ideas and reconstruct some proofs.

- The *span* of a list (set) S of vectors in V is defined to be all possible linear combinations of these vectors. In particular, the span of the empty list (set) is defined to be the zero vector space $\{0\}$.
- Similar to the fact that the sum $U_1 + U_2 + \cdots + U_m$ is the smallest vector space containing all U_i 's, $\text{span}(S)$ is the smallest subspace of V containing all the vectors in S . The proof of the two are almost exactly the same.
- A polynomial p in an indeterminate z with coefficients from a field \mathbf{F} is the (not unique) **expression** $a_0 + a_1z + \cdots + a_nz^n$ with all the a_i 's in \mathbf{F} . The *degree* of a polynomial is the largest exponent of z with nonzero coefficient, and if all the coefficients are 0, we call the polynomial the *zero polynomial*. We customarily define the degree of the zero polynomial as $-\infty$ (as in this book), -1 , or make it undefined.

P.S. The most general and rigorous definition of a polynomial over a ring is not discussed here, because our focus is vector spaces over a field.

- It is oftentimes useful to define a polynomial over a field \mathbf{F} rather as the **function** $p : \mathbf{F} \rightarrow \mathbf{F}$ with $p(z) = a_0 + a_1z + \cdots + a_nz^n$. The reason behind is that a polynomial function with coefficients from an infinite field can be written into only one polynomial expression with its unique coefficients. See Chapter 4 of this book or Appendix E, Theorem 10 of *Linear Algebra* by Friedberg, Insel, and Spence. It is the consequence of the fact that a polynomial of degree $n \geq 1$ has at most n distinct zeros. We will proceed with this definition of polynomials.
- We denote the set of all polynomials with coefficients in \mathbf{F} by $\mathcal{P}(\mathbf{F})$. Under the usual coefficient-wise addition and scalar multiplication, it is obvious that $\mathcal{P}(\mathbf{F})$ is a vector space over \mathbf{F} . In addition, the set of all polynomials with degree at most m , denoted by $\mathcal{P}_m(\mathbf{F})$ is a vector subspace of $\mathcal{P}(\mathbf{F})$.
- Linear independence and dependence can be characterized in multiple ways. A list (set) of vectors is linearly independent iff the only way to write 0 as $a_1v_1 + a_2v_2 + \cdots + a_mv_m$ is the trivial way. (We also sometimes say that some vectors themselves are linearly independent.) It is equivalent to saying that every linear combination has its only representation, or none of the vectors is in the span of the other vectors (can be written as a linear combination of them). Linear dependence is the negation of independence, but in particular we should remember that it means there exists at least one vector v_i that is in the span of the rest of the vectors.

- Obviously removing such vector v_i from S does not change the span.
- In particular, for a linearly independent list s_1, \dots, s_m properly extended to a linearly dependent list $s_1, s_2, \dots, s_m, t_1, t_2, \dots, t_n$, one such vector is always in t_1, t_2, \dots, t_n . This is because for the nontrivial way of writing $0 = a_1 s_1 + \dots + a_m s_m + b_1 t_1 + \dots + b_n t_n$, if all b 's are 0, then all a 's will also be 0. Thus, some b_i must be nonzero, and the corresponding t_i can be removed without changing the span.
 - * The special case $n = 1$ of the contraposition will be useful to us. If one new vector is added to the linearly independent list of s 's but does not belong to its span, then the resulting set is still linearly independent.
 - * From this special case we know that for a linearly dependent list v_1, \dots, v_n , at some index $k \leq n$ we have v_1, \dots, v_{k-1} being linearly independent, while $v_k \in \text{span}(v_1, \dots, v_{k-1})$. (This is the well-ordering case of the linear dependence lemma 2.21 in book that we will use again in Chapter 5 and Chapter 8.)
- Note that a linearly independent list of vectors cannot have the 0 vector.
- We now have the essential tools to prove the following important theorem, that if a vector space V can be spanned by the list L of w_1, w_2, \dots, w_n , then any linearly independent list of vectors u_1, u_2, \dots, u_m must have length $m \leq n$. The proof suggests adding a new u_i to the existing list L and then remove another w_j from L in each step. In the end all the u 's will be added and the corresponding number of w 's are removed from L . Therefore, $m \leq n$.
 - The direct consequence of this theorem is the definition of dimension. But before that we need to introduce the notion of a basis and its existence.
- A *basis* of V is a linearly independent spanning list of V . A list is a basis iff all vectors in V can be written as a unique linear combination of the basis vectors. This follows directly from the definition.
- Most of time we are concerned with vector spaces with finite bases. The vector spaces that have a finite spanning set are called *finite-dimensional vector spaces* (FDVS), and we will soon know why they bear this name.
 - First of all, we can show that every finite spanning list of a FDVS V can be reduced to a finite basis.

To prove this first claim, we construct our basis B from scratch and try to add all the elements of the spanning list L of v_1, v_2, \dots, v_n one by one into the basis, as long as the element added is not in the span of the existing vectors in B . In this way, every time a new vector is added to B , B will still be linearly independent, and this B will be linearly independent in the end. Furthermore, $\text{span}(B) = \text{span}(L) = V$ because all the v 's left out are in $\text{span}(B)$.

 - * It is also possible to start B from L and check from v_1 to v_n whether each vector is in the span of the rest of the vectors in B . If it is, then we remove it from B ; if it is not, then we keep it. In the end, $\text{span}(B) = \text{span}(L)$ and none of the vectors in

B is in the span of the others in B . However, this top-down approach is harder in practice to actually find the basis.

* This claim shows that every FDVS has a finite basis.

- Furthermore, every linearly independent list of a FDVS V can be extended to a basis.

To prove this second claim, similarly construct our basis from the given linearly independent list L . We pick a finite basis B of V consisting of b_1, \dots, b_n , and check whether the next b may be added to L (same to what we did above). In the end, all the b 's are in the span of L and L remains linearly independent.

- Choose two finite bases of a vector space. The fact that they are both linearly independent and both span V tells us that the two bases should have the same size, which we define as the *dimension* of a vector space. Thus, the vector spaces with finite bases are called *finite-dimensional*.
 - The fact that the bases of V has a fixed size implies every spanning list or linearly independent list of this size $\dim V$ is a basis itself (because the reduction/extension here is the trivial one).
- The subspace dimension $\dim U$ is always \leq the whole space dimension $\dim V$. To accomplish this we have to resort to the classic procedure: we add vectors from U that are not in the span of the existing ones in L . Thus L always remains linearly independent, and it should always be smaller in size than B , a spanning list of U . Our procedure will eventually terminate and no more vectors can be added to L , i.e., all the vectors in V is in $\text{span}(L)$. L is our desired basis, and $|L| = \dim U \leq \dim V = |B|$.
 - Note it is WRONG to assume that a finite basis of the whole space can always be reduced to a finite basis of the subspace!
 - Corollary: for FDVS V and its subspace U , if the two spaces have the same dimension, then they are the same space. This is an important result useful in many proofs.
- There are two more things that deserve mentioning. For FDVS V and its subspace U , we can always find subspace W such that $V = U \oplus W$. The idea is to find a basis B_U of U and extend it to the basis B_V of V . Using the set notation, $\text{span}(B_V \setminus B_U)$ is our desired W . We then only have to show $V = U + W$ and $U \cap W = \{0\}$, which is not difficult.
- The dimension of the sum of two vector spaces U_1 and U_2 can be expressed as follows:

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

- Despite the fact that it looks like the inclusion-exclusion formula, $+$ and \cup are quite different, thus the formula cannot be similarly generalized to higher orders. To prove the formula, start from a basis u_1, u_2, \dots, u_m of $U_1 \cap U_2$ (intersection is a subspace). U_1 and U_2 then have their respective bases B_1 and B_2 . We can show $B_1 \cup B_2$ is a basis of $U_1 + U_2$. See page 47 of the book for the full proof.

- Note that when the sum is a direct sum, then the formula is just $\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2$. We will generalize this result in §3.5.

3 Linear Maps

3.1 The Vector Space of Linear Maps

- A linear map(ping)/transformation is a homomorphism of vector spaces under addition and scalar multiplication. The linear structure is preserved after the transformation T . To use symbols, a linear map is a function $T : V \rightarrow W$ such that for all $u, v \in V$ and $\lambda \in \mathbf{F}$ (the field of V and W),

$$T(u + v) = Tu + Tv \quad \text{and} \quad T(\lambda u) = \lambda T(u).$$

- The zero transformation maps every vector of V to 0_W . The identity transformation $I_V : V \rightarrow V$ maps each element of V to itself in V .
- We define the addition and scalar multiplication of linear maps in the vector-wise way:

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

following which we can easily show that the set of linear maps $\mathcal{L}(V, W)$ from V to W is a vector space over \mathbf{F} . In addition to the linear properties, we define the *product* ST of linear maps same as $S \circ T$, given the appropriate domain and codomain. Why the product is essential will become clear later. The theory of functions tells us this product is associative [$T_1(T_2T_3) = (T_1T_2)T_3$] and the identity works as usual ($I_W T = T I_V = T$). The distributive properties [$(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$] obviously hold as well.

- The uniqueness of the linear map that takes basis vectors v_1, v_2, \dots, v_n to corresponding vectors w_1, w_2, \dots, w_n is the most fundamental theorem in this section. It is our first step toward showing that linear maps and matrices are isomorphic. To prove the theorem, for arbitrary a_i 's, let

$$T : \sum_{i=1}^n a_i v_i \mapsto \sum_{i=1}^n a_i w_i.$$

- Here are a few properties that are useful in linear maps. The proofs are quite routine.
 - $T(0_V) = 0_W$;
 - T is linear iff $T(cx + y) = cT(x) + T(y)$ for all c, x, y .

3.2 Null Spaces and Ranges

- The *nullspace/kernel* $\text{null } T$ is the preimage of 0_W . The *range/image* $\text{range } T$ is defined the same as the range of a function. The nullspace and the range are respectively subspaces of V and W (by checking the three criterion). If the nullspace (resp. range) is finite-dimensional, then its dimension is the *nullity* (resp. *rank*) of T .
 - Since $\dim U = \dim W$ for subspace U of W implies $U = W$, $\dim \text{range } T = \dim W$ implies that T is surjective.

- An easy theorem to always keep in mind is that the map of the basis vectors of V spans $\text{range } T$. We will use it in the proof of the next theorem.
- We are now endowed with all the tools to prove the **rank-nullity theorem**: for a FDVS V and T from V to W , $\text{range } T$ is finite-dimensional with

$$\dim V = \dim \text{null } T + \dim \text{range } T.$$

- As always, we need to construct a basis to work with first. Let u_1, u_2, \dots, u_n be a basis of $\text{null } T$, and thus can be extended to $u_1, \dots, u_m, v_1, \dots, v_n$ a basis of V . We prove $T(v_1, v_2, \dots, v_n)$ is a basis for $\text{range } T$. The fact that the u 's are in the nullspace shows that Tv_1, \dots, Tv_n spans $\text{range } T$ (by the “easy” theorem above). To show the elements are linearly independent, consider that

$$a_1 T(v_1) + \dots + a_n T(v_n) = 0$$

implies $a_1 v_1 + \dots + a_n v_n \in \text{null } T$. Yet we also have $b_1 u_1 + \dots + b_n u_n \in \text{null } T$, so we can set the two linear combinations equal. Because the u 's and v 's are altogether linearly independent, the coefficients a 's (and b 's) are all 0.

- The rank-nullity theorem gives rise to a number of equivalent characterizations regarding injection and surjection. First of all, there is a criterion that says T is injective iff $\text{null } T = 0$ (proof follows straight from the definition of nullspace). The results below follow.
 - A map to a smaller dimensional space cannot be injective because $\dim(\text{null } T) > 0$.
 - A map to a large dimensional space cannot be surjective because $\dim(\text{range } T) \leq \dim V < \dim W$.
 - V and W are FDVS of equal dimension, then

$$T \text{ is injective} \iff T \text{ is surjective} \iff \dim \text{range } T = \dim V.$$

- Solving systems of linear equations is a direct application of this.
 - A homogeneous system of linear equations with more variables than equations has nonzero solutions. Here we can view the system as a linear map T that maps the vector of unknown variables $x \in \mathbf{F}^n$ to $0 \in \mathbf{F}^m$ with $n > m$. The nullity of T is greater than 0 as a result.
 - Similarly, the T associated with an inhomogeneous equation with more equations than variables is not surjective, and thus some choices of constants on the right side will give no solutions to the system.

3.3 Matrices

- Consider the n -dimensional V with basis v_1, \dots, v_n and m -dimensional W with basis w_1, \dots, w_m . For $T \in \mathcal{L}(V, W)$, we can find for every v_k , a list of $A_{i,k}$'s such that

$$Tv_k = A_{1,k}w_1 + \dots + A_{m,k}w_m = \sum_{j=1}^m A_{j,k}w_j.$$

As a result, we have n lists of scalar a 's, each of length m , and we can put them into an $m \times n$ matrix A , with entries $a_{i,j}$ in the i -th row and j -th column. $\mathcal{M}(T, (v_1, \dots, v_n), (w_1, \dots, w_m))$ uniquely determines the linear map $T \in \mathcal{L}(V, W)$.

– Assume

$$\begin{aligned} Tv_1 &= A'_{1,1}w_1 + \dots + A'_{1,m}w_m, \\ &\vdots \\ Tv_n &= A'_{n,1}w_1 + \dots + A'_{n,m}w_m, \end{aligned}$$

from which we get the $m \times n$ matrix A' . Its transpose $n \times m$ A is our defined matrix representation of T . The reason for taking transpose will become clear later.

- We define matrix addition entry-wise, because under this definition, matrix representation of linear maps preserves addition, i.e. $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$ for $S, T \in \mathcal{L}(V, W)$, assuming $S + T$, S , and T share the same ordered bases.
- Similarly entry-wise scalar multiplication with matrices give us $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.
- Given the definition of matrix addition and scalar multiplication, $\mathbf{F}^{m,n}$, the set of matrices with m rows and n columns over \mathbf{F} , is a vector space of dimension mn .
- The reason why we represent linear maps from n -dimensional VS to m -dimensional VS is that we want to preserve the ordering of “ S ” and “ T ” after matrix representation, i.e., $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ instead of $\mathcal{M}(T)\mathcal{M}(S)$, under the usual definition of matrix multiplication:

$$[AC]_{j,k} = \sum_{r=1}^n A_{j,r}C_{r,k}.$$

– Remember that matrix multiplication is not commutative, but is associative and distributive.

- The book defines $A_{j,\cdot}$ and $A_{\cdot,k}$ as the j -th row and the k -th column of matrix A , and thus $[AC]_{j,k} = A_{j,\cdot}C_{\cdot,k}$. Also, k -th column of matrix product equals matrix times the k -th column of the second matrix:

$$[AC]_{\cdot,k} = AC_{\cdot,k},$$

which is an alternative to understanding matrix multiplication. Obviously there is a row

equivalent version that says j -th row of the matrix product equals j -th row times the second matrix.

- When we are considering matrix A times a column vector c , we can see Ac as the linear combination $c_1A_{\cdot,1} + \cdots + c_nA_{\cdot,n}$.

3.4 Invertibility and Isomorphic Vector Spaces

Remark. The isomorphism between linear maps and matrices is the most central theorem of linear algebra. With the machinery developed in §3.3 and §3.4, all the nice results we will prove about linear maps can be directly **translated to matrices**.

- $T \in \mathcal{L}(V, W)$ is *invertible* if there exists S such that $ST = I_V$ and $TS = I_W$. We denote this inverse S by T^{-1} . One can show that the inverse must be linear and thus belongs to $\mathcal{L}(W, V)$. By the associativity of linear maps, the inverse is unique.
- Invertibility is equivalent to bijectivity between vector spaces. Linear maps are functions, so beyond showing that the inverse is linear, the proof is exactly the same.
- A (linear) *isomorphism* is an invertible linear map; if an isomorphism exists between two vector spaces, then the vector spaces are (linearly) *isomorphic*.
- Two FDVS over \mathbf{F} are isomorphic iff they have the same dimension. Injection gives us $\dim \text{null} = 0$ and surjection gives us $\dim \text{range } T = \dim W$. The rank-nullity theorem tells us $\dim V = \dim W$. On the other hand to prove isomorphism, we show $T(a_1v_1 + \cdots + a_nv_n) = a_1w_1 + \cdots + a_nw_n$ is both injective and surjective, which is not hard.
 - The theorem above implies vector spaces of dimension n are always isomorphic to \mathbf{F}^n .
- Now we show the isomorphism between a linear map and its corresponding matrix. Given ordered bases v_1, \dots, v_n of V and w_1, w_2, \dots, w_m of W , then \mathcal{M} is a function from $\mathcal{L}(V, W)$ to $\mathbf{F}^{m,n}$. As we have noted earlier \mathcal{M} is linear. Thus, it suffices to show that \mathcal{M} is bijective.
 - For all $A \in \mathbf{F}^{m,n}$, let

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

Then this is the *unique* (recall the theorem we stressed in §3.1) T such that $A = \mathcal{M}(T)$.

- A linear map and its matrix representation are homomorphic not merely in regards to linearity, but also in regards to the preservation of the order of the product $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.
- By the theorems on isomorphism above, $\dim \mathcal{L}(V, W) = mn = (\dim V)(\dim W)$.
- We now shift our focus from the matrix representing an entire linear map to the matrix/column vector representing a vector in the same V . We fix the basis v_1, \dots, v_n , then for any $v \in V$, we

define

$$\mathcal{M}(v) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix},$$

where a_1, \dots, a_n are given by $v = a_1 v_1 + \dots + a_n v_n$. Here \mathcal{M} is an isomorphism from V to $\mathbf{F}^{n,1}$.

- Following this definition, $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(T v_k)$.
- More importantly, the linear map of a vector can now be completely described by matrix-multiplication. Symbolically, we want to prove $\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v)$ for all $v \in V$. Since $v = a_1 v_1 + \dots + a_n v_n$ and \mathcal{M} is linear,

$$\begin{aligned} \mathcal{M}(Tv) &= a_1 \mathcal{M}(T v_1) + \dots + a_n \mathcal{M}(T v_n) \\ &= a_1 \mathcal{M}(T)_{\cdot,1} + \dots + a_n \mathcal{M}(T)_{\cdot,n} \\ &= \mathcal{M}(T) \mathcal{M}(v). \end{aligned}$$

- How do we interpret this? Consider the following diagram under the appropriate bases for V and W .

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \mathcal{M} \uparrow \simeq & & \mathcal{M} \uparrow \simeq \\ \mathbf{F}^{n,1} & \xrightarrow{\mathcal{M}(T)} & \mathbf{F}^{m,1} \end{array}$$

For a vector $v \in V$, its image $\mathcal{M}(v)$ under the isomorphism \mathcal{M} of V belongs to $\mathbf{F}^{n,1}$. Matrix multiplication with $\mathcal{M}(T)$ gives us $\mathcal{M}(Tv) \in \mathbf{F}^{m,1}$, which we identify with Tv under the isomorphism \mathcal{M} of W . This is why a linear map can be described entirely by its matrix representation.

- If $V = \mathbf{F}^n$ and $W = \mathbf{F}^m$ and we are using the canonical bases, then every linear map is equivalent to a unique matrix multiplication on the left. *Every matrix-left-multiplication is a linear map, and every linear map is a matrix-left-multiplication.* Think about the Jacobian matrix representation of the total derivative in the Euclidean space. Left-multiplication by a matrix corresponds to the linear *transformation* of a vector.
- A *linear operator* is a linear map from a VS to itself. The notation $\mathcal{L}(V)$ stands for $\mathcal{L}(V, V)$.
- Suppose V is a FDVS and $T \in \mathcal{L}(V)$, then (a) T is invertible iff (b) T is injective iff (c) T is surjective. (This is not true for infinite-dimensional vector space.) This is true for functions $f : S \rightarrow S$ (where S is finite) in general, but for linear maps the proof is likely to be rather simple:
 - (a) \implies (b) we proved earlier.
 - (b) \implies (c) by checking the rank-nullity theorem. T being injective means $\dim \text{null } T = 0$, which implies $\dim \text{range } T = \dim V$.
 - (c) \implies (a) by showing T is injective, as $\dim \text{null } T = \dim V - \dim \text{range } T$, which is 0.

3.5 Products and Quotients of Vector Spaces

- The *product* $V_1 \times \cdots \times V_n$ of vector spaces V_1, \dots, V_n over \mathbf{F} is the set

$$\{(v_1, \dots, v_n) \mid v_1 \in V_1, \dots, v_n \in V_n\}.$$

The addition and scalar multiplication on the product is defined in the usual entry-wise way. It is routine to prove that this product is a vector space over \mathbf{F} as well.

- $\dim(V_1 \times \cdots \times V_n) = \dim V_1 + \cdots + \dim V_n$. Why? Choose a basis for each V_i , then put every basis vector of V_i into the i -th entry of (v_1, \dots, v_n) and let other entries be 0. These vectors are linearly independent and span the product and thus is a basis of the product.
- For subspaces U_1, \dots, U_n of V , define the linear map $\Gamma : U_1 \times \cdots \times U_n \rightarrow U_1 + \cdots + U_n$ by

$$\Gamma(u_1, \dots, u_n) = u_1 + \cdots + u_n.$$

(One can easily check that this map is indeed linear.) Then $U_1 + \cdots + U_n$ is a direct sum iff Γ is injective (or invertible, since Γ is surjective by its definition.) To prove this, we just need to show $\text{null } \Gamma = 0$ iff the only way to write $0 = u_1 + \cdots + u_n$ is to let all u 's be 0. This is quite clear.

- The previous two theorems lead to the following important one. A sum of subspaces U_i 's of V is a direct sum iff dimensions add up. This is because two isomorphic vector spaces must have the same dimension. In both directions we would encounter

$$\dim(U_1 + \cdots + U_n) = \dim(U_1 \times \cdots \times U_n) = \dim U_1 + \cdots + \dim U_n.$$

Construct the linear map Γ above from the product of subspaces makes the proof really clean. Recall that we have proved the special case $n = 2$ already.

