MATH 110 Notes

based on *Linear Algebra Done Right*, third edition

Cheng, Feng

1 Vector Spaces

- A vector space V is an abelian group under addition that has the distributive and associative properties in scalar multiplication and the multiplicative identity. (Here addition and scalar multiplication are closed operations in V.)
- $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$. It is a vector space over $\mathbb{F}$ 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 the $\{0\}$. 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.
- 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" of vectors, which is no different than using a "set" of vectors because vector addition is commutative. (We will use either a list or a set of vectors later depending on whether we have to specify the ordering of these vectors.) Finite-dimensional 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 set S of vectors in V is defined to be all possible linear combinations of these vectors. In particular, $\text{span}(\emptyset) := \{0\}$.
- 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 $\mathbb{F}$ is the (not unique) **expression** $a_0 + a_1 z + \cdots + a_n z^n$ with all the a_i's in $\mathbb{F}$. 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 just undefined.

  P.S. The most general and rigorous definition of a polynomial over a ring (sometimes as an infinite formal sum for uniqueness) is not discussed here, because our focus is vector spaces over a field.

  - It is oftentimes useful to define a polynomial over a field $\mathbb{F}$ rather as the **function** $p : \mathbb{F} \to \mathbb{F}$ with $p(z) = a_0 + a_1 z + \cdots + a_n z^n$. 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 $\mathbb{F}$ by $\mathcal{P}(\mathbb{F})$. Under the usual coefficient-wise addition and scalar multiplication, it is obvious that $\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$. In addition, the set of all polynomials with degree at most m, denoted by $\mathcal{P}_m(\mathbb{F})$ is a vector subspace of $\mathcal{P}(\mathbb{F})$.
- Linear independence and dependence can be characterized in multiple ways. A set S of vectors is linearly independent iff the only way to write 0 as $a_1 v_1 + a_2 v_2 + \cdots + a_m v_m$ 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 will not change the span.
  - In particular, for a linearly independent set $S = \{s_1, s_2, \ldots, s_m\} \subsetneq$ a linearly dependent set $T = \{s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_n\}$, one such vector is always in $T \setminus S$. This is because for the nontrivial way of writing $0 = a_1 s_1 + \cdots + a_m s_m + b_1 t_1 + \cdots + b_n t_n$, 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 S but does not belong to $\text{span}(S)$, then the resulting set is still linearly independent.
    * From this special case we know that for a linear dependent set $\{v_1, \ldots, v_n\}$, at some index $k \leq n$ we have $\{v_1, \ldots, v_{k-1}\}$ being linearly independent, while $v_k \in \text{span}(\{v_1, \ldots, v_{k-1}\})$.

(This is the well-ordering case of the linear dependence lemma 2.21 in book that we will use again in Chapter 5.)

  - Note that a linearly independent set 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 $S = \{w_1, w_2, \ldots, w_n\}$, then all linearly independent sets of vector $\{u_1, u_2, \ldots, u_m\} \subseteq V$ must have $m \leq n$. The proof suggests adding a new u_i into S and then remove another w_j from S in each step. In the end all the u's will be added and the corresponding number of w's are removed from S. 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 set of V. A set 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 set of a FDVS V