- An *affine subset* of V is a subset of V of the form $v + U = \{v + u \mid u \in U\}$ for some vector $v \in V$ and some subspace U of V . Here the affine subset $v + U$ is said to be *parallel* to U .
 - $v + U$ is a subspace of V iff $v \in U$. If $v \in U$, then $v + U = U$, subspace of V . If $v \notin U$, then $0 \notin v + U$, and thus $v + U$ cannot be a subspace of V .
- Under the same setting, the quotient space V/U is the set of all affine subsets of V parallel to U , i.e., $V/U = \{v + U \mid v \in V\}$.
- For subspace U of V and $v, w \in U$, then

$$(a) \ v - w \in U \iff (b) \ v + U = w + U \iff (c) \ (v + U) \cap (w + U) \neq \emptyset.$$

(a) \implies (b) as $v + u = w + ((v - w) + u) \in w + U$ (and the same for the other direction).
 (b) \implies (c) because $v + U$ is not empty, while (c) \implies (a) because $v + u_1 = w + u_2$ for some u_1 and u_2 , and we then move $v - w$ to the left side.

- The theorem above tells us that two affine subset to U are either the same or disjoint.
- The addition and scalar multiplication on V/U are defined by

$$(v + U) + (w + U) = (v + w) + U \quad \text{and} \quad \lambda(v + U) = \lambda v + U.$$

We should pay particular attention to this definition because **an affine subset do not have a unique representation** (for $v \neq v'$, it is possible that $v + U = v' + U$ still)! Therefore, one has to show that different representations $v + U$ and $w + U$ of the same affine subsets of U still gives us a **unique** $(v + w) + U$, so that the addition function above is well-defined (and similar for scalar multiplication). Suppose $v + U = v' + U$ and $w + U = w' + U$, then essentially we want to prove $(v + w) + U = (v' + w') + U$. By assumption $v - v' \in U$, $(v + w) - (v' + w') \in U$, and we reach the conclusion (and similar for scalar multiplication).

- V/U with the two operations above is a vector space. Showing this is routine, but one has to remember that V/U is not a subspace of V . Despite all the elements of the quotient space is in the whole space, the additive identity is now $0 + U = U$, and the additive inverse of $v + U$ is $-v + U$. (The quotient space is discussed with respect to a fixed subspace.) Therefore, we have to check the 8 VS properties.

- Also, by the previous theorem, for any $v \in U$, $v + U$ equals the 0 vector $(0+U)$ and is thus not a basis vector for V/U .

- To find the dimension of V/U , we appeal to the trick of constructing a linear map again. (One should check its linearity) The **quotient map** $\pi : V \rightarrow V/U$ is given by

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

For FDVS V and its subspace U , $\dim V/U = \dim V - \dim U$. π is a map onto V/U obviously, and null $\pi = U$. Therefore the formula follows.

- A basis from V/U combined with a basis from U form a basis of V . Suppose v_1, \dots, v_n are the basis vectors of V/U and u_1, \dots, u_m are the basis vectors of U . If we write

$$a_1 v_1 + \dots + a_n v_n + b_1 u_1 + \dots + b_m u_m = 0,$$

then necessarily $(a_1 v_1 + \dots + a_n v_n) \in U$, which implies that

$$0 + U = (a_1 v_1 + \dots + a_n v_n) + U = a_1(v_1 + U) + \dots + a_n(v_n + U).$$

Because the $v + U$'s form a basis of V/U , the coefficients a_1, \dots, a_n must be 0. Since u_1, \dots, u_m form a basis of U , their coefficients b_1, \dots, b_m must also be 0. And because $v_1, \dots, v_n, u_1, \dots, u_m$ is of length $n + m$, by the previous theorem this indeed is a basis of V .

- Now we define the induced map $\tilde{T} : V/(\text{null } T) \rightarrow W$ of $T : V \rightarrow W$ by

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

$u + \text{null } T = v + \text{null } T$ implies $u - v \in \text{null } T$, and thus $Tu - Tv = 0$. Therefore the induced map is well defined. Checking the linearity of \tilde{T} is routine. Note that $T = \tilde{T} \circ \pi$.

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ & \searrow \pi \quad \nearrow \tilde{T} & \\ & V/\text{null } T & \end{array}$$

- The definition leads to the isomorphism theorem that exists for different algebraic structures. For $T \in \mathcal{L}(V, W)$, \tilde{T} is injective and $\text{range } \tilde{T} = \text{range } T$. The latter simply follows from the definition, while the first requires us to prove $\text{null } \tilde{T} = \{0 + \text{null } T\}$, the identity. $\tilde{T}(v + \text{null } T) = Tv = 0$ gives us $v - 0 \in \text{null } T$. Hence, $v + \text{null } T = 0 + \text{null } T$, and the conclusion follows.
 - The two claims combined tell us $V/(\text{null } T) \simeq \text{range } T$ under the isomorphism \tilde{T} .

3.6 Duality

- A *linear functional* on V is a linear map belonging to $\mathcal{L}(V, \mathbf{F})$. The *dual space* of V , denoted usually by V^* or V' , is the vector space $\mathcal{L}(V, \mathbf{F})$.
- We proved earlier that $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$. Therefore, $\dim V' = \dim V$ for finite-dimensional V .
- The *dual basis* of v_1, \dots, v_n , basis of V , is a list of vectors $\varphi_1, \dots, \varphi_n$ in V' , where each φ_i is a linear functional defined by

$$\varphi_i(v_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

These φ_i 's are designed to form a basis of V' . It suffices to show that $\varphi_1, \dots, \varphi_n$ is linearly independent in V' . Consider

$$a_1\varphi_1 + \dots + a_n\varphi_n = 0.$$

Since $(a_1\varphi_1 + \dots + a_n\varphi_n)(v_j) = a_j$, all a_j 's are 0.

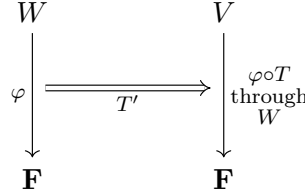
- Remember that the dual basis vector φ_i of e_1, \dots, e_n in \mathbf{F}^n has

$$\varphi_i(x_1, \dots, x_n) = \varphi_i(x_1e_1 + \dots + x_ne_n) = x_i.$$

Namely, the i -th dual basis vector projects (x_1, \dots, x_n) to its i -th coordinate.

- For v_1, \dots, v_n basis of V and its associated dual basis $\varphi_1, \dots, \varphi_n$ of V' ,

- first, every $\psi \in V'$ can be written as $\sum_{i=1}^n \psi(v_i)\varphi_i$. Since $(\sum_{i=1}^n \psi(v_i)\varphi_i)(v_j) = \psi(v_j)$, the sum and ψ agrees on all basis vectors and are thus equal;
- second, every $v \in V$ can be written as $\sum_{i=1}^n \varphi_i(v)v_i$, since for $v = a_1v_1 + \cdots + a_nv_n$, every $a_j = \varphi_j(v)$.
- The **dual map** $T' \in \mathcal{L}(W', V')$ of $T \in \mathcal{L}(V, W)$ is given by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$. We have to be very clear about what this definition means. Here T' maps a function to a function.



T' is a composition of linear maps and is thus a linear map from V' to \mathbf{F} . To show that T' is linear from W' to V' , one just follows the definition.

- The following algebraic properties hold for dual maps:

$$\begin{aligned}
 * & (S + T)' = S' + T', \\
 * & (\lambda T)' = \lambda T', \\
 * & (ST)' = T'S'
 \end{aligned}$$

under the appropriate S, T , and λ . The first two are routine, and the third one is manipulation of function compositions.

- We define **annihilator** U^0 of set $U \subseteq V$ as $\{\varphi \in V' \mid \varphi(U) = 0\}$, the subset of all linear functionals on V whose image of subset U is 0.
- There is a good example (3.104) in the book about annihilators. For the dual basis $\varphi_1, \dots, \varphi_5$ in $(\mathbf{F}^5)'$ that corresponds to e_1, \dots, e_5 in \mathbf{F}^5 , $(\text{span}(e_1, e_2))^0 = \text{span}(\varphi_3, \varphi_4, \varphi_5)$, since φ_i takes out the i -th coordinate.
- As expected U^0 is a subspace of V' . The zero linear functional takes $U \subseteq V$ to 0. $(\varphi + \psi)(U) = 0$ and $\lambda\varphi(U) = 0$ are omitted as usual.
- For FDVS V and its subspace U ,

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

The book provides two proofs. The first proof takes the usual “choose a basis and extend it” approach. The second proof is more obscure: consider the inclusion map $i \in \mathcal{L}(U, V)$ such that $i(u) = u$ for all $u \in U$, then $i' \in \mathcal{L}(V', U')$, which gives us

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

Note that $i'(\varphi) = \varphi \circ i$ for $\varphi \in V'$, giving us $\text{null } i' = U^0$. The nullspace of i' are the set of φ 's

such that $\varphi \circ i = 0$, and therefore φ should map the entire U to 0.

Now we show $U = \text{range } i'$. A linear map defined on a subspace can always be extended to the whole space, and thus we may extend any $\varphi \in U'$ to $\psi \in V'$. The dual map i' maps ψ back to φ , so $\varphi \in U' \implies \varphi \in \text{range } i'$, showing us what we desire.

- The nullspace and range of the dual map T' can now be characterized in terms of the concepts we have introduced.
 - For $T \in \mathcal{L}(V, W)$, $\text{null } T' = (\text{range } T)^0$:

$$\varphi \in \text{null } T' \iff (\varphi \circ T)(v) = \varphi(Tv) = 0 \iff \varphi \in (\text{range } T)^0.$$

- If V, W are finite-dimensional, then we have

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

- $\dim \text{range } T' = \dim \text{range } T$ easily follows from the rank-nullity theorem and the two bullet points above.
- Also we can prove $\text{range } T' = (\text{null } T)^0$. Showing $\text{range } T' \subseteq (\text{null } T)^0$ is standard (just assume T' maps φ to $\psi \in \text{range } T'$), while the other direction is quite tricky. The book assumes finite-dimension here and showed the two spaces have the same dimension:

$$\dim \text{range } T' = \dim \text{range } T = \dim V - \dim \text{null } T = \dim(\text{null } T)^0.$$

We have to make a side note here. When we proved the first bullet point, we did not assume finite-dimension of V and W , so it is natural to consider if our proof of $\text{range } T' \supseteq (\text{null } T)^0$ also does not require the finite-dimension assumption either.

We may formulate our question this way: for $T \in \mathcal{L}(V, W)$ and $\varphi \in V'$ with $\varphi(\text{null } T) = 0$, can we always find a $\psi \in W'$ such that $\varphi = \psi \circ T$? This turns out to be true if we assume the axiom of choice from set theory, which has important corollaries in the theory of infinite-dimensional vector spaces. We will not get into details here, but it is interesting to know about this.

- The corollary of the previous claims are the following:

$$T \text{ is surjective} \iff T' \text{ is injective} \quad \text{and} \quad T \text{ is injective} \iff T' \text{ is surjective}.$$

Proof of this is quite standard, and we show the first one here only:

$$\begin{aligned} T \in \mathcal{L}(V, W) \text{ surjective} &\iff \text{range } T = W \\ &\iff (\text{range } T)^0 = 0 \iff \text{null } T' = 0 \iff T' \text{ injective.} \end{aligned}$$

- The *row rank* and *column rank* of matrix A are defined to be the dimension of the span of the rows of A in $\mathbf{F}^{1,n}$ and of the columns of A in $\mathbf{F}^{m,1}$, respectively. As we know from lower-division matrix algebra, they are the same and represent the number of pivots in a reduced row echelon form of a matrix. We now prove this fact with the tools we have.
- $\mathcal{M}(T') = (\mathcal{M}(T))^t$. Consider the dual basis ψ_j ($1 \leq j \leq m$) of W' and φ_j ($1 \leq k \leq n$) of V' . Then

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

which we evaluate at v_k :

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

On the other hand, we have

$$(\psi_j)(Tv_k) = \psi_j \left(\sum_{r=1}^m A_{r,k} w_r \right) = \sum_{r=1}^m A_{r,k} \psi_j(w_r) = A_{j,k}.$$

- First, $\dim \text{range } T = \text{column rank of } \mathcal{M}(T)$. From bases v_1, \dots, v_n and w_1, \dots, w_m of V and W , we have $\text{span}(Tv_1, \dots, Tv_n)$ is isomorphic to $\text{span}(\mathcal{M}(Tv_1), \dots, \mathcal{M}(Tv_n))$ under the isomorphism \mathcal{M} . Thus, the two VS have the same dimensions, with

$$\dim \text{span}(\mathcal{M}(Tv_1), \dots, \mathcal{M}(Tv_n)) = \text{column rank of } \mathcal{M}(T).$$

Recall that $\text{span}(Tv_1, \dots, Tv_n) = \text{range } T$, thus $\dim \text{range } T = \text{column rank of } \mathcal{M}(T)$, as desired.

- For $A \in \mathbf{F}^{m,n}$, define $T \in \mathcal{L}(V, W)$ as the left-multiplication by A map (i.e., $Tx = Ax$). Then $A = \mathcal{M}(T)$ with respect to the canonical bases. It follows that

$$\begin{aligned} \text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\ &= \dim \text{range } T = \dim \text{range } T' \text{ (which we proved a moment ago)} \\ &= \text{column rank of } \mathcal{M}(T') = \text{column rank of } A^t = \text{row rank of } A. \end{aligned}$$

Therefore, *rank* is a property of a finite matrix.

4 Polynomials

Remark. We used facts about polynomials in high school algebra. These facts were intuitively “understood” back then, but we never had the chance to rigorously prove them. An abstract algebra course delves deep into some materials on polynomials, which we still take for granted in this book. For example, note that when evaluating a polynomial $p(z)q(z)$ at $z = a$, although we still directly plug a in and compute $p(a)q(a)$ to get the result, this is formally not rigorous. This “plugging-in” step requires the machinery of *evaluation homomorphism* to be justified. As another example, one should show that for polynomials with coefficients from a field (more generally from an integral domain), $\deg p + \deg q = \deg pq$ holds, before using it.

- The product $pq \in \mathcal{P}(\mathbf{F})$ of polynomials $p(z) = \sum_{i=0}^m a_i z^i$ and $q(z) = \sum_{j=0}^n b_j z^j$ in $\mathcal{P}(\mathbf{F})$ is defined by

$$(pq)(z) = p(z)q(z) = \sum_{i=0}^m \sum_{j=0}^n a_i b_j z^{i+j}$$

for all $z \in \mathbf{F}$ (“expanding out and collecting like terms”). We can interchange the Sigma notation and see that $p(z)q(z) = q(z)p(z)$. Two polynomials commute under multiplication.

- The division algorithm for polynomials is similar to the one in number theory. For $p, s \in \mathcal{P}(\mathbf{F})$ with $s \neq 0$, then there exist a pair of unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ such that $p = sq + r$ and $\deg r < \deg s$.
 - For $\deg p < \deg s$, then $q = 0$ and $r = p$. If $\deg p = n \geq \deg s = m$, we define $T : \mathcal{P}_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F}) \rightarrow \mathcal{P}_n(\mathbf{F})$ by

$$T(q, r) = sq + r.$$

Clearly T is a linear map (by the definition of polynomial addition and multiplication), and we basically want to show T is bijective. For $sq + r = 0$, we must have $q = 0$ and $r = 0$ because otherwise $\deg sq \geq m > \deg r$. Also, notice that

$$\dim(\mathcal{P}_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F})) = n + 1 = \dim \mathcal{P}_n(\mathbf{F}).$$

- Each zero of a polynomial corresponds to a degree-1 factor, i.e., for $p \in \mathcal{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$, $p(\lambda) = 0$ iff $\exists q \in \mathcal{P}(\mathbf{F})$ such that for all $z \in \mathbf{F}$, $p(z) = (z - \lambda)q(z)$. To prove the \implies direction, division algorithm gives us $p(z) = (z - \lambda)q(z) + r$ (here r is a constant) as $(z - \lambda)$ is of degree 1. Thus, $p(\lambda) = 0$ gives us $r = 0$.
- A polynomial $p \in \mathcal{P}(\mathbf{F})$ has at most as many zeros as its degree (≥ 0). $\deg = 0$ and 1 can be checked, and we proceed with induction on higher orders. If p has no zeros, then done. If p has one zero, let

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

where p is of degree $m + 1$ and q is of degree m (with at most m zeros).

- Recall the remark on polynomials we made in Chapter 2. Now we prove that polynomial functions over an infinite field \mathbf{F} have a unique representation. Suppose $f, g \in \mathcal{P}(\mathbf{F})$ and $f(z) = g(z)$ for all $z \in \mathbf{F}$, we want to show that $f = g$.

Define $h = f - g$, and then we have $\forall z \in \mathbf{F}, h(z) = 0$. Suppose h is a nonzero polynomial of degree n , then by the h has at most n zeros by the theorem above, yet \mathbf{F} is an infinite field. Contradiction! Thus $h = f - g$ is the zero polynomial, and $f = g$ as a result.

- In the case of $\mathbf{F} = \mathbf{C}$ or \mathbf{R} , the book gives another proof of the contraposition that if one coefficient out of

$$h(z) = a_0 + a_1z + \cdots + a_mz^m$$

is not zero, then $h(z) \neq 0$ for some $z \in \mathbf{F}$. Assume $m \neq 0$ and pick

$$z = \frac{|a_0| + \cdots + |a_{m-1}|}{|a_m|} + 1 \geq 1.$$

Thus, $z^j \leq z^{m-1}$ for all $j \leq m-1$. By triangle inequality and $|a_m|z > |a_0| + \cdots + |a_{m-1}|$,

$$|a_0 + a_1z + \cdots + a_{m-1}z^{m-1}| \leq (|a_0| + \cdots + |a_{m-1}|)z^{m-1} < |a_mz^m|.$$

Therefore $a_0 + a_1z + \cdots + a_{m-1}z^{m-1} \neq -a_mz^m$.

- The proof of the fundamental theorem of algebra is left to a complex analysis class. Every nonconstant polynomial on \mathbf{C} has a zero, which directly implies that a polynomial p of degree m has m nondistinct roots, which we sketch below.
 - Every nonconstant polynomial over \mathbf{C} has a unique factorization (up to the order of factors) of the form

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

where $c, \lambda_1, \dots, \lambda_m \in \mathbf{C}$. We call this the unique factorization theorem.

- To prove this, by the fundamental theorem of algebra we can factorize $p(z)$ into $(z - \lambda)q(z)$, and again factorize $q(z)$, and so on. We have proved the existence of such factorization.
- Now we proceed to prove the uniqueness. c is clearly unique as the highest coefficient of $p(z)$. Consider $(z - \lambda_1) \cdots (z - \lambda_m) = (z - \tau_1) \cdots (z - \tau_m)$. We want to show that the λ 's and τ 's are the same up to their ordering. Let $z = \lambda_1$, one of the $\lambda_1 - \tau_i$'s on the RHS must be 0. WLOG we assume $\lambda_1 = \tau_1$, and $(z - \lambda_2) \cdots (z - \lambda_m) = (z - \tau_2) \cdots (z - \tau_m)$ holds for all $z \in \mathbf{F}$ except for λ_1 . Actually

$$(z - \lambda_2) \cdots (z - \lambda_m) - (z - \tau_2) \cdots (z - \tau_m) = 0$$

must hold for all $z \in \mathbf{F}$ because we would have a nonzero polynomial with infinite many zeros. Thus, we can proceed to show $\lambda_2 = \tau_2$, and so on.

- For polynomials $p \in \mathcal{P}(\mathbf{C})$ with real coefficients, the complex conjugate of a root is still a root. This follows from $\lambda^m = \bar{\lambda}^m$.
- For nonconstant $p \in \mathcal{P}(\mathbf{R})$, p has a unique factorization into linear and quadratic polynomials with the highest coefficients being 1, i.e.,

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1x + c_1) \cdots (x^2 + b_Mx + c_M), \quad (1)$$

where the c at the front, the λ 's, and the b 's and c 's at the end are all reals and each quadratic factors have negative discriminant.

- (1) would be obviously true if all zeros are reals by the unique factorization theorem. Suppose p has one pair of nonreal complex zeros λ and $\bar{\lambda}$, then

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

What we want to show is that such $(x - \lambda)$ appear exactly the same number of times $(x - \bar{\lambda})$ in the factorization.

Our idea is to use induction on the degree of p . We assume real nonconstant polynomials of degree less than $\deg p$ can be factorized according to (1). If we prove that q (of degree $\deg p - 2$) has real coefficients, then p can be factorized according to (1).

Now for all $x \in \mathbf{R}$,

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

Therefore, for all $x \in \mathbf{R}$, $q(x) \in \mathbf{R}$. If we write $q(x) = a_0 + a_1x + \cdots + a_{n-2}x^{n-2}$ ($n = \deg p$ and the a 's are complex-valued), then taking the imaginary parts on both sides gives us

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

for all $x \in \mathbf{R}$. Hence the imaginary-part coefficients must all be 0, telling us q is of real coefficients.

Now we have a factorization of p given by (1), and we want to show its uniqueness. For the quadratic factors with negative discriminants at the end, each can be uniquely factorized into $(x - \lambda_j)(x - \bar{\lambda}_j)$. Therefore, if we have two different factorizations of p of the form in (1), then we would have two different standard factorizations (\star) of p over \mathbf{C} (as one of factors over \mathbf{R} being different implies that one of the factors over \mathbf{C} is different). This contradicts the unique factorization theorem.

5 Eigenvalues, Eigenvectors, and Invariant Subspaces

5.1 Invariant Subspaces

- For $T \in \mathcal{L}(V)$, the subspace U of V is *invariant under T* if $T(U) \subseteq U$, or alternatively, $T|_U$ is an operator on U .
- It can be easily verified that $\{0\}$, V , $\text{null } T$, and $\text{range } T$ are invariant subspaces of V under T .
- Here we focus on invariant subspaces of dimension 1. Any 1-dimensional subspace U has some nonzero vector $u \in V$ as the basis (in fact any nonzero vector in U can be the basis). To check if T is invariant under U , we only need to check if the basis vectors are mapped into U (so that the $\text{range } T \subseteq U$). Since $Tu \in U$ iff $Tu = \lambda u$, a subspace with dimension 1 is invariant under T iff a (any) nonzero vector u in the subspace has $Tu = \lambda u$.
- For $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{F}$ is an *eigenvalue* of T if there exists a nonzero $v \in V$ such that $Tv = \lambda v$. The v here is called the *eigenvector*.

Any nonzero vector in a 1-dimension subspace invariant under T is an eigenvector. If u is an eigenvector of T , then $\text{span}(u)$ is invariant under T .

- We may move λv to the left side: consider the linear operator λI that maps v to λv . Therefore, $Tv = \lambda v$ is equivalent to $(T - \lambda I)v = 0$. Since $v \neq 0$, this is equivalent to $T - \lambda I$ is not injective/surjective/invertible (since it is an operator on V).
 - A special case: T is not invertible iff 0 is an eigenvalue of T . This will become useful in Chapter 8.
- Eigenvectors corresponding to distinct eigenvalues are linearly independent. The usual proof of this starts with 2 distinct eigenvalues and proceeds with induction, but the book uses the well-ordering principle instead.

Suppose the v_1, \dots, v_n corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$ are linearly dependent, then at $k \leq n$ we have $v_k \in \text{span}(v_1, \dots, v_{k-1})$ with v_1, \dots, v_{k-1} being linearly dependent. Thus,

$$v_k = a_1 v_1 + \dots + a_{k-1} v_{k-1}, \quad (2)$$

and by applying T to both sides, we have

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

If instead, we multiply (2) by λ_k on both sides, then

$$\lambda_k v_k = a_1 \lambda_k v_1 + \dots + a_{k-1} \lambda_k v_{k-1}.$$

Subtracting the two equations above give us

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

Since v_1, \dots, v_{k-1} are linearly independent, all the coefficients are 0 and thus all the a_i 's are the same (because the λ 's are distinct). However, this implies $v_k = 0$, which contradicts with v_k being an eigenvector. Therefore, the set of eigenvectors are linearly independent.

- The direct corollary of this is that the number of eigenvalues of $T \in \mathcal{L}(V)$ cannot exceed the $\dim V$.
- Here we restate the $T|_U \in \mathcal{L}(U)$ as the *restriction operator* of T . The *quotient operator* $T/U \in \mathcal{L}(V/U)$ is given by $(T/U)(v + U) = Tv + U$ for any $v \in V$. Of course for the quotient operator we need to verify it is well-defined for different representation of the same quotient space, which is very similar to what we have done before and is thus omitted. These two operators provide nice information about the original T .

5.2 Eigenvectors and Upper-Triangular Matrices

Remark. (In the upcoming fourth edition, this section is rewritten in terms of minimal polynomials.) Most theorems in this section rely heavily on induction. This is because we will be considering bases and matrices of arbitrary finite dimensions.

- Linear operators allow power operations. Given $T \in \mathcal{L}(V)$, we define T^m ($m \in \mathbf{Z}^+$) as T composite with itself m times, T^0 as the identity operator I_V , and T^{-m} as $(T^{-1})^m$, given T^{-1} exists.
 - Obviously $T^m T^n = T^{m+n}$ and $(T^m)^n = T^{mn}$. (If T is invertible then we can m, n are allowed to be all integers.)
- $p(T)$ is the polynomial p taking T as its indeterminate (although T is not \mathbf{F}). For $p(z) = a_0 + a_1 z + \dots + a_n z^n$ over \mathbf{F} , we define $p(T) = a_0 I + a_1 T + \dots + a_n T^n$.
 - If we fix an operator T , the evaluation map from $\mathcal{P}(\mathbf{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.
 - Recall that $p(z)q(z) = q(z)p(z)$ for polynomials. It follows directly that $p(T)q(T) = q(T)p(T)$, because we are only replacing the indeterminate symbol z by the operator symbol T . Thus, when computing the polynomial of a linear map, the order of the factors does not matter.
- It turns out that every operator on a finite-dimensional and nonzero complex vector space necessarily has an eigenvalue. (This is not necessarily true on real vector spaces. See example 5.8(a) on the book: $T(w, z) = (-z, w)$ for $T \in \mathcal{L}(\mathbf{R}^2)$ over $\mathbf{F} = \mathbf{R}$.)

The proof employs the fundamental theorem of algebra and is quite tricky. Choose a nonzero $v \in V$ ($\dim V = n$) and consider the list $v, Tv, \dots, T^n v$, which must be linearly dependent. Thus, we have

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

where the a_i 's are not all 0. In particular, a_1, \dots, a_n cannot all be 0 since $v \neq 0$ will then give us $a_0 = 0$. Thus, we have a nonconstant polynomial $a_0 + a_1 z + \dots + a_m z^m$ ($m \leq n$), which

can be factorized into

$$c(z - \lambda_1) \cdots (z - \lambda_m)$$

over \mathbf{C} . For

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

the nonzero vector $w = (T - \lambda_{j+1} I) \cdots (T - \lambda_m I)v$ will be taken to 0 by $(T - \lambda_j I)$. $(T - \lambda_j I)$ is thus not injective and λ_j is an eigenvalue.

- We construct the matrix of an operator always with respect to only one ordered basis of V . The matrix $\mathcal{M}(T, (v_1, \dots, v_n))$ is a square matrix. In the j -th column of $\mathcal{M}(T)$, each entry corresponds to each coefficient from $a_1 v_1 + \cdots + a_n v_n = T v_j$. A square matrix with all entries being 0 below the diagonal is called an *upper-triangular matrix*, which is crucial in the theory of eigenvalues.
- How are invariant subspaces and upper-triangular matrices connected? Suppose $T \in \mathcal{L}(V)$ and v_1, \dots, v_n is a basis of V , then the following are equivalent:
 - (a) $\mathcal{M}(T)$ is upper-triangular;
 - (b) $T v_j \in \text{span}(v_1, \dots, v_j)$ for each j from 1 to n ;
 - (c) $\text{span}(v_1, \dots, v_j)$ is invariant under T for each j from 1 to n .

The first two are equivalent follow straight from the definition. (b) \implies (c) because for every i between 1 and j , $T v_i \in \text{span}(v_1, \dots, v_j)$. (c) \implies (b) is obvious as well.

- For FDVS V over \mathbf{C} and $T \in \mathcal{L}(V)$, then $\mathcal{M}(T)$ is upper-triangular with respect to some basis of V . The book gives two proofs relying on the induction on $\dim V$. *Remember this result will be improved in Chapter 8, where we will show that $\mathcal{M}(T)$ with respect to some basis is not only upper-triangular.*

The first proof first shows that $U = \text{range}(T - \lambda I)$ is invariant under T , which means that U has an upper-triangular matrix with respect to a basis u_1, \dots, u_m of U (by the inductive hypothesis, and is equivalent to $T u_j \in \text{span}(u_1, \dots, u_j)$ for all j).

We extend u_1, \dots, u_m to $u_1, \dots, u_m, v_1, \dots, v_n$, basis of V . Now set $T v_k = (T - \lambda I)v_k + \lambda v_k$. Since $(T - \lambda I)v_k \in U$,

$$T v_k \in \text{span}(u_1, \dots, u_m, v_1, \dots, v_k)$$

for all $1 \leq k \leq n$. This combined with $T u_j \in \text{span}(u_1, \dots, u_j)$ for all j shows that $u_1, \dots, u_m, v_1, \dots, v_n$ is the basis of V that has an upper-triangular matrix.

The second proof is nonstandard but quite interesting. It sets $U = \text{span}(v_1)$ and looks at V/U of dimension $n - 1$ instead. $T/U \in \mathcal{L}(V/U)$ by the inductive hypothesis should have (for all j between 2 and n)

$$(T/U)(v_j + U) \in \text{span}(v_2 + U, \dots, v_j + U).$$

This implies that $T v_j \in \text{span}(v_1, \dots, v_j)$ for j between 1 and n . Recall in §3.5 we showed that

v_2, \dots, v_n combined with the basis vector v_1 from U is a basis of V , and the conclusion follows.

- Upper-triangular matrices help us determine **whether T is invertible**. If $T \in \mathcal{L}(V)$ has an upper-triangular $\mathcal{M}(T)$ with respect to some basis of V , then T is invertible iff all entries on the diagonal of $\mathcal{M}(T)$ are nonzero.

$$\begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}.$$

Suppose the λ 's on the diagonal in the matrix above are all nonzero, then $Tv_1 = \lambda_1 v_1$ implies that $T(v_1/\lambda_1) = v_1$, so that $v_1 \in \text{range } T$. Looking at the next column of the matrix we can continue to show that $T(v_2/\lambda_2) = av_1 + v_2$, which implies that $v_2 \in \text{range } T$. Following this fashion we can conclude that v_1, \dots, v_n are all in $\text{range } T$. Since $\dim V = n$, T is surjective and thus invertible.

Suppose in the other direction that T is invertible, then in the first place $\lambda_i \neq 0$ because otherwise $Tv_1 = 0$ (Since $v_1 \neq 0$, T cannot be injective). For $1 < j \leq n$ suppose $\lambda_j = 0$, “looking” at the j -th column above we immediately know that $T|_{\text{span}(v_1, \dots, v_j)}$ is not injective (it maps into $\text{span}(v_1, \dots, v_{j-1})$ of dimension $j-1 < j$). Thus T (with respect to the whole V) is not injective (thus not invertible) as well. Thus, all the λ 's on the diagonal must be nonzero.

- By the previous theorem, we can determine all the eigenvalues of T if we are fortunate to find a basis that corresponds to an upper-triangular $\mathcal{M}(T)$. $\mathcal{M}(T - \lambda I)$ is of the form

$$\begin{bmatrix} \lambda_1 - \lambda & & & * \\ & \lambda_2 - \lambda & & \\ & & \ddots & \\ 0 & & & \lambda_n - \lambda \end{bmatrix}.$$

One of the $\lambda_i - \lambda$'s is 0 iff $T - \lambda I$ is not invertible iff λ is an eigenvalue. Therefore, all the entries λ_i 's on the diagonal are exactly all the eigenvalues.

- Note that if T has no eigenvalue at all, then by this claim T has no upper-triangular matrix with respect to any basis of V .

5.3 Eigenspaces and Diagonal Matrices

- A *diagonal matrix* has all its entries off the diagonal 0.
- The *eigenspace* of $T \in \mathcal{L}(V)$ corresponding to $\lambda \in \mathbf{F}$ is given by

$$E(\lambda, T) = \text{null}(T - \lambda I) = \{v \in V \mid Tv = \lambda v\}.$$

The eigenspace is basically the vector subspace of V with all eigenvectors corresponding to v , along with the 0 vector.

- The sum of eigenspaces is a direct sum. We consider $u_1 + \cdots + u_m = 0$, where each eigenvector u_i is from the eigenspace $E(\lambda_i, T)$. Because eigenvectors corresponding to distinct eigenvalues are linearly independent, all the u_i 's must be 0. Furthermore, since

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

is a direct sum, we have

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

Recall the theorem in §3.5 that says a sum is a direct sum iff the dimensions add up.

- An operator T is *diagonalizable* if the operator has a diagonal matrix with respect to some basis of T .
- For T on FDVS V with dimension n and distinct eigenvalues $\lambda_1, \dots, \lambda_m$, diagonalizability of T is equivalent to the following four conditions:
 - (a) V has a basis consisting of eigenvectors of T ;
 - (b) V can be decomposed into n 1-dimension subspaces U_1 to U_n , where $V = U_1 \oplus \cdots \oplus U_n$;
 - (c) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;
 - (d) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T))$

The proof to this theorem is a little involved but not hard at all. One should get an intuitive idea of what each equivalent condition means.

- We consider the whole list of eigenvalues $\lambda_1, \dots, \lambda_n$ with repetition ($n \geq m$). Looking at the matrix

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

for a moment would tell us diagonalizability and (a) are equivalent.

- (a) \implies (b): Let each $U_j = \text{span}(v_j)$, each of which are invariant under T . Since v_1, \dots, v_n is a basis of V , any $v \in V$ can be written uniquely as a linear combination of v_1, \dots, v_n , and thus a unique sum of elements from each U_j . Therefore, V is a direct sum of the U_j 's. Trace this proof backward to show (b) \implies (a).
- We now show (a) \iff (c) \iff (d). (c) \iff (d) follows from the theorem in §3.5.
 - * (a) \implies (c) because V has a basis of n eigenvectors, each of which must belong to any of the m distinct eigenspaces $E(\lambda_j, T)$. Therefore, V is the sum of all these eigenspaces (and thus the direct sum of them).

- * To show (c) & (d) \implies (a), choose a basis B_j from each $E(\lambda_j, T)$. Therefore, we get a list of vectors v_1, \dots, v_n of length $n = \dim V$ by (d). To show the list is linearly independent, consider $0 = u_1 + \dots + u_n$, where each u_j is from $E(\lambda_j, T)$. All the u_j 's must be 0 because V is a direct sum of the eigenspaces. If $0 = u_j$, since u_j can be expressed as a linear combination of the basis vectors in B_j , the coefficients in the linear combination must be 0. Therefore our constructed list of eigenvectors v_1, \dots, v_n is indeed a basis of V .
- If $T \in \mathcal{L}(V)$ has $\dim V$ number of distinct eigenvalues, then T is necessarily diagonalizable. Having $\dim V$ number of distinct eigenvalues tells us we have a basis of eigenvectors of T . With respect to this basis $\mathcal{M}(T)$ is a diagonal matrix.

6 Inner Product Spaces

Remark. In this section V is by default an inner product space.

6.1 Inner Products and Norms

- For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n) \in \mathbf{R}^n$ (the real n -dimensional space), the *dot product* $x \cdot y := x_1 y_1 + \dots + x_n y_n$ is a scalar in \mathbf{R} . The norm $\|x\|$ on \mathbf{R}^n is usually given by $\sqrt{x_1^2 + \dots + x_n^2}$, which is the square root of the dot product of x with itself.
- The usual norm $\|z\| := \sqrt{|z_1|^2 + \dots + |z_n|^2}$ on \mathbf{C}^n (the complex n -dimensional space) is also a scalar in \mathbf{R} .
- Generally on both real and complex n -dimensional spaces, we define the *Euclidean inner product* by

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

This generalizes the definition of the dot product on \mathbf{R}^n . Moreover, we have

$$\langle (z_1, \dots, z_n), (z_1, \dots, z_n) \rangle = z_1 \overline{z_1} + \dots + z_n \overline{z_n} = |z_1|^2 + \dots + |z_n|^2 = \|z\|^2,$$

so that $\|z\|$ becomes the square root of the inner product of z with itself.

We regard \mathbf{R}^n and \mathbf{C}^n by default as the n -dimensional real and complex Euclidean spaces endowed with the Euclidean inner product.

- Observation on our definition of the Euclidean inner product suggests how we could possibly define a general abstract inner product on arbitrary real or complex vector spaces. The *inner product* is a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbf{F}$ (\mathbf{R} or \mathbf{C}) with the following properties:
 - Positive definiteness: $\langle v, v \rangle \geq 0$ (**real** and nonnegative), which achieves “=” iff $v = 0$;
 - Linearity in the first slot: $\langle \lambda u + v, w \rangle = \lambda \langle u, w \rangle + \langle v, w \rangle$; (This means $\varphi_w(v) := \langle v, w \rangle$ with respect to a fixed w gives us a linear functional on V .)
 - Conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

The properties that directly ensue include:

- Conjugate symmetry gives us “conjugate linearity” (anti-linearity) in the second slot:

$$\langle u, \lambda v + w \rangle = \overline{\lambda} \langle u, v \rangle + \langle u, w \rangle;$$

- For fixed $w \in V$, $f : v \mapsto \langle v, w \rangle$ is a linear map from V to \mathbf{F} because of linearity in the first slot;
- $\langle 0, v \rangle = 0$ because linear maps take 0 to 0; and $\langle v, 0 \rangle = 0$ by conjugate symmetry.
- The *norm* $\|v\| := \sqrt{\langle v, v \rangle}$. It has properties
 - $\|v\| = 0$ iff $v = 0$, and

$$- \|\lambda v\| = |\lambda| \|v\|.$$

- Two vectors are *orthogonal* if their inner product is 0. Therefore, 0 is orthogonal to any vector and 0 is the only vector that is orthogonal to itself.
- Introducing these concepts provide us with a list of useful equations and inequalities about *inner product spaces* (vector spaces endowed with an inner product) that we know as facts in the Euclidean spaces.
 - The Pythagorean theorem now holds for orthogonal vectors $u, v \in V$:

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

Expanding the LHS, we have $\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle$, where the two terms in the middle are 0. The converse of this theorem holds in real inner product spaces because then we would have $\langle u, v \rangle + \langle v, u \rangle = 2\langle u, v \rangle = 0$, showing that the two vectors are orthogonal.

- The method of orthogonal decomposition is extended to inner product spaces. Suppose $u, v \in V$ with v being nonzero, we wish write u as the sum of scalar multiple of v and another vector w that is orthogonal to v .

Basically we want to choose a scalar $c \in \mathbf{F}$ such that

$$0 = \langle u - cv, v \rangle = \langle u, v \rangle - c\|v\|^2.$$

Let $c = \frac{\langle u, v \rangle}{\|v\|^2}$ and we have $w = u - cv$ and we have our orthogonal decomposition

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

- The well-known Cauchy-Schwarz inequality states that $|\langle u, v \rangle| \leq \|u\| \|v\|$, which reaches equality iff one of u, v is a scalar multiple of the other. This can be proved using the orthogonal decomposition theorem we have above, but we choose to use the more standard method. When $v = 0$, the result is immediate. By the equality condition given, for $v \neq 0$ we may consider

$$0 \leq \|u - cv\|^2 = \langle u, u \rangle - \bar{c}\langle u, v \rangle - c\langle v, u \rangle + c\bar{c}\langle v, v \rangle.$$

Set $c = \|\langle u, v \rangle / \langle v, v \rangle\|$, then the RHS of the “=” becomes $\|u\|^2 - \frac{|\langle u, v \rangle|^2}{\|v\|^2}$, as desired. The equality is achieved if and only if $u = cv$ or $v = 0$.

Two special cases should be kept in mind.

- * On real numbers, we have $|x_1 y_1 + \cdots + x_n y_n|^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$.
- * One can define an inner product on function spaces such as the vector space V of

continuous functions from $[a, b]$ to \mathbf{F} . One can define the inner product on V as such:

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx.$$

This is an important inner product that will be useful in later courses. If $\mathbf{F} = \mathbf{R}$, then we simply have

$$\langle f, g \rangle = \int_a^b f(x)g(x) dx.$$

By the Cauchy-Schwarz inequality, we then have (over \mathbf{R})

$$\left| \int_a^b f(x)g(x) dx \right|^2 \leq \left(\int_a^b (f(x))^2 dx \right) \left(\int_a^b (g(x))^2 dx \right).$$

- The triangle inequality states that $\|u + v\| \leq \|u\| + \|v\|$, which reaches equality iff one of u, v is a nonnegative real multiple of the other. To prove, expanding the LHS gives us

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

Taking the square root gives us the result. To have equality in the third row of our derivation, we need $|\langle u, v \rangle| = \|u\|\|v\|$, which implies the equality in the second row because the norm is a real number. Therefore, the triangle inequality achieves equality iff $|\langle u, v \rangle| = \|u\|\|v\|$.

Note that $|\langle u, v \rangle| = \|u\|\|v\|$ implies $u = cv$ or $v = 0$ by the Cauchy-Schwarz inequality. Also, it is easy to show $u = cv$ or $v = 0$ implies $|\langle u, v \rangle| = \|u\|\|v\|$. Thus, the triangle inequality achieves equality iff one of u, v is a scalar multiple of the other.

- The parallelogram equality says that $\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$. To prove, expand the LHS and after some cancellations, we get the RHS.

6.2 Orthonormal Bases

- A list (set) of vectors is *orthonormal* if each vector is of norm 1 and is orthogonal to all other vectors in the list. That is to say for an orthonormal list of vectors $e_1, \dots, e_n \in V$,

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

- Note that the canonical basis in \mathbf{F}^n is orthonormal.

- Given an orthonormal list e_1, \dots, e_n , by repeated use of the Pythagorean theorem, we get

$$\|a_1e_1 + \dots + a_ne_n\|^2 = |a_1|^2 + \dots + |a_n|^2$$

for all scalars a_1, \dots, a_n .

- It follows from the theorem above that if we set $a_1e_1 + \dots + a_ne_n = 0$, then all the a_i 's must be 0, showing that an orthonormal list of vectors e_1, \dots, e_n are linearly independent. Therefore, an orthonormal list of length $\dim V$ is an *orthonormal basis*, i.e., an orthonormal list (set) that is also a basis.
- Let u_1, \dots, u_n be a basis of a vector space U , we may endow a natural inner product on U (which you can verify) given by

$$\langle a_1u_1 + \dots + a_nu_n, b_1u_1 + \dots + b_nu_n \rangle = a_1b_1 + \dots + a_nb_n.$$

What is special about this inner product is that $\langle u_i, u_i \rangle = 1$ for all i , so that the list of u_i 's becomes an orthonormal basis of the inner product space U .

- Now any $v \in V$ can be written as a unique linear combination of the orthonormal basis vectors e_1, \dots, e_n . In fact, we have

$$v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n.$$

Consider $v = a_1e_1 + \dots + a_ne_n$ and take an inner product with e_j on both sides. Then, $\langle v, a_j \rangle = a_j$, as desired. Note that $\|v\|^2 = |\langle v, e_1 \rangle|^2 + \dots + |\langle v, e_n \rangle|^2$.

- There is an important corollary to the theorem above. For an orthonormal basis of V e_1, \dots, e_n , every

$$Te_k = \langle Te_k, e_1 \rangle e_1 + \dots + \langle Te_k, e_n \rangle e_n.$$

Therefore, we have $\mathcal{M}(T)_{j,k} = \langle Te_k, e_j \rangle$ for the matrix representation of T with respect to e_1, \dots, e_n .

- The Gram-Schmidt process can now be proved. For a linearly independent list v_1, \dots, v_n in V , define $e_1 = \frac{v_1}{\|v_1\|}$, and then for $2 \leq j \leq n$, define e_j inductively by

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

Then for any $j \leq n$, e_1, \dots, e_j is still orthonormal and has the same span as v_1, \dots, v_j .

We use induction for this proof. $j = 1$ obviously holds. Suppose for $1 < j \leq n$ we already have

$$\text{span}(v_1, \dots, v_{j-1}) = \text{span}(e_1, \dots, e_{j-1}).$$

By the linear independence of the list of v 's, $v_j \notin \text{span}(v_1, \dots, v_{j-1})$ and thus

$$v_j \notin \text{span}(e_1, \dots, e_{j-1}).$$

Therefore, $v_j - \langle v_j, e_1 \rangle e_1 - \dots - \langle v_j, e_{j-1} \rangle e_{j-1}$ is nonzero, and we can indeed divide it by its norm to get a new vector e_j with norm 1.

For $1 \leq k < j$, we can show that the new vector e_j is orthogonal to the existing e_k 's, as

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

$v_j \in \text{span}(e_1, \dots, e_j)$ holds by construction and hence $\text{span}(v_1, \dots, v_j)$ is a subspace of $\text{span}(e_1, \dots, e_j)$. The two spaces have the same dimension and are thus equal. This finishes our inductive step.

- This algorithm for constructing an orthonormal basis from a given basis leads to two useful corollaries.
 - * Every FD inner product space has an orthonormal basis. Proof is obvious.
 - * If V is FD, then every orthonormal list of vectors can be extended to an orthonormal basis. We start with e_1, \dots, e_m and extend the list to a basis $e_1, \dots, e_m, v_{m+1}, \dots, v_n$ of V . It is easily observed that if we apply the Gram-Schmidt process from e_1 to v_n , then e_1 to e_m will remain themselves, yet v_{m+1}, \dots, v_n will be orthonormalized to f_{m+1}, \dots, f_n . Thus the e 's and the f 's become the basis extended from the original list of e 's.
- If $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V , then it has an upper-triangular matrix with respect to some orthonormal basis of V .

(In the case of $\mathbf{F} = \mathbf{C}$, since the condition always holds, T must have an upper-triangular matrix with respect to some orthonormal basis of V . This is known as Schur's theorem.)

Recall having upper-triangular $\mathcal{M}(T)$ is equivalent to $\text{span}(v_1, \dots, v_j)$ being invariant under for each $1 \leq j \leq n$. Our claim is ensured by the Gram-Schmidt process.
- We briefly mentioned a while ago that an inner product with the second slot fixed is a linear functional. In fact there is a **one-to-one correspondence** between every $u \in V$ and every $\varphi \in V'$. The Riesz representation theorem states that for finite-dimensional inner product space V and **any** $\varphi \in V'$, there exists a **unique** $u \in V$ such that $\varphi(v) = \langle v, u \rangle$ for every $v \in V$. This is a very important theorem that leads to the definition of adjoint in Chapter 7.

To prove, choose an arbitrary orthonormal basis e_1, \dots, e_n . Then for any v ,

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

Therefore $\overline{\varphi(e_1)} e_1 + \dots + \overline{\varphi(e_n)} e_n$ is our desired u .

To show uniqueness, suppose $\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$, then $\langle v, u_1 - u_2 \rangle = 0$ for all v . And by setting $v = u_1 - u_2$ we have $u_1 = u_2$.

(If for all v , $\langle v, u_1 \rangle = \langle v, u_2 \rangle$ or $\langle u_1, v \rangle = \langle u_2, v \rangle$, then $u_1 = u_2$; we will use this frequently in Chapter 7.)

- By uniqueness, whatever what orthonormal basis e_1, \dots, e_n we choose, the expression

$$\overline{\varphi(e_1)} e_1 + \dots + \overline{\varphi(e_n)} e_n$$

remains the same.

- Thus, if we choose any $u = a_1 e_1 + \dots + a_n e_n$ in U and define a linear functional φ such that

$$\varphi(e_1) = \overline{a_1}, \dots, \varphi(e_n) = \overline{a_n},$$

then φ here is the **unique** linear functional such that $\varphi(v) = \langle v, u \rangle$.

6.3 Orthogonal Complements and Minimization Problems

- For $U \subseteq V$ (we assume U is nonempty by default), the *orthogonal complement* of U (denoted by U^\perp) is the set of all vectors in V that are orthogonal to every vector in U , i.e.,

$$U^\perp = \{v \in V \mid \forall u \in U, \langle v, u \rangle = 0\}.$$

- It is good to see an example in \mathbf{R}^3 : if U is a line in \mathbf{R}^3 through the origin, then all the vectors perpendicular to the line U form the orthogonal complement of U , the plane through the origin perpendicular to the line U . Similarly, if on the other hand, U is a plane in \mathbf{R}^3 through the origin, then its orthogonal complement is the line through the origin perpendicular to the plane U .
- Given $U \subseteq V$, U^\perp is a subspace of V . Check $0 \in U^\perp$ and U^\perp is closed under linear operation.
- It is easy to verify and understand the following properties of orthogonal complements:
 - (a) $0^\perp = V$.
 - (b) $V^\perp = \{0\}$.

- (c) For $U \subseteq V$, $U \cap U^\perp = \begin{cases} \{0\} & \text{if } 0 \in U, \\ \emptyset & \text{if } 0 \notin U. \end{cases}$
- (d) If $U \subseteq W \subseteq V$, then $W^\perp \subseteq U^\perp$.

And we skip their proofs.

- If in particular U is a finite-dimensional subspace of V (V does not have to be finite-dimensional), then $V = U \oplus U^\perp$.

First we prove $V = U + U^\perp$. Let e_1, \dots, e_m be the orthonormal basis of U . Any $v \in V$ can be written into

$$\{\langle v, e_1 \rangle e_1 + \dots + \langle v, e_m \rangle e_m\} + \{v - \langle v, e_1 \rangle e_1 - \dots - \langle v, e_m \rangle e_m\}.$$

Here $\langle v, e_1 \rangle e_1 + \dots + \langle v, e_m \rangle e_m \in U$, and one can easily show that for $w = v - \langle v, e_1 \rangle e_1 - \dots - \langle v, e_m \rangle e_m$, $\langle w, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0$ (for all j between 1 and m).

Therefore, w is orthogonal to $\text{span}(e_1, \dots, e_m) = U$, meaning that $w \in U^\perp$. Since U is a subspace and is thus nonempty, $U \cap U^\perp = \{0\}$, showing that $U + U^\perp$ is a direct sum.

– It follows directly that $\dim U^\perp = \dim V - \dim U$.

- Let U again be a finite-dimensional subspace of V , then $U = (U^\perp)^\perp$. For the “ \subseteq ” direction, any $u \in U$ is orthogonal to any vectors $v \in U^\perp$, while *all* vectors in $(U^\perp)^\perp$ are orthogonal to any $v \in U^\perp$. (Note that we can prove “ \supseteq ” this way because we do not know if U contains all the vectors orthogonal to U^\perp .)

For the “ \supseteq ” direction, suppose $v \in (U^\perp)^\perp \subseteq V$. Then v can be uniquely written as $u + w$, where $u \in U$ and $w \in U^\perp$. We have just proved that $U \subseteq (U^\perp)^\perp$, meaning that $u \in (U^\perp)^\perp$, subspace of V . Therefore, $w = v - u \in (U^\perp)^\perp$ and $w \in U^\perp$ simultaneously. Since U is nonempty, $w = 0$ and thus $v = u \in U$.

- The *orthogonal projection* P_U of V to its finite-dimensional subspace U is the function $P_U : V \rightarrow U$ defined by $P_U(v) = u$, where $u \in U$ and $w \in U^\perp$ are the pair of *unique* elements such that $v = u + w$. It is easy to verify $P_U \in \mathcal{L}(V)$ (note that $U \subseteq V$).
- Also, for orthonormal basis e_1, \dots, e_m , $P_U(v) = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_m \rangle e_m$, which we have showed earlier. In general, if we want to find $P_U(v)$, we should first use Gram-Schmidt process to find an orthonormal basis of U and then use the expression above to find $P_U(v)$.
- We give a list of properties of P_U below.
 - (a) $\forall u \in U, P_U(u) = u$; [for $u = u + 0$]
 - (b) $\forall w \in U^\perp, P_U(w) = 0$; [for $w = 0 + w$]
 - (c) $\text{range } P_U = U$; [for $\text{range } P_U \subseteq U$ by definition, and (a) above; thus P_U is surjective onto U]
 - (d) $v - P_U(v) \in U^\perp$; [true because $w = v - P_U(v)$ is in U^\perp]
 - (e) $\text{null } P_U = U^\perp$; [$U^\perp \subseteq \text{null } P_U$ by (b); also, for v such that $P_U(v) = 0$, meaning that $v - 0 = v \in U^\perp$ by (c)]

(f) P_U is an idempotent operator: $P_U^2 = P_U$; [by (b)]

(g) $\|P_U(v)\| \leq \|v\|$; [since $\|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2$]

- For finite-dimensional subspace U of an inner product space V and a fixed $v \in V$, the minimum distance between v and all vectors on U is achieved at $u = P_U(v)$ (so that $v - u$ is orthogonal to the subspace U), as expected. Formally, we want to show that for all $u \in U$,

$$\|v - P_U(v)\| \leq \|v - u\|,$$

which achieves equality iff $u = P_U(v)$.

The proof obviously employs the Pythagorean theorem. We have

$$\|v - P_U(v)\|^2 \leq \|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 = \|v - u\|^2,$$

because $v - P_U(v) \in U^\perp$ and $P_U(v) - u \in U$, so that the two are orthogonal. The equality condition is obvious.

7 Operators on Inner Product Spaces

Remark. In this chapter we regard V by default as FDVS over $\mathbf{F} = \mathbf{R}$ or \mathbf{C} .

7.1 Self-Adjoint and Normal Operators

- For $T \in \mathcal{L}(V, W)$, the *adjoint* of T is the function $T^* : W \rightarrow V$ such that

$$\langle Tv, w \rangle_W = \langle v, T^*w \rangle_V.$$

Pay attention to the fact that the inner product on the left is on W , while the inner product on the right is on V . What we have above is well-defined by the Riesz representation theorem in §6.2. Fix w , then $\langle Tv, w \rangle$ becomes a linear functional $\psi_{T,w}(v)$ on V (because T is linear), which uniquely determines a $u \in V$ such that $\langle v, u \rangle = \psi_{T,w}(v)$. This unique u is our T^*w . Thus, this T^* always exists and is unique given T .

To calculate T^* based on T , fix the second slot w in $\langle Tv, w \rangle$ and convert $\langle Tv, w \rangle_W$ into the form $\langle v, f(w) \rangle_V$, where the function f of w is our desired T^* .

- As expected, $T^* \in \mathcal{L}(W, V)$, since

$$\begin{aligned} \langle v, T^*(\lambda w_1 + w_2) \rangle &= \langle Tv, \lambda w_1 + w_2 \rangle = \overline{\lambda} \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\ &= \overline{\lambda} \langle v, T^*w_1 \rangle + \langle v, T^*w_2 \rangle \\ &= \langle v, \lambda T^*w_1 + T^*w_2 \rangle. \end{aligned}$$

The above is true for all v , and thus the second slot must be the same, showing that T^* is linear.

- Here are some properties of the adjoint. Move the adjoint from the second slot to the first slot and then back to the second slot will give these results.

- (a) $(S + T)^* = S^* + T^*$;
- (b) $(\lambda T)^* = \overline{\lambda} T^*$;
- (c) $(T^*)^* = T$;
- (d) $I^* = I$;
- (e) $(ST)^* = T^* S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$.

We prove (c) here and omit the others. For any $v \in V$ and $w \in W$,

$$\langle w, (T^*)^*v \rangle = \langle T^*w, v \rangle = \overline{\langle v, T^*w \rangle} = \overline{\langle Tv, w \rangle} = \langle w, Tv \rangle.$$

Because the first slot w is fixed, the second slot must be the same for all v and hence $(T^*)^* = T$.

As you may have noticed from (a), (b), & (c), the map $*$: $\mathcal{L}(V, W) \rightarrow \mathcal{L}(W, V)$ turns out

to be a **conjugate-linear bijective map**. The map is injective because

$$S^* = T^* \implies S = (S^*)^* = (T^*)^* = T;$$

and the map is surjective because for any $S \in \mathcal{L}(W, V)$, we always have $S^* \in \mathcal{L}(V, W)$ such that $(S^*)^* = S$.

- The nullspace and range of the adjoint of T is connected to the orthogonal complement of $\text{null } T$ and $\text{range } T$. For $T \in \mathcal{L}(V, W)$, we have

- (a) $\text{null } T^* = (\text{range } T)^\perp$;
- (b) $\text{range } T = (\text{null } T^*)^\perp$; [take the orthogonal complement of both sides in (a)]
- (c) $\text{null } T = (\text{range } T^*)^\perp$; [replace T^* in (a) by $(T^*)^* = T$]
- (d) $\text{range } T^* = (\text{null } T)^\perp$. [take the orthogonal complement of both sides in (d)]

For (a), $w \in \text{null } T^* \iff T^*(w) = 0 \iff \langle v, T^*(w) \rangle = 0$ for all $v \in V \iff \langle Tv, w \rangle = 0$ for all $v \in V \iff w \in (\text{range } T)^\perp$.

- With respect to an orthonormal basis e_1, \dots, e_n of V and an orthonormal basis f_1, \dots, f_m of W , $\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$ is the conjugate transpose of $\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$. (If we are considering operators $T \in \mathcal{L}(V)$ instead, by default we use the same basis on the domain V and on the range of V .)

The entries in k -th column of $\mathcal{M}(T)$ are exactly the coefficients in

$$\langle Te_k, f_1 \rangle f_1 + \dots + \langle Te_k, f_m \rangle f_m = Te_k.$$

Therefore, $\mathcal{M}(T)_{j,k} = \langle Te_k, f_j \rangle$. Similarly we can conclude that $\mathcal{M}(T^*)_{k,j} = \langle T^* f_j, e_k \rangle = \langle f_j, Te_k \rangle$, where $1 \leq j \leq m$ and $1 \leq k \leq n$. It is clear that $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ (with switched bases) are conjugate transposes of one another.

Note that in the usual case when we are considering an operator T with respect to a single basis, then $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ are simply conjugate transposes of one another (because we do not need to switch bases).

- An operator $T \in \mathcal{L}(V)$ is **self-adjoint** (or **Hermitian**) if $T = T^*$. That is to say, T is self-adjoint iff $\langle Tv, w \rangle = \langle v, Tw \rangle$ for all $v, w \in V$.
 - It follows from definition that the sum of two self-adjoint operators is self-adjoint and the product of a real scalar and a self-adjoint operator is self-adjoint.
 - As the book points out, over \mathbf{C} we can draw a parallel between the adjoint of an operator and the complex conjugation of a number. In some abstract sense saying that an operator is self-adjoint is similar to saying that a number is equal to its conjugation and is thus real. Self-adjoint operators are closely tied to real numbers, as we will see.
- Eigenvalues of self-adjoint operators are real. Consider the self-adjoint operator T on V and

$Tv = \lambda v$ for nonzero $v \in V$, then

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

telling us $\lambda = \bar{\lambda}$.

- (†) For a **complex** inner product space V and $T \in \mathcal{L}(V)$, if $\langle Tv, v \rangle = 0$ for all $v \in V$, then $T = 0$. We prove this via some calculation tricks: for all $u, w \in V$,

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

Therefore, $\langle Tu, w \rangle = 0$ for all $u, w \in V$. Taking $w = Tu$ gives us $T = 0$.

- For operator T on a **complex** inner product spaces V , T is self-adjoint iff for every $v \in V$, $\langle Tv, v \rangle \in \mathbf{R}$.

Let us look at the equation (this is the general trick for proving something is real)

$$\langle Tv, v \rangle - \overline{\langle Tv, v \rangle} = \langle Tv, v \rangle - \langle v, Tv \rangle = \langle (T - T^*)v, v \rangle.$$

If $\langle Tv, v \rangle \in \mathbf{R}$ for every v , then $T - T^* = 0$ by (†). If in the other direction $T = T^*$, then $\langle Tv, v \rangle = \overline{\langle Tv, v \rangle}$ and is thus real for every v .

- If $T = T^*$ on a (real or complex) inner product space V and $\langle Tv, v \rangle = 0$ for all v , then $T = 0$. By (†), now we only have to check this fact on the real inner product space. As one may check,

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

We let $w = Tu$ and get $T = 0$.

- An operator T on V is **normal** if $TT^* = T^*T$. Obviously self-adjoint operators $T = T^*$ are normal.
- T is normal iff $\|Tv\| = \|T^*v\|$ for all $v \in V$.

To prove this, first note that $TT^* - T^*T$ is self-adjoint because $(TT^* - T^*T)^* = (TT^*)^* - (T^*T)^* = TT^* - T^*T$. Therefore,

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

– As a corollary, $Tv = 0 \iff T^*v = 0$, showing that $\text{null } T = \text{null } T^*$.

- All the eigenvalues of $T \in \mathcal{L}(V)$ are exactly the complex conjugates of all the eigenvalues in

T^* . We basically need to prove that λ is an eigenvalue of T iff $\bar{\lambda}$ is an eigenvalue of T^* :

$$\begin{aligned}
\lambda \text{ is not an eigenvalue of } T &\iff T - \lambda I \text{ is injective} \\
&\iff \text{null}(T - \lambda I) = \{0\} \\
&\iff \text{range}(T - \lambda I)^* = \{0\}^\perp = V \\
&\iff (T - \lambda I)^* = T^* - \bar{\lambda}I \text{ is surjective} \\
&\iff \bar{\lambda} \text{ is not an eigenvalue of } T^*.
\end{aligned}$$

- If additionally T is normal, T and T^* share the same set of eigenvectors. Suppose T has eigenvector v with $T^*v = \lambda v$, then T^* has the same eigenvector v with $Tv = \bar{\lambda}v$.

We first check $T - \lambda I$ is normal given T is normal:

$$(T - \lambda I)(T - \lambda I)^* = (T - \lambda I)(T^* - \bar{\lambda}I) = (T^* - \bar{\lambda}I)(T - \lambda I) = (T - \lambda I)^*(T - \lambda I).$$

Then, $0 = \|(T - \lambda I)v\| = \|(T - \lambda I)^*v\| = \|(T^* - \bar{\lambda}I)v\|$, showing that v is an eigenvector of T^* corresponding to the eigenvalue $\bar{\lambda}$.

- Eigenvectors corresponding to distinct eigenvalues of the normal operator T are orthogonal. Consider $Tu = \alpha u$ and $Tv = \beta v$ with $\alpha \neq \beta$. Also, by the last theorem, $T^*v = \bar{\beta}v$. If we now look at $(\alpha - \beta)\langle u, v \rangle$, we have

$$\alpha\langle u, v \rangle - \beta\langle u, v \rangle = \langle Tu, v \rangle - \langle u, T^*v \rangle = 0,$$

telling us that u and v are orthogonal.

7.2 The Spectral Theorem

Now we extend the diagonalizability theorem proved in Chapter 5 from general vector spaces to inner product spaces.

- The complex spectral theorem says that for operator T on an complex inner product space V , the following are equivalent:

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

The equivalency between (b) and (c) is by looking at the diagonal matrix with respect to an orthonormal basis, as we did back in §5.3.

(c) \implies (a): If $\mathcal{M}(T, (e_1, \dots, e_n))$ is diagonal, then $\mathcal{M}(T^*, (e_1, \dots, e_n))$ would be the conjugate transpose $\mathcal{M}(T)$ and is thus still diagonal. Because diagonal matrices commute,

$$\mathcal{M}(TT^*) = \mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T) = \mathcal{M}(T^*T),$$

showing that T is a normal operator.

(a) \implies (c): Given T is normal, by Schur's theorem (since $\mathbf{F} = \mathbf{C}$) we can find an orthonormal basis e_1, \dots, e_n such that

$$\mathcal{M}(T) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{bmatrix}.$$

The idea is to show all the entries above the diagonal are 0. From the first column of $\mathcal{M}(T)$ we have $\|Te_1\|^2 = |a_{11}|^2$, while from the first column of $\mathcal{M}(T^*)$ (the conjugate of the first row of $\mathcal{M}(T)$) we have

$$\|T^*e_1\|^2 = |a_{11}|^2 + \cdots + |a_{1n}|^2.$$

Since $\|Te_1\|^2 = \|T^*e_1\|^2$, a_{12}, \dots, a_{1n} are all 0.

We can proceed in this way to show that a_{23}, \dots, a_{2n} are also all 0 and so on. Therefore, $\mathcal{M}(T)$ is indeed a diagonal matrix, as desired.

Remark. Schur's theorem gives us a shortcut to prove (a) \implies (c) by comparing $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$. By contrast, in the spectral theorem for T on real inner product spaces, (a) \implies (c) is not easy to prove because in the first place we do not know over \mathbf{R} (unlike over \mathbf{C}) whether T has an upper-triangular matrix with respect to some basis of V . Neither do we know whether any eigenvalue exists for $T \in \mathcal{L}(V)$ over \mathbf{R} .

Therefore, a stronger condition on T is required for the similar spectral theorem on real inner product spaces, and as you may guess, T is now required to be self-adjoint in (a). We leave (b) and (c) unchanged in the real spectral theorem. We can show that if T is self-adjoint, then T must have an eigenvalue, from which we can prove that (a) \implies (b).

- Self-adjoint operators has at least an eigenvalue. Recall how we proved operators on complex vector spaces always have an eigenvalue in §5.2. The approach is exactly the same, but we use the polynomial factorization theorem over \mathbf{R} instead of \mathbf{C} .

- We can show that for self-adjoint $T \in \mathcal{L}(V)$, $T^2 + bT + cI$ is injective under $b, c \in \mathbf{R}$, $b^2 - 4c < 0$. For an arbitrary nonzero $v \in V$,

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

showing that $(T^2 + bT + cI)v \neq 0$ for all nonzero vector v and is thus injective.

Once we have shown $T^2 + bT + cI$ is injective, for any $v \in V$,

$$\begin{aligned} 0 &= a_0 + a_1Tv + \cdots + a_nT^n v \\ &= (a_0 + a_1T + a_nT^n)v \\ &= c(T^2 + b_1T + c_1I) \cdots (T^2 + b_MT + c_MI)(T - \lambda_1I) \cdots (T - \lambda_mI)v. \end{aligned}$$

Since the first M terms in the factorization are all injective, one of the λ 's is our desired eigenvalue of T . For the full proof, see this section in the book.

- For self-adjoint $T \in \mathcal{L}(V)$ and the subspace U of V invariant under T , we have

(a) U^\perp is invariant under T ;

Let $v \in U^\perp$, then for all $u \in U$, we have

$$\langle Tv, u \rangle = \langle v, Tu \rangle = 0.$$

Since this holds for all u , $Tv \in U^\perp$ and thus U^\perp is invariant under T .

(b) $T|_U \in \mathcal{L}(U)$ is self-adjoint;

For any $u, v \in U$,

$$\langle T|_U u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, T|_U v \rangle.$$

(c) $T|_{U^\perp} \in \mathcal{L}(U)$ is self-adjoint.

Since U^\perp is invariant from (a), we can repeat what we did in proving (b).

- The real spectral theorem says that for operator T on an real inner product space V , the following are equivalent:

(a) T is self-adjoint.

(b) V has an orthonormal basis consisting of eigenvectors of T .

(c) T has a diagonal matrix with respect to some orthonormal basis of V .

For (b) \implies (c), check out §5.3.

For (c) \implies (a), the diagonal matrix $\mathcal{M}(T)$ is equal to its transpose $\mathcal{M}(T^*)$, meaning that $T = T^*$.

(a) \implies (b) by induction on $\dim V$:

If $\dim V = 1$, then since T has an eigenvalue, we have an eigenvector of norm 1. This eigenvector itself is a basis of V .

Suppose (a) \implies (b) holds for all real inner product spaces with dimension less than $n = \dim V$ ($n > 1$). Given T is self-adjoint, we must have an eigenvector u with norm 1. $U := \text{span}(u)$ is thus invariant under T , and therefore $T|_{U^\perp}$ is self-adjoint, where U^\perp is of dimension $n - 1$.

The inductive hypothesis suggests there is an orthonormal basis of U^\perp consisting of eigenvectors of $T|_{U^\perp}$. Since u has norm 1 and are perpendicular to these eigenvectors of $T|_{U^\perp}$,

the combined list of vectors gives an orthonormal basis of U (as $V = U \oplus U^\perp$), which completes the inductive step.

7.3 Positive Operators and Isometries

- $T \in \mathcal{L}(V)$ is *positive* if T is self-adjoint and $\langle Tv, v \rangle \geq 0$ for all $v \in V$.
 - For example, we can verify that P_U is a positive operator on V .
- An operator R is a *square root* of T if $R^2 = T$.
- For $T \in \mathcal{L}(V)$, the following are equivalent:
 - (a) T is positive;
 - (b) T is self-adjoint and all eigenvalues of T are nonnegative;
 - (c) T has a positive square root;
 - (d) T has a self-adjoint square root;
 - (e) there is an operator R such that $T = R^*R$.

We provide some key steps to the proof below.

(a) \implies (b): T is positive implies T is self-adjoint, which implies it must have an eigenvalue λ . Then for an eigenvector v such that $Tv = \lambda v$,

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

(b) \implies (c): By the spectral theorems we have an orthonormal basis e_1, \dots, e_n corresponding to nonnegative eigenvalues $\lambda_1, \dots, \lambda_n$. Construct $R \in \mathcal{L}(V)$ such that $Re_j = \sqrt{\lambda_j}e_j$ for every e_j . Then $R^2e_j = \lambda e_j = Te_j$, so that $R^2 = T$. Show that R is a positive operator by the definition completes the proof.

(c) \implies (d) \implies (e) is trivial.

(e) \implies (a): $T^* = (R^*R)^* = R^*R = T$, and $\langle Tv, v \rangle = \langle Rv, Rv \rangle \geq 0$.

- Each positive operator T on V has a **unique** positive square root, which we will denote by \sqrt{T} in place of R . From the proof of (b) \implies (c) above, we then have for any eigenvector v and its corresponding eigenvalue λ of T , $\sqrt{\lambda}$ is the eigenvalue of \sqrt{T} corresponding to the same eigenvector v .

The idea is simple, for positive operator T , each eigenvalue v of T has $Tv = \lambda v$ for some nonnegative λ . Once we show $Rv = \sqrt{\lambda}v$ for $R^2 = T$, since the spectral theorems say that V has a basis consisting entirely of these eigenvalues of T , we can show that R is the unique square root.

By the spectral theorem, there is an orthonormal basis e_1, \dots, e_n consisting of eigenvectors of the positive operator R . Let $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ be the corresponding eigenvalues. If we now consider

$$v = a_1e_1 + \dots + a_ne_n,$$

applying R^2 to both sides gives

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

while applying T to both sides gives

$$Tv = a_1\lambda e_1 + \cdots + a_n\lambda e_n.$$

The two expressions above are equal, and since e_1, \dots, e_n is a basis, we have $a_j(\lambda - \lambda_j) = 0$ for all j . Thus,

$$v = \sum_{\{j:\lambda_j=\lambda\}} a_j e_j, \text{ and then } Rv = \sum_{\{j:\lambda_j=\lambda\}} a_j \sqrt{\lambda} e_j = \sqrt{\lambda} v.$$

- An operator $S \in \mathcal{L}(V)$ is an *isometry* if $\|Sv\| = \|v\|$ for all $v \in V$. (an operator that preserves norm)
- There are many equivalent characterizations of isometries. The following are equivalent for $S \in \mathcal{L}(V)$:
 - (a) S is an isometry;
 - (b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$; (*This says that norm-preserving is equivalent to inner-product-preserving.*)
 - (c) Se_1, \dots, Se_n is orthonormal for every orthonormal list of vectors e_1, \dots, e_n in V ;
 - (d) there exists an orthonormal basis e_1, \dots, e_n of V such that Se_1, \dots, Se_n is orthonormal;
 - (e) $S^*S = I$;
 - (f) $SS^* = I$;
 - (g) S^* is an isometry;
 - (h) S is invertible and $S^{-1} = S^*$.

(a) \implies (b) because inner products can be calculate from norms. On real and complex inner product spaces, respectively,

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

(b) \implies (c): replace u and v by e_j and e_k

(c) \implies (d): trivial

(d) \implies (e): For orthonormal basis of e_1, \dots, e_n that makes Se_1, \dots, Se_n orthonormal,

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

By the fact that all $u, v \in V$ can be expressed as a linear combination of these e 's, one can show $\langle S^*Su, v \rangle - \langle u, v \rangle = \langle (S^*S - I)u, v \rangle = 0$. If we set $v = (S^*S - I)u$, we will have

$\|(S^*S - I)u\| = 0$ for all u , and thus $S^*S = I$.

(e) \implies (f): It is not hard to prove the general case that $ST = I$ iff $TS = I$ for operators $S, T \in \mathcal{L}(V)$.

(f) \implies (g) is standard: $\|S^*v\|^2 = \langle SS^*v, v \rangle = \langle Iv, v \rangle = \|v\|^2$.

(g) \implies (h): We can replace S by S^* in (a) and similarly get $SS^* = I$ [from (a) \implies (e)] and $S^*S = I$ [from (a) \implies (f)], showing that S is invertible with $S^{-1} = S^*$.

(h) \implies (a) is again standard: $\|Sv\| = \langle S^*Sv, v \rangle = \langle Iv, v \rangle = \|v\|^2$.

- An isometry on a real (resp. complex) inner product space is often called an *orthogonal* (resp. *unitary*) *operator*.
- Every isometry is normal by (e) and (f), and actually we can extend properties of normal operators to describe properties of isometry. The complex case is given below and the real case will be given later.
- If there is an orthonormal basis e_1, \dots, e_n of V all being eigenvectors of $S \in \mathcal{L}(V)$, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ all having absolute value 1, then S must be an isometry. First, $Sv = \langle v, e_1 \rangle Se_1 + \dots + \langle v, e_n \rangle Se_n = \lambda_1 \langle v, e_1 \rangle e_1 + \dots + \lambda_n \langle v, e_n \rangle e_n$. Since the λ 's all have absolute value 1, $\|Sv\| = |\langle v, e_1 \rangle|^2 + \dots + |\langle v, e_n \rangle|^2 = \|v\|^2$.
- If we let V be a **complex** inner product space, then the above example S is the only type of isometry allowed.

For an isometry $S \in \mathcal{L}(V)$ over \mathbf{C} , since S is normal, by the complex spectral theorem we know that there is an orthonormal basis e_1, \dots, e_n consisting of eigenvectors of S . For every j between 1 and n , the eigenvalue λ_j corresponding to e_j has

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

Therefore, under \mathbf{C} , S is an isometry is equivalent to saying that there is an orthonormal basis of V consisting of eigenvectors of S with corresponding eigenvalues all of absolute value 1.

7.4 Polar Decomposition and Singular Value Decomposition

Remark. In the fourth edition, this section is written as well. SVD will instead be used to prove the polar decomposition theorem.

- The polar decomposition theorem states that for $T \in \mathcal{L}(V)$, we can find an isometry S such that $T = S\sqrt{T^*T}$. (breaking T into a composition of S with $\sqrt{T^*T}$)

Note that T^*T by its form is already a positive operator, so we know it has a positive square root.

There is a good analogy of this theorem with the polar decomposition of a complex number. For nonzero $z \in \mathbf{C}$, we have $z = \left(\frac{z}{|z|}\right) \sqrt{\bar{z}z}$, where the first factor is a point on the unit circle and the second factor resembles T^*T . To prove the theorem, we need several steps:

- *Step I.* Define the function $S_1 : \text{range } \sqrt{T^*T} \rightarrow \text{range } T$ by $S_1(\sqrt{T^*T}v) = Tv$.

What is special about $\sqrt{T^*T}$ is that $\|Tv\| = \|\sqrt{T^*T}v\|$ for all v , and one uses this to prove that the function S_1 is well-defined and injective. One should also verify that S_1 is a linear map and thus a linear isometry. The linear map is surjective by definition.

- *Step II.* Our idea is to extend S_1 on $\text{range } T$ to S on the entire V .

Because $S_1 : \text{range } \sqrt{T^*T} \rightarrow \text{range } T$ is now bijective, $\dim \sqrt{T^*T} = \dim \text{range } T$, from which we know $\dim(\text{range } \sqrt{T^*T})^\perp = \dim(\text{range } T)^\perp$ because both T^*T and T are linear operators on V . Thus we have an orthonormal basis e_1, \dots, e_m of $(\text{range } \sqrt{T^*T})^\perp$ and another orthonormal basis f_1, \dots, f_m of $(\text{range } T)^\perp$. If we let $S_2(e_1) = f_1, \dots, S_2(e_m) = f_m$, the fact that the e 's and the f 's are two orthonormal bases gives us an isometry S_2 .

- *Step III.* Merge S_1 and S_2 into an isometry S on V .

Because v can be uniquely written into as the sum of $u \in \text{range } \sqrt{T^*T}$ and $w \in (\text{range } \sqrt{T^*T})^\perp$, defining

$$Sv = S_1u + S_2w$$

gives $S(\sqrt{T^*T}v) = S_1(\sqrt{T^*T}v) = Tv$ for all $v \in V$, so that $T = S\sqrt{T^*T}$. Furthermore, by the Pythagorean theorem and S_1 and S_2 are isometries themselves, one can show $\|Sv\|^2 = \|v\|^2$ for all v .

- The *singular values* of $T \in \mathcal{L}(V)$ are the eigenvalues of $\sqrt{T^*T}$, which each eigenvalue repeating $\dim E(\lambda, \sqrt{T^*T})$ times.
 - Naturally the singular values should all be nonnegative because $\sqrt{T^*T}$ is a positive operator (the positive square root of the positive operator T^*T).
 - Also, recall that the sum of the dimensions of all eigenspaces is the dimension of the whole space, the number of singular values of $T \in \mathcal{L}(V)$ is equal to $\dim V$.
- The singular values of $T \in \mathcal{L}(V)$ are the nonnegative square roots of the eigenvalues of T^*T , with each of these eigenvalues repeated $\dim E(\lambda, T^*T)$ times.

T^*T is positive, so by the spectral theorems there is an orthonormal basis e_1, \dots, e_n and nonnegative $\lambda_1, \dots, \lambda_n$ such that $T^*Te_j = \lambda_j e_j$ for every j . Thus $\sqrt{T^*T}e_j = \sqrt{\lambda_j}e_j$, which we emphasized in the last section. Note that this establishes a one-to-one correspondence between every λ_j and every $\sqrt{\lambda_j}$, and hence $\dim E(\lambda, T^*T) = \dim E(\lambda, \sqrt{T^*T})$.

- The singular value decomposition theorem says the following: suppose $T \in \mathcal{L}(V)$ has singular values s_1, \dots, s_n , then there exists two orthonormal bases e_1, \dots, e_n and f_1, \dots, f_n of V such that

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

for every $v \in V$.

The spectral theorems applied to the self-adjoint operator $\sqrt{T^*T}$ gives us an orthonormal basis e_1, \dots, e_n such that $Te_j = s_j v_j$ for all j . Since for all v , $v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n$,

applying T to both side leads to

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

Since we have isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$ by the polar decomposition theorem, applying S to both sides of the above equation leads further to

$$S\sqrt{T^*T}v = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n.$$

Because S is an isometry, we know that $f_1 := Se_1, \dots, f_n := Se_n$ is an orthonormal basis of V as well. Therefore, we could simplify the equation above to

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

as desired.

- This theorem introduces a new way to decompose an operator/matrix on an inner product space known as the *singular value decomposition* (SVD). Replace v in the equation above by each individual e_j , and we have $Te_j = s_j f_j$ for every j between 1 and n . Thus we have

$$\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_n)) = \begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{bmatrix},$$

a diagonal matrix representation of T with respect to *two* different bases rather than one basis.

8 Operators on Complex Vector Spaces

Remark. We assume V is a nonzero FDVS in this chapter, and $T \in \mathcal{L}(V)$. (Although some results are possible to be extended to the infinite-dimensional setting, it is not that useful here.) We take n as $\dim V$ by default in addition.

8.1 Generalized Eigenvectors and Nilpotent Operators

- $\{0\} = \text{null } T^0 \subseteq \text{null } T^1 \subseteq \text{null } T^2 \cdots \subseteq \text{null } T^k \subseteq \text{null } T^{k+1} \subseteq \cdots$ holds with a quite obvious proof.
- Furthermore, once $\text{null } T^m = \text{null } T^{m+1}$ for some $m \geq 0$, $\text{null } T^m = \text{null } T^{m+1} = \text{null } T^{m+2} = \cdots$ and so on. To prove this we need to further show that $\text{null } T^{m+k} \supseteq \text{null } T^{m+k+1}$, which is not hard either.
 - As a corollary, if $\text{null } T = 0$, then all powers of T are injective.
 - When do we hit the first m such that $\text{null } T^m = \text{null } T^{m+1}$? This m always exists, and $m \leq \dim V$.
- For T on V with dimension n , we have $\text{null } T^n = \text{null } T^{n+1} = \cdots$.
For the proof, assume that $\text{null } T^n \neq \text{null } T^{n+1}$, then by the previous two claims we have

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

Every proper subset means the dimension of the next one on the right must be one more than the dimension of the previous one on the left. Therefore, at the point $\text{null } T^n$ is at least of dimension n , from which we can increase no further.

From now on we will call

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

the *nullspace power chain*, in which

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

for some k between 0 and n .

- *Note that all the results above hold for the ranges of integer powers of T , with “ \supseteq ” replacing “ \subseteq ” accordingly. See the exercises of this section in the book.*
- Say T is on V with dimension n , $V = \text{null } T^n \oplus \text{range } T^n$. (The claim is actually true for all m such that $T^m = T^{m+1}$.) The proof divides into two steps. First, we need to show that $\text{null } T^n \cap \text{range } T^n = \{0\}$. Second, we need to show that the direct sum is actually V .

Suppose $v \in \text{null } T^n \cap \text{range } T^n$, then $T^n v = 0$ and there exists u such that $T^n u = v$.

Therefore,

$$0 = T^{2n}v = T^n(T^n u) = T^{2n}u,$$

showing that $u \in \text{null } T^{2n} = \text{null } T^n$. Therefore $v = T^n u = 0$.

Now we have $\dim(\text{null } T^n \oplus \text{range } T^n) = \dim \text{null } T^n + \dim \text{range } T^n = \dim V$. Since $\text{null } T^n \oplus \text{range } T^n$ is a subspace of V , the direct sum is V itself.

- We now generalize the concept of eigenvectors so that we will show soon that a complex vector space can be written into a direct sum of generalized eigenspaces. The *generalized eigenvector* of T corresponding to eigenvalue λ is a nonzero vector $v \in V$ such that $(T - \lambda I)^j v = 0$ for some $j \in \mathbf{Z}^+$. (Note that when generalizing eigenvectors, no new eigenvalues are introduced because $T - \lambda I$ being injective implies that $(T - \lambda I)^n$ is injective.)

The *generalized eigenspace* of T corresponding to λ is the set

$$G(\lambda, T) = \{0\} \cup \text{the set of generalized eigenvectors of } T \text{ corresponding to } \lambda.$$

Obviously $E(\lambda, T) \subseteq G(\lambda, T)$. (The E is actually a subspace of G because E is closed under the induced operations from G again induced from V .) When λ is not an eigenvalue, $G(\lambda, T) = \{0\}$.

- We now state that $G(\lambda, T) = \text{null}(T - \lambda I)^n$, where $n = \dim V$. (In particular, $G(0, T) = \text{null } T^n$) You should already have an idea that the proof uses the nullspace power chain.
- A list of generalized eigenvectors corresponding to distinct eigenvalues still needs to be linearly independent. Let $\lambda_1, \dots, \lambda_m$ be the list of eigenvalues and v_1, \dots, v_m be the list of corresponding eigenvectors we are considering. Suppose $0 = a_1 v_1 + \dots + a_m v_m$.

Let k be the largest nonnegative integer such that $w := (T - \lambda_1 I)^k v_1 \neq 0$ (clearly $k < n$). Then

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

which says that $Tw = \lambda_1 w$. Therefore for any λ , $(T - \lambda I)w = (\lambda_1 - \lambda)w$, which implies that

$$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w.$$

Consider the operator $(T - \lambda_1 I)^k (T - \lambda_2 I)^n \dots (T - \lambda_m I)^n$ and apply it the two sides of $0 = a_1 v_1 + \dots + a_m v_m$, then all the terms on the RHS vanish except for the first one:

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

Since w is nonzero and the λ_1 is distinct from the other λ 's, $a_1 = 0$. In the same way, all the coefficients a_i must be zero and thus the eigenvectors are linearly independent.

- A operator N on a nonzero FDVS V is *nilpotent* if $N^k = 0$ for some $k \in \mathbf{Z}^+$.
 - A nilpotent operator cannot be injective/surjective.

Similar to functions on finite sets, the composition of injective (surjective/bijective) linear maps must be injective (surjective/bijective) on a finite dimensional vector space. If N is injective, i.e., $\text{null } N = \{0\}$, then $\text{null } N^k = \{0\} \neq V$.

– The only eigenvalue of a nilpotent operator is 0. Proof is obvious.

- Given nilpotent operator N on V , $N^{\dim V} = 0$, again by the nullspace power chain. (The reverse direction is obviously true as well.)

- $\mathcal{M}(N)$ with respect to some basis of V has the form
$$\begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{bmatrix},$$
 a strictly upper-triangular matrix.

By the nullspace power chain we can extend a basis of $\text{null } N$ to a basis of $\text{null } N^2$, which can be extended to $\text{null } N^3$ and so on. The first few columns concerning N (basis vector of $\text{null } N$) must be all 0. The next few columns concerning N (basis vectors of $\text{null } N^2$) have nonzero entries strictly above the diagonal because

$$N(\text{basis vector of } \text{null } N^2) \in \text{null } N, \text{ since } N^2 v = N(Nv) = 0.$$

Thus, N (basis vector of $\text{null } N^2$) can be written into a linear combination of the previous vectors in $\text{null } N$. Proceeding in this fashion gives us a strictly upper-triangular $\mathcal{M}(N)$.

8.2 Decomposition of an Operator

- We noted before that $\text{null } T$ and $\text{range } T$ are invariant under T . It is true in general that for $p \in \mathcal{P}(\mathbf{F})$, $\text{null } p(T)$ and $\text{range } p(T)$ are invariant under T .

For any $v \in \text{null } p(T)$, $(p(T))(Tv) = T(p(T)v) = T(0) = 0 \implies Tv \in \text{null } p(T)$; for any $v = p(T)u$, $Tv = T(p(T)u) = p(T)(Tu) \in \text{range } p(T)$.

- For operator T on a **complex** vector space V and its distinct eigenvalues $\lambda_1, \dots, \lambda_m$, we have
 - (a) $V = G(\lambda_1, T) \oplus \dots \oplus G(\lambda_m, T)$;
 - (b) $G(\lambda_j, T)$ is invariant under T for all j ;
 - (c) $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent for all j .

For (b), given $p(z) = (z - \lambda_j)^n$, we know $\text{null}(T - \lambda_j I)^n = G(\lambda_j, T)$ is invariant under T . (c) holds straight from definition.

For (a), we prove by induction on $n = \dim V$. If $n = 1$, then all nonzero vectors of V are eigenvectors of V . Thus $V = G(\lambda_1, T)$. For $n > 1$, suppose (a) holds for all complex vector spaces of dimension less than n . Then

$$V = \text{null}(T - \lambda_1 I)^n \oplus \text{range}(T - \lambda_1 I)^n = G(\lambda_1, T) \oplus U, \quad (3)$$

if we let $U = \text{range}(T - \lambda_1 I)^n$. This U is invariant under T by applying the last theorem to $p(z) = (z - \lambda_1)^n$. Since $G(\lambda_1, T) \neq \{0\}$, we have $\dim U < n$, so that we can (a) is true in

$T|_U \in \mathcal{L}(U)$:

$$U = G(\lambda_1, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U).$$

Since $G(\lambda_1, T)$ contains all generalized eigenvectors corresponding to λ_1 by definition, the eigenvectors of $T|_U$ can only correspond to $\lambda_2, \dots, \lambda_m$.

We now intend to prove that $G(\lambda_k, T|_U) = G(\lambda_k, T)$ for all k between 2 and m , i.e., U contains all the generalized eigenvectors of T corresponding to λ_k . “ \subseteq ” is clear because, and we prove “ \supseteq ”: suppose $v \in G(\lambda_k, T) \subseteq U$, then by (3) above, $v = v_1 + u$ for some $v_1 \in G(\lambda_1, T)$ and $u \in U$. Furthermore, u can be broken up into $v_2 + \cdots + v_m$ for each $v_j \in G(\lambda_j, T|_U) \subseteq G(\lambda_j, T)$ [which we proved just now], so that $v = v_1 + v_2 + \cdots + v_m$. Because generalized eigenvectors corresponding to distinct eigenvalues are linearly independent, all the v_1 to v_m on the RHS must be 0 except for v_k . In particular $v_1 = 0$, so that $v = u \in U$. Therefore, $v \in G(\lambda_k, T)$, as desired.

- Note that if T is diagonalizable, since $E(\lambda_j, T) \subseteq G(\lambda_j, T)$, comparing the (a) here and $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ gives us $E(\lambda_j, T) = G(\lambda_j, T)$ for all j , or otherwise the G ’s would not give a direct sum.
- Because all the generalized eigenspaces $G(\lambda_j, T)$ are vector spaces, we can choose a basis from each of $G(\lambda_1, T)$ to $G(\lambda_m, T)$. Put all these bases together and by (a), we have a basis of V consisting of generalized eigenvectors of T .
- For an eigenvalue λ of $T \in \mathcal{L}(V)$, its *algebraic multiplicity* is defined as $\dim G(\lambda, T) = \dim \text{null}(T - \lambda I)^{\dim V}$, and its *geometric multiplicity* is defined as $\dim E(\lambda, T) = \dim \text{null}(T - \lambda I)$. By default, when the book says “multiplicity” it means specifically the algebraic multiplicity.
 - Since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ for operator T on a **complex** vector space V , $\dim V$ is the sum of multiplicities of all eigenvalues of T .

- A *block diagonal matrix*, as its name tells, is a matrix of the form
$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$
 where the A_i ’s are square matrices on the diagonal and all other entries outside these A_j ’s are 0. The block diagonal matrix is often related to the direct sum of invariant subspaces, as its form indicates.

- We give the block diagonal matrix (with upper triangular blocks) decomposition theorem. Suppose T is an operator over a complex vector space V . For the list of all distinct eigenvalues $\lambda_1, \dots, \lambda_m$ of T , each with corresponding multiplicity d_1, \dots, d_m . Then there exists a basis of V with respect to which T has a block diagonal matrix of the form
$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$
 with each

A_j being upper-triangular and having λ_j on all diagonal entries, i.e., $A_j = \begin{bmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{bmatrix}$.

Since each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent, its matrix with respect to some basis of $G(\lambda_j, T)$ is strictly upper-triangular. Therefore, $(T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$ is in the form of each A_j . Combining all these bases together gives a block diagonal matrix $\mathcal{M}(T)$ described above.

- As a preliminary lemma to the upcoming theorem, for nilpotent operator $N \in \mathcal{L}(V)$, $I + N$ has a square root.

The Taylor expansion of $\sqrt{1+x}$ looks like

$$1 + a_1 x + a_2 x^2 + \cdots,$$

and we guess that the square root of $I + N$ should look like $I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}$, in which the higher order terms vanish because $N^m = 0$ for some m (N is nilpotent). We square this expression:

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

We can let $2a_1 = 1$, $2a_2 + a_1^2 = 0$, $2a_3 + 2a_1 a_2 = 0$, and so on so that all the coefficients of the expression above are 0 (because the value of a_1 gives the value of a_2 , and the value of a_1, a_2 gives the value of a_3, \dots). In this manner we have found a square root of $I + N$: $I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}$.

- For complex vector space V and invertible $T \in \mathcal{L}(V)$, T must have a square root.

First, we look at nilpotent operators that arises from each generalized eigenspaces:

$$N_j := T|_{G(\lambda_j, T)} - \lambda_j I|_{G(\lambda_j, T)}.$$

Then, since T is invertible, none of the λ_j 's can be 0, so $T|_{G(\lambda_j, T)} = \lambda_j (\frac{N_j}{\lambda_j} + I|_{G(\lambda_j, T)})$. Clearly by the lemma above, the square root of $\frac{N_j}{\lambda_j} + I|_{G(\lambda_j, T)}$ times the square root of the complex number λ_j gives a square root R_j of $T|_{G(\lambda_j, T)}$.

We know $v \in V$ can always be written as a unique sum of $u_1 + \cdots + u_m$ with each u_j in $G(\lambda_j, T)$. Define R such that

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

for all $v \in V$, which we can verify is a square root operator of T on V , as follows.

Note that $R_1 u_1$ is still in $G(\lambda_1, T)$, $R(R_1 u_1) = R_1(R_1 u_1) = (R_1)^2 u_1$. Taking this as an

example, we have

$$Rv = (R_1)^2 u_1 + \cdots + (R_m)^2 u_m = T(u_1 + \cdots + u_m) = Tv,$$

as desired.

8.3 Characteristic and Minimal Polynomials

Remark. By default in this section, q is the characteristic polynomial, and p is the minimal polynomial.

- For an operator T over a complex vector space V , let $\lambda_1, \dots, \lambda_m$ be all distinct eigenvalues of T , each with corresponding multiplicity d_1, \dots, d_m . Then the *characteristic polynomial* of T is defined to be

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

- The characteristic polynomial must be of degree $\dim V$ because $d_1 + \cdots + d_m = \dim V$.
- The zeros of the characteristic polynomial are the eigenvalues of T .
- Now we prove the important **Cayley-Hamilton theorem** on complex vector spaces. With the tools of generalized eigenspaces we introduced in the last two sections, the proof becomes straightforward. Let T be an operator on a complex vector space V . Let q be the characteristic polynomial of T , then $q(T) = 0$.

Let $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and d_1, \dots, d_m be the corresponding dimensions of the generalized eigenspaces $G(\lambda_j, T)$, i.e., the multiplicities of these eigenvalues. We know the restricted operator $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent, and thus it raised to the power of $\dim G(\lambda_j, T)$ is the zero operator: $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$.

Since every vector in V is a sum of vectors in the individual generalized eigenspaces, it suffices to prove that each $q(T)|_{G(\lambda_j, T)} = 0$. Consider $q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}$, in which the factor on the RHS are commutative. For any chosen j between 1 and m , we may move the factor $(T - \lambda_j I)^{d_j}$ to the very right. Since for all $v \in G(\lambda_j, T)$, $(T - \lambda_j I)^{d_j} v = 0$, $q(T)|_{G(\lambda_j, T)}$ is the zero operator, as desired.

- A *monic polynomial* is polynomial with highest-term coefficient being 1.
- For $T \in \mathcal{L}(V)$, then there exists a unique monic polynomial p of smallest degree such that $p(T) = 0$. We call it the *minimal polynomial* of the operator T . This definition and the forthcoming theorems apply to operators on both real and complex vector spaces.

Note $\dim \mathcal{L}(V) = \dim V \times \dim V = n^2$, if we define $n = \dim V$. Thus, the list of powers of T of length $n + 1$

$$I, T, T^2, \dots, T^{n^2}$$

is linearly dependent. By the well-ordering principle applied to the linear dependence lemma, there exists a least m such that

$$I, T, T^2, \dots, T^{m-1}$$

is linearly independent, while T^m is in its span. Therefore, we have

$$a_0I + a_1T + \cdots + a_{m-1}T^{m-1} + T^m = 0$$

for some $a_0, \dots, a_{m-1} \in \mathbf{F}$. Accordingly we define $p(z) = a_0 + a_1z + \cdots + a_{m-1}z^{m-1} + z^m \in \mathcal{P}(\mathbf{F})$, which is a monic polynomial of the smallest possible degree such that $p(T) = 0$.

For the uniqueness part, by the choice of m if we have another minimal polynomial q , then $\deg q = m$ as well. Consider $(p - q)(z)$, which is of degree less than m and has $(p - q)(T) = p(T) - q(T) = 0$. It follows that $p = q$ because otherwise p cannot be the minimal polynomial.

- We remark that the Cayley-Hamilton theorem tells us that the minimal polynomial has to be of degree at most $\dim V$, an upper bound stronger than $(\dim V)^2 = \dim \mathcal{L}(V)$, over \mathbf{C} . This holds over \mathbf{R} as well, as we will see in the next chapter.
- Keep in mind that the minimal polynomial has *real coefficients when V is a real vector space, and has complex coefficients when V is a complex vector space.*
- $q(T) = 0$ iff q is a polynomial multiple of the minimal polynomial p of T (i.e., there exists $s \in \mathcal{P}(\mathbf{F})$ such that $q = ps$).

“ \Leftarrow ” is obvious. For “ \Rightarrow ”, suppose we have $q(z)$ such that $q(T) = 0$, then by the division algorithm with respect to the minimal polynomial $p(z)$, we have $q = ps + r$ for a (unique) pair of s and r with $\deg r < \deg p$. Then

$$0 = q(T) = p(T)s(T) + r(T) = r(T).$$

If r is not the zero polynomial, then we may divide r by its highest coefficient to get a monic polynomial \tilde{r} such that $\tilde{r}(T) = 0$ still. Yet $\deg \tilde{r} < \deg p$, which leads to contradiction. Therefore, $r = 0$, and we have $q = ps$.

- It follows directly that for an operator T on a **complex** vector space, the characteristic polynomial must be a polynomial multiple of the minimal polynomial. (We have not defined the characteristic polynomials in the case $\mathbf{F} = \mathbf{R}$, but the result is similarly true. See the next chapter.) This is an important criterion for us to find the minimal polynomial of an operator.
- The following result is another important criterion for finding the minimal polynomial. For an operator T on a real/complex vector space V , the zeros of the minimal polynomial p (over \mathbf{R}/\mathbf{C}) are **precisely** the eigenvalues of T .

Suppose $\lambda \in \mathbf{F}$ is a zero of p , so that $p(z) = (z - \lambda)q(z)$ for a monic polynomial q (because p is monic). Replace z by T , and we have

$$0 = p(T) = (T - \lambda I)q(T).$$

Since $\deg q < \deg p$, for some $v \in V$ $q(T)v \neq 0$ so that q is not a minimal polynomial. Therefore, $T - \lambda I$ is not injective, which means that λ is an eigenvalue of T .

For the other direction, suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T instead, then $\exists v \in V \setminus \{0\}$ such that $Tv = \lambda v$. We have used a number of times already that $T^j v = \lambda^j v$ for arbitrary integer exponent j . Thus

$$\begin{aligned} 0 &= p(T)v = (a_0I + a_1T + \cdots + a_{m-1}T^{m-1} + T^m)v \\ &= (a_0 + a_1\lambda + \cdots + a_{m-1}\lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v. \end{aligned}$$

Since $v \neq 0$, $p(\lambda) = 0$, as desired.

8.4 Jordan Form

- Let $N \in \mathcal{L}(V)$ be nilpotent. Then there exist vectors $v_1, \dots, v_n \in V$ and nonnegative integers m_1, \dots, m_n such that

- (a) $N^{m_1}v_1, \dots, Nv_1, v_1,$
 $N^{m_2}v_2, \dots, Nv_2, v_2,$
 \dots
 $N^{m_n}v_n, \dots, Nv_n, v_n$ is a basis of V ;
- (b) $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$.

Induction on $\dim V$: if $\dim V = 1$, the only nilpotent operator is the 0 operator. We may take v_1 to be any nonzero vector in V and $m_1 = 0$.

Assume $\dim V > 1$ and the claim holds on all vector spaces of smaller dimensions. Since N is nilpotent, N is not surjective and thus $\text{range } N$ is proper subspace of V with dimension smaller than V . We may then use the induction hypothesis on $N|_{\text{range } N} \in \mathcal{L}(\text{range } N)$. [In the trivial case $\text{range } N = \{0\}$, N is just 0 operator, and we may choose v_1, \dots, v_n to be any basis of V and let $m_1, \dots, m_n = 0$.]

By the induction hypothesis we have $v_1, \dots, v_n \in \text{range } N$ and nonnegative integers m_1, \dots, m_n such that

$$\begin{aligned} &N^{m_1}v_1, \dots, Nv_1, v_1, \\ &N^{m_2}v_2, \dots, Nv_2, v_2, \\ &\dots \\ &N^{m_n}v_n, \dots, Nv_n, v_n \end{aligned} \tag{4}$$

is a basis of $\text{range } N$, and

$$N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0. \tag{5}$$

By $v_j \in \text{range } N$ we know there exists $u_j \in V$ such that $v_j = Nu_j$. It follows that $N^{k+1}u_j =$

$N^k v_j$ for each j and each nonnegative k .

Now we prove that

$$\begin{aligned} & N^{m_1+1}u_1, \dots, Nu_1, u_1, \\ & \dots \\ & N^{m_n+1}u_n, \dots, Nu_n, u_n \end{aligned} \tag{6}$$

is linearly independent in V . Consider a linear combination of the vectors in the list above and apply N to it. By $N^{k+1}u_j = N^k v_j$ it is easy to see that the result is a linear combination of (4) and $N^{m_1+1}v_1, \dots, N^{m_n+1}v_n$, with coefficients left unchanged. Therefore by (5) the linear combination of terms in (4) and of $N^{m_1}v_1, \dots, N^{m_n}v_n$ must both be 0. Since (4) is a basis of $\text{range } N$, all these coefficients must be 0. Thus the list is linearly independent, and we can extend it to a basis

$$\begin{aligned} & N^{m_1+1}u_1, \dots, Nu_1, u_1, \\ & \dots \\ & N^{m_n+1}u_n, \dots, Nu_n, u_n \\ & w_1, \dots, w_p \end{aligned} \tag{7}$$

of V .

Note that $Nw_j \in \text{range } N = \text{the span of (4)}$, and each vector in (4) is N applied to a vector in (6). Therefore, there is a vector x_j in the span of (6) such that $Nx_j = Nw_j$ for every j .

Now let $u_{n+j} = w_j - x_j$, which follows that $Nu_{n+j} = 0$. Meanwhile

$$\begin{aligned} & N^{m_1+1}u_1, \dots, Nu_1, u_1, \\ & \dots \\ & N^{m_n+1}u_n, \dots, Nu_n, u_n, \\ & u_{n+1}, \dots, u_{n+p} \end{aligned}$$

spans V because each x_j and u_{n+j} is contained in its span, meaning that each w_j is in its span. Since the length of the list is same as (7), it is the desired basis for V , which finishes the induction.

- For $T \in \mathcal{L}(V)$, a basis of V is called a *Jordan basis for T* if with respect to this basis $\mathcal{M}(T)$ is a block diagonal matrix

$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{bmatrix},$$

where each block A_j is an upper-triangular matrix of the form

$$A_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}.$$

From the end of §5.2 we know that all the λ_j 's are precisely the eigenvalues of T because $\mathcal{M}(T)$ is upper-triangular. The λ_j 's do not have to be distinct.

- Existence of Jordan Basis: for an operator T on a complex vector space V , there is a basis of V that is a Jordan basis for T .

Consider a nilpotent operator N and $v_1, \dots, v_n \in V$ as given by the previous theorem. What is the special about the result we just proved is that for each j , N sends the first vector in the list $N^{m_j}v_j, \dots, Nv_j, v_j$ to 0 and the rest to the previous vector. In terms of matrix, the previous theorem gives a basis of V with respect to which $\mathcal{M}(N)$ is block diagonal, with each diagonal blocks of the form

$$\begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}.$$

Thus Jordan bases exist for nilpotent operators.

In general for $T \in \mathcal{L}(V)$, we know

$$V = G(\lambda_1, T) \oplus \dots \oplus G(\lambda_m, T),$$

where the λ_j 's are distinct eigenvalues of T . In §8.2 we proved that each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. With respect to the basis described in the last paragraph we have

$$\mathcal{M}(T|_{G(\lambda_j, T)}) = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}.$$

Therefore, adjoining the described bases of each nilpotent $(T - \lambda_j I)|_{G(\lambda_j, T)}$ gives us the Jordan basis for T , with respect to which $\mathcal{M}(T)$ has the *Jordan canonical form*.

9 Operators on Real Vector Spaces

Remark. We continue to assume V is a nonzero FDVS in this chapter, and $T \in \mathcal{L}(V)$.

9.1 Complexification

Remark. In this section, u, v by default denotes vectors in V , and a, b are real numbers to represent an arbitrary complex number $a + bi$. V is a nonzero finite-dimensional real vector space.

- For a real vector space V , its **complexification**, denote by $V_{\mathbf{C}}$, is $V \times V$. Customarily we write $(u, v) \in V_{\mathbf{C}}$ as $u + iv$, although it sometimes makes more sense to

(1) Addition is defined entrywise, i.e.,

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2),$$

or equivalently, $(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$.

- (2) **Complex** scalar multiplication is defined by seeing $u + iv \in V_{\mathbf{C}}$ as a complex number, i.e.,

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbf{R}$, or equivalently, $(a + bi)(u, v) = (au - bv, av + bu)$.

We can show that $V_{\mathbf{C}}$ is a complex vector space. Clearly it is abelian group under addition. The multiplicative identity is $1 + 0i$, and because complex number multiplication follows the distributive and associative properties, scalar multiplication defined on $V_{\mathbf{C}}$ here also follows these properties. Since now $V_{\mathbf{C}}$ is a vector space, the original real vector space V is naturally embedded into $V_{\mathbf{C}}$: each $u \in V$ is naturally identified with each $u + 0i \in V_{\mathbf{C}}$.

- For a real vector space V , if v_1, \dots, v_n is its basis, then the list is also a basis of the complex vector space $V_{\mathbf{C}}$. It follows directly that $\dim V_{\mathbf{C}} = \dim V$.

Note that by the definition of scalar multiplication, $(0 + 1i)(u + i0) = 0 + iu$ for any $u \in V$. Therefore, $\text{span}(v_1, \dots, v_n)$ contains all the vectors v_1, \dots, v_n and iv_1, \dots, iv_n , so that all $u + iv \in V_{\mathbf{C}}$ can be expressed.

It might make more sense to think about this in terms of the (\cdot, \cdot) notation. The list of vectors $(v_1, 0), \dots, (v_n, 0)$ spans the first slot, while the list $(0, v_1), \dots, (0, v_n)$ spans the second slot in $V \times V$.

Regarding the linear independence of v_1, \dots, v_n , suppose $a_1v_1 + \dots + a_nv_n = 0$ with coefficients from \mathbf{C} , then

$$(\text{Re } a_1)v_1 + \dots + (\text{Re } a_n)v_n = 0 \quad \text{and} \quad (\text{Im } a_1)v_1 + \dots + (\text{Im } a_n)v_n = 0.$$

Since the list of v_i 's are linearly independent on the real vector space V , the Re 's and Im 's above are all zero, and thus the a_i 's are all zero, finishing the proof.

- $T_{\mathbf{C}}$, the *complexification of an operator* T on a real vector space V , is an operator on the complex vector space $V_{\mathbf{C}}$ given by $T_{\mathbf{C}}(u + iv) = Tu + iTv$. In particular, $T_{\mathbf{C}}u = Tu$ for $u \in V$.

Now we show that given $T \in \mathcal{L}(V)$, $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$: first,

$$\begin{aligned} T_{\mathbf{C}}((u_1 + iv_1) + (u_2 + iv_2)) &= T_{\mathbf{C}}((u_1 + u_2) + i(v_1 + v_2)) \\ &= T(u_1 + u_2) + iT(v_1 + v_2) \\ &= (Tu_1 + iTv_1) + (Tu_2 + iTv_2) \\ &= T_{\mathbf{C}}(u_1 + iv_1) + T_{\mathbf{C}}(u_2 + iv_2); \end{aligned}$$

and second, given $a + bi \in \mathbf{C}$ (since $V_{\mathbf{C}}$ is a **complex** vector space),

$$\begin{aligned} T_{\mathbf{C}}((a + bi)(u + iv)) &= T_{\mathbf{C}}(au - bv + i(bu + av)) \\ &= T(au - bv) + iT(bu + av) \\ &= (a + bi)Tu + (ia - b)Tv \\ &= (a + bi)Tu + (a + bi)iTv \\ &= (a + bi)T_{\mathbf{C}}(u + iv). \end{aligned}$$

- For T on a real vector space V with basis v_1, \dots, v_n , $\mathcal{M}(T) = \mathcal{M}(T_{\mathbf{C}})$ with respect to the same basis v_1, \dots, v_n , because $T_{\mathbf{C}}v_i = Tv_i$ for all i between 1 and n . *This gives us the insight that $T_{\mathbf{C}}$ can be thought as the complete extension of T from V over \mathbf{R} to the complexified space $V \times V$ over \mathbf{C} , so that the matrix of T is still preserved.*
- We mentioned that a real vector space can have no eigenvalues at all and therefore no invariant subspace of dimension 1. However, we can conclude that an invariant subspace of dimension 1 or 2 always exists in a finite-dimensional real vector space. The proof follows quite mystically from the trick of complexification.

For an operator T on a real vector space V , its complexification $T_{\mathbf{C}}$ on the complex vector space $V_{\mathbf{C}}$ must have an eigenvalue $a + bi$. We can simplify $T_{\mathbf{C}}(u + iv) = (a + bi)(u + iv)$ into

$$Tu + iTv = (au - bv) + (av + bu)i,$$

from which we get $Tu = au - bv$ and $Tv = av + bu$. Let U be the span of $\{u, v\}$ in V , which is clearly invariant under T with dimension ≤ 2 .

- Clearly by repeated application of $T_{\mathbf{C}}$ to $u + iv$ we have $(T_{\mathbf{C}})^n(u + iv) = T^n u + iT^n v$. It directly follows that $p(T_{\mathbf{C}}) = (p(T))_{\mathbf{C}}$.
- The minimal polynomial of $T_{\mathbf{C}}$ is equal to the minimal polynomial of T on a real vector space. We use $p \in \mathcal{P}(\mathbf{R})$ to denote the minimal polynomial of T .

First from $p(T) = 0$ we immediately know that $p(T_{\mathbf{C}}) = 0$.

We now show that p is indeed the *minimal* polynomial of $T_{\mathbf{C}}$. Suppose $q \in \mathcal{P}(\mathbf{C})$ is a monic

polynomial such that $q(T_{\mathbf{C}}) = 0$. Then for any $u \in V$ (not in $V_{\mathbf{C}}$), $q(T_{\mathbf{C}})u = 0$, i.e., given

$$q(z) = (a_0 + b_0) + (a_1 + b_1)z + \cdots + (a_{m-1} + b_{m-1})z^{m-1} + z^m,$$

we have

$$\begin{aligned} 0 = q(T_{\mathbf{C}})u &= (a_0 + b_0i)u + (a_1 + b_1i)Tu + \cdots + (a_{m-1} + b_{m-1}i)T^{m-1}u + T^m u \\ &\quad \text{(equal to } q(T)u) \\ &= (a_0 + a_1T + \cdots + a_{m-1}T^{m-1} + T^m)u + i(b_0 + b_1T + \cdots + b_{m-1}T^{m-1})u \quad (\dagger) \end{aligned}$$

Take $r(z) = a_0 + a_1z + \cdots + a_{m-1}z^{m-1} + z^m$. Then r is a monic polynomial in $\mathcal{P}(\mathbf{R})$ with $r(T)u = 0$ for all $u \in V$. Thus $\deg q = m = \deg r \geq \deg p$, showing that p is of the minimum degree and thus the *unique* minimal polynomial of $V_{\mathbf{C}}$.

- We give another more involved way of thinking about the proof, but it helps you understand the theorem. Consider q instead as the the minimal polynomial of $T_{\mathbf{C}}$. Again restrict $q(T_{\mathbf{C}})$ to V and look at (\dagger) . If $b_0 + b_1z + \cdots + b_{m-1}z^{m-1}$ is not a zero polynomial, then we can divide it by its highest-degree coefficient to get a monic polynomial $s(z)$ such that $s(T)u = 0$ for all $u \in V$, with degree $< m$. Therefore, q must be of real coefficients [$q(z) \in \mathcal{P}(\mathbf{R})$]. Since we know p is the minimal polynomial of T on V , q must be just p because we have showed that p is a monic polynomial such that $p(T_{\mathbf{C}}) = 0$.
- $\lambda \in \mathbf{R}$ is an eigenvalue of T on a real vector space V iff it is an eigenvalue of $T_{\mathbf{C}}$.
The (real) eigenvalues of T are the (real) zeros of the minimal polynomial of T , which are the real zeros of the minimal polynomial of $T_{\mathbf{C}}$ and thus the real eigenvalues of $T_{\mathbf{C}}$. Proof straight from the definition is also viable.
- For $\lambda \in \mathbf{C}$ and any nonnegative j , for T on a real vector space V ,

$$(T_{\mathbf{C}} - \lambda I)^j(u + iv) = 0 \iff (T_{\mathbf{C}} - \bar{\lambda} I)^j(u - iv) = 0.$$

WLOG we only need to prove “ \implies ”. Proof is based on induction and some calculation tricks, which you can refer to the book.

- Take $j = 1$, and it follows directly that the eigenvalues of $T_{\mathbf{C}}$ come in pairs: if $\lambda \in \mathbf{C}$ is one, then $\bar{\lambda}$ is also one.
- The multiplicity of $\lambda \in \mathbf{C}$ is equal to the multiplicity of $\bar{\lambda}$ (both considered as an eigenvalue of $T_{\mathbf{C}}$). Suppose $u_1 + iv_1, \dots, u_m + iv_m$ is a basis of $G(\lambda, T_{\mathbf{C}})$, then $u_1 - iv_1, \dots, u_m - iv_m$ is a basis of $G(\bar{\lambda}, T_{\mathbf{C}})$. Both generalized eigenspaces are of the same dimension, and thus λ and $\bar{\lambda}$ have the same multiplicities.
- An operator T on an odd-dimension real vector space V has an eigenvalue. Because nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs of of 2, and each eigenvalue in such a pair share the same multiplicity, the sum of multiplicities of the nonreal eigenvalues is even.

However, the sum of multiplicities of all eigenvalues is $\dim V_{\mathbf{C}} = \dim V$, which is odd, there must be a real eigenvalue of $T_{\mathbf{C}}$, which is an eigenvalue of T .

Remark. We defined in §8.3 the minimal polynomial altogether for operators on both real and complex vector spaces, because the whether \mathbf{F} is the real or the complex field does not make any difference for the properties and theorems about the minimal polynomial. For example, the coefficients of the minimal polynomial for an operator on real (resp. complex) vector space must be real (resp. complex). But this was not the case for the characteristic polynomial. For example, we know $\dim V$ is equal to the sum of multiplicities of all eigenvalues of an operator T on a complex vector space, but not on a real vector space. However, with the tool of complexification we can now define the characteristic polynomial for operators on real vector spaces, and make sense of the theorems we have proved about characteristic polynomials in the context of real vector spaces.

- For T on a real vector space V , the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real.

We know the multiplicities of nonreal eigenvalues of $T_{\mathbf{C}}$ in a conjugate pair are the same. Thus we can the characteristic polynomial of $T_{\mathbf{C}}$ should include the following:

$$(z - \lambda)^m(z - \bar{\lambda})^m = (z^2 - 2(\operatorname{Re} \lambda)z + |\lambda|^2),$$

which is of real coefficients. The characteristic polynomial of $T_{\mathbf{C}}$ should therefore be of the product of expressions above with expressions $(z - t)^d$, where t is a real eigenvalue of $T_{\mathbf{C}}$ and d is its multiplicity. The coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ should therefore be all real,

- and we define the characteristic polynomial of $T_{\mathbf{C}}$ to be the *characteristic polynomial* of T on a **real** vector space V . Properties to be noted include:
 - (a) the coefficients of it are all real by definition;
 - (b) it has degree $\dim V = \dim V_{\mathbf{C}}$;
 - (c) its **real zeros** are precisely the eigenvalues of T , because the (real) zeros of $T_{\mathbf{C}}$ are precisely the (real) eigenvalues of $T_{\mathbf{C}}$, which are precisely the eigenvalues of T .
- The **Cayley-Hamilton theorem** now holds on real vector spaces: for T on either a real/complex vector space V and the characteristic polynomial q of T , $q(T) = 0$.

Since we have $q(T_{\mathbf{C}}) = 0$, by $q(T_{\mathbf{C}}) = (q(T))_{\mathbf{C}}$, we get $q(T) = 0$.

- Also, we can conclude in general for T either on a real/complex vector space V , let p be the minimal polynomial and q be the characteristic polynomial of T , then
 - (a) $\deg p \leq \dim V$; (since $\deg q = \dim V$)
 - (b) q is a polynomial multiple of p . (since $q(T) = 0$ by Cayley-Hamilton)

9.2 Operators on Real Inner Product Spaces

Remark. The real spectral theorem that we proved in §7.2 offers a complete description of self-adjoint (but not normal) operators on real vector spaces. And in §7.3, we gave a completed description of isometries only on complex (but not real) vector spaces. The following section will extend these results with the tools we now have.

- For T on a real inner product space V with dimension 2, the following are equivalent:

- (a) T is normal but not self-adjoint;
- (b) The matrix of T with respect to any orthonormal basis of V is of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $b \neq 0$.
- (c) The matrix of T with respect to some orthonormal basis of V is of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $b > 0$.

(a) \implies (b): Suppose T is normal but not self-adjoint, then let e_1, e_2 be an orthonormal basis of V , with respect to which $\mathcal{M}(T, (e_1, e_2)) = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$, and its transpose $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathcal{M}(T^*, (e_1, e_2))$. Thus $Te_1 = ae_1 + be_2$ and $T^*e_1 = ae_1 + ce_2$, from which we have

$$\|Te_1\| = a^2 + b^2 \quad \text{and} \quad \|T^*e_1\| = a^2 + c^2.$$

Since T is normal, the two norms above should be equal, and $b = c$ or $-c$. If $b = c$ then $\mathcal{M}(T) = \mathcal{M}(T^*)$, so that T is self-adjoint. Thus $\mathcal{M}(T, (e_1, e_2)) = \begin{bmatrix} a & -b \\ b & d \end{bmatrix}$.

Since T is normal, $\mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T)$:

$$\begin{bmatrix} a & -b \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ -b & d \end{bmatrix} = \begin{bmatrix} a & b \\ -b & d \end{bmatrix} \begin{bmatrix} a & -b \\ b & d \end{bmatrix}. \quad (*)$$

Then we have $ab - bd = -ab + bd$. Note that if $b = 0$ then T is self-adjoint, so $b \neq 0$. Thus $a - d = -a + d$, which gives $a = d$.

(b) \implies (c) is trivial. Suppose $\mathcal{M}(T) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with respect to an orthonormal basis e_1, e_2 . Since $b \neq 0$, one of b and $-b$ is greater than 0. If $b > 0$, the proof is finished; if not, then then consider $\mathcal{M}(T, (e_1, -e_2)) = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$. Since $-b > 0$, this completes the proof.

(c) \implies (a): Since $b \neq 0$, the matrix is not equal to its transpose and hence T is not self-adjoint. Because $(*)$ is satisfied by $d = a$, the matrices of T and T^* commute. Thus T is normal.

- We now prove a lemma that will be useful in describing all normal operators on real inner product spaces. (Recall in §7.2 we had to show a similar result in order to prove the real

spectral theorem.) Suppose T is a normal operator on an inner product space V (real or complex), and U is a subspace of V that is invariant under T , then

(a) U^\perp is invariant under T .

Let e_1, \dots, e_m be an orthonormal basis of U , from which we can extend to a basis of V : $e_1, \dots, e_m, f_1, \dots, f_n$. Since $V = U \oplus U^\perp$, by a claim we mentioned in Chapter 2, U^\perp has f_1, \dots, f_n as a basis.

Because U is invariant under T , we should have

$$\mathcal{M}(T, (e_1, \dots, e_m, f_1, \dots, f_n)) = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix},$$

where A is a $m \times m$ matrix. Because the e_i 's are orthonormal,

$$\sum_{i=1}^m \|Te_j\|^2 = \text{sum of the squares of the absolute values of all entries of } A.$$

Similar to what we have done in our last theorem, if we look at the matrix of T^* ,

$$\sum_{i=1}^m \|T^*e_j\|^2 = \text{sum of the squares of the absolute values of all entries of } A \text{ and } B.$$

Equating the two formulas above (by T is normal) tells us that B is a zero matrix. This shows that U^\perp is invariant under T (because all Tf_j 's are in the span of the basis vectors of U^\perp).

(b) U is invariant under T^* .

This is easy. Since

$$\mathcal{M}(T, (e_1, \dots, e_m, f_1, \dots, f_n)) = \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix},$$

we have

$$\mathcal{M}(T^*, (e_1, \dots, e_m, f_1, \dots, f_n)) = \begin{bmatrix} \overline{A^t} & 0 \\ 0 & \overline{C^t} \end{bmatrix},$$

(where $\overline{\cdot}^t$ is our notation for the conjugate transpose of a matrix). Thus U is invariant under T^* .

(c) $(T|_U)^* = T^*|_U$.

Set $S = (T|_U) \in \mathcal{L}(U)$, and we want to show $S^*v = T^*v$ for all $v \in U$. Fix $u \in U$, and for all $v \in U$ we have $\langle u, S^*v \rangle = \langle Su, v \rangle = \langle Tu, v \rangle = \langle u, T^*v \rangle$ (here we consider $\langle \cdot, \cdot \rangle$ as an inner product on V). By (b), $T^*v \in U$, and so we can restrict this inner product to the subspace U . It follows that $S^* = T^*$. [More formally, $\langle u, (S^* - T^*)v \rangle = 0$, where u is an arbitrary vector in U , and S^*v, T^*v are both in U as well. Let $u = (S^* - T^*)v$, and

we have $(S^* - T^*)v = 0$ for all $v \in U$.]

(d) $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are both normal.

$(T|_U)^*(T|_U) = (T^*|_U)(T|_U) = (T|_U)(T^*|_U) = (T|_U)(T|_U)^*$ [essentially this holds because restriction does not change the properties of maps]. Because U^\perp is also invariant under T , by the same token we can prove that $T|_{U^\perp}$ is normal as well.

Since a normal operator restricted to an invariant subspace is still normal, we can use our description of normal but not self-adjoint operators of dimension 2 to describe normal operators on real inner product spaces of arbitrary dimensions by induction, from which we can completely describe isometries on real inner product spaces.

- For T on a real inner product space V , T is normal iff there is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is 1×1 or 2×2 of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, with $b > 0$.

We first mention an easy lemma: for block diagonal matrices A and B with the same block sizes on their diagonals, AB is still block diagonal with each block of AB being the product of the corresponding block in A times the corresponding block in B .

(\Leftarrow) Each 1×1 block commutes with its transpose (itself), and the 2×2 block of the given form must also commute with its transpose [showed earlier in this chapter; see (*)]. By the preceding lemma, the block diagonal matrix of T should commute with its transpose, showing that T commutes with T^* . This completes the proof.

(\Rightarrow) We use induction. The case $\dim V = 1$ is trivial. When $\dim V = 2$, if T is self-adjoint, then by the real spectral theorem we should find two 1×1 blocks on the diagonal and two 0s off the diagonal; if T is not self-adjoint, then the first theorem in the section applies.

Assume $\dim V > 2$, and that the result holds on real inner product spaces of smaller dimensions. In §9.1 we showed that every operator must have invariant subspaces of dimension 1 or 2. Let U be a dimension 1 subspace of V invariant under T , if a dimension 1 subspace exists (equivalently we can say if T has an eigenvector in V , then let U be the span of this eigenvector); if not, let U be a dimension 2 subspace of V invariant under T , and in this case $T|_U$ cannot be a self-adjoint operator (because a self-adjoint operator must have an eigenvalue, which was proved in Chapter 7). *This construction maximizes the number of invariant subspaces of dimension 1.*

If $\dim U = 1$, then consider a vector in U of norm 1. This vector forms an orthonormal basis of U , with respect to which $T|_U$ has a 1×1 matrix. If $\dim U = 2$, then the matrix of $T|_U$ with respect to some orthonormal basis of V is of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $b > 0$ (because $T|_U$ is a normal but not self-adjoint operator).

Now since $\dim U^\perp < \dim V$, and $T|_{U^\perp}$ is normal on U^\perp by the preceding lemma, the inductive hypothesis tells us that the desired orthonormal basis of U^\perp exists. Adjoining the orthonormal basis of U we created to this orthonormal basis of U^\perp gives us an orthonormal

basis of the whole V , with respect to which $\mathcal{M}(T)$ is still of the form desired.

- Now we can describe every isometry on a real inner product space. Let S be an operator on a real inner product space V , then S is an isometry iff 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 1×1 matrix containing 1 or -1 , or is a 2×2 rotation matrix of the form $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, where $\theta \in (0, \pi)$.

If we assume that S is an isometry, then S is normal tells us that there is an orthonormal basis of S with respect to which $\mathcal{M}(S)$ has the special block diagonal form. If $\lambda \in \mathbf{R}$ is an entry in the 1×1 blocks on the diagonal of $\mathcal{M}(S)$, then $Se_j = \lambda e_j$ for some e_j of norm 1. Because S is an isometry, $|\lambda| = 1$.

Now we consider 2×2 blocks of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $b > 0$ on the diagonal of $\mathcal{M}(S)$. This means that two orthonormal vectors e_j and e_{j+1} make $Se_j = ae_j + be_{j+1}$. It follows that

$$1 = \|e_j\|^2 = \|Se_j\|^2 = a^2 + b^2.$$

By $b > 0$ we can now reparameterize $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ into $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ by a single $\theta \in (0, \pi)$.

Conversely, we suppose that there is an orthonormal basis of V with respect to which $\mathcal{M}(S)$ is of the special block diagonal form as stated in the theorem. For clarity we refer to the basis as e_1, \dots, e_n . From the block diagonal form, we know V can be expressed as $U_1 \oplus \dots \oplus U_m$, where each U_j is of dimension 1 or 2, and S is invariant under each U_j . It follows that for every $v \in V$, $v = u_1 + \dots + u_m$, where each $u_j \in U_j$. Note that all the u_j 's must be pairwise orthogonal, because the basis vector(s) of one U is orthogonal to the basis vector(s) of another U . Note also that for every vector $v_j \in U_j$, if $\dim U_j = 1$, then we have $Sw_j = v_j$ or $Sw_j = -v_j$. If $\dim U_j = 2$, then given

$$\mathcal{M}(S|_{U_j}) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

for some θ with respect to the orthonormal basis vectors e_k and e_{k+1} of U , by matrix computation we see that

$$S(ae_k + be_{k+1}) = [a \cos(\theta) - b \sin(\theta)]e_k + [a \sin(\theta) + b \cos(\theta)]e_{k+1}.$$

This implies that

$$\|S(ae_k + be_{k+1})\|^2 = [a \cos(\theta) - b \sin(\theta)]^2 + [a \sin(\theta) + b \cos(\theta)]^2 = a^2 + b^2 = \|ae_k + be_{k+1}\|^2$$

for every vector $ae_k + be_{k+1} \in U_j$. Our work above in general shows that $S|_{U_j} \in \mathcal{L}(U)$ is an

isometry. Therefore, for any $v \in V$

$$\begin{aligned}\|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\ &= \|Su_1\|^2 + \cdots + \|Su_m\|^2 \quad (\text{by orthogonality, because each } Su_j \in U_j) \\ &= \|u_1\|^2 + \cdots + \|u_m\|^2 \\ &= \|u_1 + \cdots + u_m\|^2 \quad (\text{again by orthogonality}) \\ &= \|v\|^2,\end{aligned}$$

as desired.

10 Trace and Determinant

Remark. V still remains a nonzero FDVS in this last chapter.

Change of Basis

- Our focus finally comes to matrices themselves. We use I_n to denote the $n \times n$ identity matrix, and A^{-1} to denote the inverse matrix of a square matrix A .

You may appeal to the fact that there is one-to-one correspondence between every square matrix and an operator with respect to a fixed basis to show that, or simply prove directly, that $AI = IA = A$ for a square matrix A , and that there is a unique inverse B (such that $AB = BA = I$) for an invertible square matrix A .

Remark. The change of basis formula is one of the most important result in linear algebra, and we now consider the matrix of an operator with respect to two different bases (instead of only one basis). The reason is clear, because we are now changing the basis of our operator/matrix. Recall the notation $\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n))$ is the matrix of writing each Tu_j as a linear combination of v_1, \dots, v_n in the j -th column.

- Suppose $u_1, \dots, u_n, v_1, \dots, v_n$, and w_1, \dots, w_n are all bases of V . For $S, T \in \mathcal{L}(V)$, we have

$$\begin{aligned} & \mathcal{M}(ST, (u_1, \dots, u_n), (w_1, \dots, w_n)) \\ &= \mathcal{M}(S, (v_1, \dots, v_n), (w_1, \dots, w_n)) \mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n)). \end{aligned}$$

We proved this result for linear maps in Chapter 3, and we restate it here.

- For two bases u_1, \dots, u_n and v_1, \dots, v_n of V , the two matrices $\mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$ and $\mathcal{M}(I, (v_1, \dots, v_n), (u_1, \dots, u_n))$ are inverses of one another. (Note that the I here is the identity operator on V .)

This follows from the last claim and

$$I_n = \mathcal{M}(II, (u_1, \dots, u_n), (u_1, \dots, u_n)) = \mathcal{M}(II, (v_1, \dots, v_n), (v_1, \dots, v_n)).$$

- Now we can prove the **change of basis formula**. For $T \in \mathcal{L}(V)$, where V has two bases u_1, \dots, u_n and v_1, \dots, v_n . Let $A = \mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$. Then

$$\mathcal{M}(T, (u_1, \dots, u_n)) = A^{-1} \mathcal{M}(T, (v_1, \dots, v_n)) A.$$

By the previous two claims, $A^{-1} \mathcal{M}(T, (v_1, \dots, v_n)) = \mathcal{M}(T, (v_1, \dots, v_n), (u_1, \dots, u_n))$, which when multiplied by A on the right gives $\mathcal{M}(T, (u_1, \dots, u_n))$, as desired.

Remark. A number of times in this book we have discussed finding a basis of an operator with respect to which there is a “nice-looking” matrix (e.g., a diagonal, a block diagonal, or a Jordan matrix). Given a matrix $P = \mathcal{M}(T, (u_1, \dots, u_n))$, if we have shown that we can always find a kind

of “nice-looking” matrix Q of the operator T with respect to a basis (v_1, \dots, v_n) , then we can always decompose P into $A^{-1}QA$, where $A = \mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$.

P and Q are oftentimes called *similar matrices* when $P = A^{-1}QA$. They are similar in the sense that $Q = (A^{-1})^{-1}P(A^{-1})$ as well.

Now we look back at the diagonalization computation from our first course in linear algebra. Consider the linear operator $T \in \mathcal{L}(\mathbf{F}^n)$ of multiplying a $n \times n$ matrix P on the left. Then $P = \mathcal{M}(T, (e_1, \dots, e_n))$, and $A^{-1} = \mathcal{M}(I, (v_1, \dots, v_n), (e_1, \dots, e_n))$. We know (v_1, \dots, v_n) is a basis of eigenvectors of \mathbf{F}^n , and thus the j -th column of A^{-1} is exactly the j -th eigenvector $v_j \in \mathbf{F}^n$.

Recall when diagonalizing a matrix P , we used to find all the eigenvalues $\lambda_1, \dots, \lambda_n$ and their corresponding eigenvectors v_1, \dots, v_n , and then adjoin all the eigenvectors to get the $A^{-1} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and its inverse A , that $Q = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ needs to multiply on the left and right respectively. These computation steps are finally justified.

Trace and Determinant of an Operator

- The *trace of an operator* T on a complex (resp. real) vector space is the sum of eigenvalues of T (resp. $T_{\mathbf{C}}$), with each eigenvalue repeated according to its multiplicity. The *determinant of an operator* T on a complex (resp. real) vector space is the product of the eigenvalues of T (resp. $T_{\mathbf{C}}$), with each eigenvalue repeated according to its multiplicity.
- Let $\dim V = n$, and $\lambda_1, \dots, \lambda_n$ be all the eigenvalues of $T/T_{\mathbf{C}}$ (counting repeats according to multiplicities), then the characteristic polynomial of T should be

$$(z - \lambda_1) \cdots (z - \lambda_n).$$

By polynomial expansion (or Vieta’s formulas), $\text{trace } T = \lambda_1 + \cdots + \lambda_n$ is the negative coefficient of z^{n-1} in the characteristic polynomial of T , and $\det T = \lambda_1 \cdots \lambda_n$ is $(-1)^n$ times the constant term of the characteristic polynomial of T . The characteristic polynomial of T now explicitly should look like

$$z^n - (\text{trace } T)z^{n-1} + \cdots + (-1)^n(\det T).$$

- An operator on V is invertible iff its determinant is nonzero.
Note that T on a complex vector space is invertible iff T does not have 0 as an eigenvalue, which happens iff the constant term of the characteristic polynomial is nonzero, i.e., $\det T \neq 0$. If T is instead on a real vector space, then we need one additional step in the middle: T does not have 0 as an eigenvalue iff $T_{\mathbf{C}}$ does not have 0 as an eigenvalue.
- We now relate the idea of the determinant to the characteristic polynomial. The following statement can be taken as an alternative definition of the characteristic polynomial:

For T on V , the characteristic polynomial of T equals $\det(zI - T)$.

Suppose V is a complex vector space. Let $\lambda \in \mathbf{C}$, and for now we see z as a fixed complex indeterminate. Then λ is an eigenvalue of T iff $z - \lambda$ is an eigenvalue of $zI - T$, as can be seen from

$$-(T - \lambda I) = (zI - T) - (z - \lambda)I.$$

Raise both sides to the power $n = \dim V$ and take nullspaces. Then the multiplicities of λ in T should match the multiplicities of $z - \lambda$ in $zI - T$. Let $\lambda_1, \dots, \lambda_n$ be all eigenvalues of T (including repeats according to multiplicities), then $z - \lambda_1, \dots, z - \lambda_n$ are all the eigenvalues of $zI - T$ (including repeats according to multiplicities). Therefore, for the complex indeterminate z ,

$$\det(zI - T) = (z - \lambda_1) \cdots (z - \lambda_n),$$

where the RHS is exactly the characteristic polynomial of T .

If V is a real vector space instead, then by definition the characteristic polynomial of T is the characteristic polynomial of $T_{\mathbf{C}}$, and $\det(zI - T)$ is the product of eigenvalues of $(zI - T)_{\mathbf{C}} = zI_{\mathbf{C}} - T_{\mathbf{C}}$ (because when z is a complex number, $zI - T$ is a polynomial in T). Then we can use the claim above in the complex case.

Trace of a Matrix

- The *trace of a square matrix* A , which we denote by $\text{trace } A$, is defined as the sum of the diagonal entries of A . We will show that the trace of an operator $T \in \mathcal{L}(V)$ is exactly the trace of $\mathcal{M}(T)$ with respect to any basis of V .
- The trace of the transpose of a matrix is the same as the trace of the matrix itself.
- $\text{trace}(AB) = \text{trace}(BA)$ for matrix A of dimension $m \times n$ and B of dimension $n \times m$. Proof is easy.
- Similar matrices have the same trace by the claim above. Let $P = A^{-1}QA$, then

$$\text{trace } P = \text{trace}(A^{-1}QA) = \text{trace}(QA^{-1}A) = \text{trace } Q.$$

– In particular, for two bases u_1, \dots, u_n and v_1, \dots, v_n of V ,

$$\text{trace } \mathcal{M}(T, (u_1, \dots, u_n)) = \text{trace } \mathcal{M}(T, (v_1, \dots, v_n)).$$

- By the theorem above, the trace of the matrix of an operator is irrespective of the basis chosen. *We will soon see that the determinant of the matrix of an operator is irrespective of the basis chosen as well.* From the block diagonal decomposition of T on a complex vector space (from §8.2), we may conclude that $\text{trace } \mathcal{M}(T)$ should be the sum of eigenvalues of T (repeated according to their multiplicities), which is exactly $\text{trace } T$.

If T is on a real vector space, then consider $T_{\mathbf{C}}$. Remember earlier we said that $\mathcal{M}(T) = \mathcal{M}(T_{\mathbf{C}})$, so that $\text{trace } \mathcal{M}(T) = \text{trace } \mathcal{M}(T_{\mathbf{C}})$. By definition $\text{trace } T$ now is the sum of eigenval-

ues of $T_{\mathbf{C}}$ (repeated according to their multiplicities), which is exactly $\text{trace } \mathcal{M}(T_{\mathbf{C}})$ by the last paragraph. The two cases combined finish the proof.

- When computing the trace (and also the determinant, as we will see later) of an operator, we may just look at any of its matrix without computing the eigenvalues and their multiplicities.
- Trace is additive: $\text{trace}(S + T) = \text{trace } S + \text{trace } T$ for $S, T \in \mathcal{L}(V)$. This holds obviously if we look at $\mathcal{M}(S)$ and $\mathcal{M}(T)$ with respect to the same basis of V .
 - Note that meanwhile for any $\lambda \in \mathbf{F}$, $\text{trace}(\lambda T) = \lambda \text{trace}(T)$ because $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$ with respect to the same basis of V .

These facts tell us that the trace of an operator is a linear functional on $\mathcal{L}(V)$.

- There do not exist operators $S, T \in \mathcal{L}(V)$ such that $ST - TS = I$. This statement has nothing to do with traces on the surface, but interestingly the use of trace makes this claim of nonexistence obvious.

Consider $\mathcal{M}(S)$ and $\mathcal{M}(T)$ with respect to the same basis of V , then

$$\begin{aligned} \text{trace}(ST - TS) &= \text{trace}(ST) - \text{trace}(TS) \\ &= \text{trace}(\mathcal{M}(ST)) - \text{trace}(\mathcal{M}(TS)) \\ &= \text{trace}(\mathcal{M}(S)\mathcal{M}(T)) - \text{trace}(\mathcal{M}(T)\mathcal{M}(S)) = 0, \end{aligned}$$

yet $\text{trace } I = 1 \cdot \dim V = \dim V \neq 0$.

Determinant of a Matrix

- We use (m_1, m_2, \dots, m_n) as a shorthand for the permutation $\begin{pmatrix} 1 & 2 & \cdots & n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix}$. Note this is not the cycle notation. The *sign of a permutation* (m_1, m_2, \dots, m_n) is defined to be
 - i) 1 if the number of pairs of integers (j, k) with $1 \leq j \leq k \leq n$ such that j appears after k in the list (m_1, m_2, \dots, m_n) is even;
 - ii) -1 if the number of such pairs is odd.

In other words, the sign of (m_1, \dots, m_n) , which we denote by $\text{sgn}(m_1, \dots, m_n)$, is 1 (resp. -1) if the natural order of the list has been changed an even (resp. odd) number of times.

Examples 10.26 and 10.28 in the book provide good intuition to why we should define the sign of a permutation in this way, and how this helps the definition of the determinant of a matrix.

- Interchanging two entries in a permutation (i.e., a single transposition) multiplies the sign of the permutation by -1 .

Consider two permutations, where the second permutation can be obtained from the first in one transposition. This two interchanged entries from b, a to a, b first leads $+1$ or -1 pairs not in the natural order.

Consider any entry c between the two interchanged entries b and a .

- i) If $b < c < a$ originally, then $a > c > b$ now, so $+2$ pairs not in the natural order.
- ii) If $b > c$ and $c < a$ (or $b < c$ and $c > a$) originally, then switching b and a does not change the number of pairs not in the natural order.
- iii) If $b > c > a$ originally, then similar to i) we have -2 pairs not in the natural order after switching b and a .

Therefore, the parity of the number of pairs not in the natural order changes in a transposition. This means that the sign of a permutation is multiplied by -1 in a transposition.

- A corollary crucial to the study of permutations in abstract algebra: a permutation must be either an even permutation or an odd permutation, i.e., it can be broken into only an odd number of transpositions or an even number of transpositions. (Recall that every permutation can be broken into transpositions.)

We know $(1, \dots, n)$ has sign 1. If $\text{sgn}(m_1, \dots, m_n) = 1$ (resp. -1), then the number of transpositions must be even (resp. odd).

The sign of a permutation proves that the parity of a permutation is well-defined. The two equivalent definitions have their use in their own areas. Here because the determinant of a matrix is a number, we use the sign property. But in abstract algebra when we are dealing with permutations as functions, the parity property is more important.

- The Leibniz formula for the definition of determinants: for the $n \times n$ square matrix

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{bmatrix},$$

$\det A$, the *determinant of the square matrix* A , is given by

$$\sum_{(m_1, \dots, m_n) \in S_n} \text{sgn}(m_1, \dots, m_n) A_{m_1,1} \cdots A_{m_n,n},$$

where S_n is the symmetric group of permutations on n letters.

- It is equivalent to define $\det A$ as

$$\sum_{(k_1, \dots, k_n) \in S_n} \text{sgn}(k_1, \dots, k_n) A_{1,k_1} \cdots A_{n,k_n}. \quad (8)$$

For fixed m_1, \dots, m_n , let $\sigma = (m_1, \dots, m_n)$. Then

$$A_{m_1,1} \cdots A_{m_n,n} = A_{1,\sigma^{-1}(1)} \cdots A_{n,\sigma^{-1}(n)}.$$

Remember that $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$: since the identity permutation $\iota = \sigma^{-1}\sigma$ is even, σ^{-1} and σ must have the same parity, and thus the two have the same sign. Therefore, we

have shown that

$$\operatorname{sgn}(\sigma)A_{m_1,1} \cdots A_{m_n,n} = \operatorname{sgn}(\sigma^{-1})A_{1,\sigma^{-1}(1)} \cdots A_{n,\sigma^{-1}(n)}.$$

Clearly summing σ over S_n and summing σ^{-1} over S_n are the same, and thus we have our alternative definition (8).

- The determinant of an upper-triangular matrix equals the product of all the diagonal entries.

This is pretty obvious. Except for $A_{1,1} \cdots A_{n,n}$ with $\operatorname{sgn}(\iota) = 1$, all the other products in the sum vanish because one of the A_{j,m_j} in each product is always below the diagonal.

- The determinant of a matrix is equal to that of its transpose.

For fixed m_1, \dots, m_n , again let $\sigma = (m_1, \dots, m_n)$. Note that

$$A_{1,m_1} \cdots A_{n,m_n} = A_{\sigma^{-1}(1),1} \cdots A_{\sigma^{-1}(n),n},$$

and so essentially we need to show that $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$, which we already know.

- Interchanging two columns/rows of a matrix changes the sign of the determinant.

This is easy. Consider the product $A_{m_1,1} \cdots A_{m_n,n}$. For matrix B obtained from A by switching two columns, the sign in B associated with the same product $A_{m_1,1} \cdots A_{m_n,n}$ is -1 multiplied by the original $\operatorname{sgn}(m_1, \dots, m_n)$ in A .

Interchanging two rows of A can be seen as interchanging two columns of A^t . By the fact that the determinant of a matrix is equal to that of its transpose, we see that interchanging two rows of a matrix as well changes the sign of the determinant as well.

– Matrices with two equal columns/rows have 0 determinant. (Consider switching the two equal columns.)

- For square matrix $A = \begin{bmatrix} A_{\cdot,m_1} & \cdots & A_{\cdot,m_n} \end{bmatrix}$ and permutation (m_1, \dots, m_n) ,

$$\det \begin{bmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n} \end{bmatrix} = \operatorname{sgn}(m_1, \dots, m_n) \det A.$$

When (m_1, \dots, m_n) is an even (resp. odd) permutation, $\begin{bmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n} \end{bmatrix}$ is obtained from A by switching columns an even (resp. odd) number of times, exactly when $\operatorname{sgn}(m_1, \dots, m_n) = 1$ (resp. -1).

- Let $k, n \in \mathbf{Z}^+$ with $1 \leq k \leq n$. Fix the $n \times 1$ columns $A_{\cdot,1}, \dots, A_{\cdot,n}$ except for $A_{\cdot,k}$. The function $f : A_{\cdot,k} \mapsto \det \begin{bmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n} \end{bmatrix}$ is a linear map from \mathbf{F}^n to \mathbf{F} .

Recall from §3.4 that linear maps from \mathbf{F}^n to \mathbf{F}^m can be uniquely represented by a matrix-left-multiplication. Since $m = 1$ here, basically we want to show that $\det \begin{bmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n} \end{bmatrix}$ is a linear combination of the entries of $A_{\cdot,k}$. This is clearly true by the Leibniz formula.

- For square matrices A, B of the same size, $\det(AB) = \det(A)\det(B) = \det(BA)$.

$$\begin{aligned}
\det(AB) &= \det \begin{bmatrix} AB_{\cdot,1} & \cdots & AB_{\cdot,n} \end{bmatrix} \\
&= \det \begin{bmatrix} A(\sum_{m_1=1}^n B_{m_1,1}e_{m_1}) & \cdots & A(\sum_{m_n=1}^n B_{m_n,n}e_{m_n}) \end{bmatrix} \\
&= \det \begin{bmatrix} \sum_{m_1=1}^n B_{m_1,1}Ae_{m_1} & \cdots & \sum_{m_n=1}^n B_{m_n,n}Ae_{m_n} \end{bmatrix} \\
&\quad (\text{because the } B_{m_i,i} \text{'s are constants}) \\
&= \sum_{m_1=1}^n \cdots \sum_{m_n=1}^n B_{m_1,1} \cdots B_{m_n,n} \det \begin{bmatrix} Ae_{m_1} & \cdots & Ae_{m_n} \end{bmatrix} \\
&\quad (\text{by repeated applications of the linearity of } \det \text{ as a function of one column})
\end{aligned}$$

Since the determinant of a matrix with two equal columns is 0, the sum above is equivalent to

$$\sum_{(m_1, \dots, m_n) \in S_n} B_{m_1,1} \cdots B_{m_n,n} \det \begin{bmatrix} Ae_{m_1} & \cdots & Ae_{m_n} \end{bmatrix}.$$

Therefore,

$$\begin{aligned}
\det(AB) &= \sum_{(m_1, \dots, m_n) \in S_n} B_{m_1,1} \cdots B_{m_n,n} (\operatorname{sgn}(m_1, \dots, m_n)) \det(A) \\
&= \det(A) \sum_{(m_1, \dots, m_n) \in S_n} \operatorname{sgn}(m_1, \dots, m_n) B_{m_1,1} \cdots B_{m_n,n},
\end{aligned}$$

which shows that $\det(AB) = \det(A) \det(B) = \det(B) \det(A) = \det(BA)$.

- By $\det(AB) = \det(BA)$, similar matrices have the same determinant, just like the trace.
 - In particular, the determinant of the matrix of an operator T is irrespective of the basis chosen, which we can simply write as $\det \mathcal{M}(T)$.
 - It is also the case that the determinant of an operator is exactly the determinant of its matrix.

The proofs of these results are the same as the proofs of the same results about the trace.

- While trace is additive, determinant is multiplicative: $\det(ST) = \det(S) \det(T) = \det(TS)$ for $S, T \in \mathcal{L}(V)$. Just like the similar result about trace, the result here holds obviously if we fix a basis of V and work with $\mathcal{M}(S)$ and $\mathcal{M}(T)$.

Geometric Interpretation of the Determinant

- For an isometry S on an inner product space V , $|\det S| = 1$.

In the case where $\mathbf{F} = \mathbf{C}$, recall from §7.3 that all eigenvalues of S has absolute value 1. Their product (counting multiplicities), which is $\det S$, must have absolute value 1 as well.

In the case where $\mathbf{F} = \mathbf{R}$, the book offers two proofs:

- The first proof uses the result that for real inner product space V , it can be routinely verified from definition that

$$\langle u + iv, x + iy \rangle = \langle u, x \rangle + \langle v, y \rangle + (\langle v, x \rangle - \langle u, y \rangle)i$$

for $u, v, x, y \in V$ defines a complex inner product on $V_{\mathbf{C}}$.

Consider $S_{\mathbf{C}}(u + iv) = Su + iSv$ and $S_{\mathbf{C}}(x + iy) = Sx + iSy$. To show that $S_{\mathbf{C}}$ is inner-product-preserving, we look at $\langle S_{\mathbf{C}}(u + iv), S_{\mathbf{C}}(x + iy) \rangle$:

$$\begin{aligned} \langle Su + iSv, Sx + iSy \rangle &= \langle Su, Sx \rangle + \langle Sv, Sy \rangle + (\langle Sv, Sx \rangle - \langle Su, Sy \rangle)i \\ &= \langle u, x \rangle + \langle v, y \rangle + (\langle v, x \rangle - \langle u, y \rangle)i \\ &\quad \text{[exactly because } S \text{ is inner-product-preserving]} \\ &= \langle u + iv, x + iy \rangle. \end{aligned}$$

This shows that $S_{\mathbf{C}}$ is an isometry on $V_{\mathbf{C}}$, and so by the case $\mathbf{F} = \mathbf{R}$, $|\det S_{\mathbf{C}}| = 1$. Since $\det S$ is $\det S_{\mathbf{C}}$ for S on a real inner product, the proof is completed.

In summary, we have chosen an inner product for the complexification $V_{\mathbf{C}}$ such that $S_{\mathbf{C}}$ is an isometry on it. This *enables* us to use the complex case we have already established.

- The second proof uses the result that the determinant of a block diagonal matrix is equal to the product of the determinants of each block. The proof of this is some complicated calculations involving the Leibniz formula. In §9.2 we described every isometries on a real inner product space by a block diagonal matrix of a specific form (when the right basis is chosen). The diagonal blocks must be $\begin{bmatrix} 1 \end{bmatrix}$, $\begin{bmatrix} -1 \end{bmatrix}$, or the 2×2 rotational matrix, which has determinant 1. Therefore $|\det S| = |\det \mathcal{M}(S)| = 1$.
- The sign of the $\det T$ for an invertible operator T on a real inner product space V can be geometrically comprehended. (An invertible operator T must not have 0 as an eigenvalue, and thus $\det T$ must be either positive or negative.)

By the polar decomposition theorem, $T = S\sqrt{T^*T}$ for an isometry $S \in \mathcal{L}(V)$. It follows that $\det(T) = \det(S) \det(\sqrt{T^*T})$.

First recall that $\sqrt{T^*T}$ is the unique positive square root of T^*T , and then must have all its eigenvalues nonnegative. By the real spectral theorem, because $\sqrt{T^*T}$ is self-adjoint, with respect to an orthonormal basis of V the matrix $\mathcal{M}(\sqrt{T^*T})$ becomes diagonal and has all eigenvalues repeated according to their multiplicities. Therefore $\det \sqrt{T^*T}$ is the product of these eigenvalues of $\sqrt{T^*T}$ (repeated accordingly), which is nonnegative. Since $\det T \neq 0$, $\det \sqrt{T^*T} > 0$, which means that the sign of $\det T$ is exactly the sign of $\det S$.

In the second proof of $|\det S| = 1$ above, we see that whether $\det S = 1$ or -1 depends on whether the number of $\begin{bmatrix} -1 \end{bmatrix}$ blocks on the diagonal is even or odd. Recall the construction of the block diagonal matrix in §9.2 maximizes the number of invariant subspaces of dimension 1, and thus the dimension of $E(-1, S) = \{v \in V \mid Sv = -v\}$ is

exactly the number of $\begin{bmatrix} -1 \end{bmatrix}$ blocks on the diagonal. We may conclude that whether $\det T$ is positive or negative depends on whether $\dim E(-1, S)$, i.e., the subspace of V on which vectors reverse directions through S , has even or odd dimension.

The book continues with a pseudo-proof of the change of variables theorem. The pseudo-proof clearly illustrates where the determinant in the formula comes from. Some measure theory is required to make the proof actually rigorous. Please check out the book for the exposition.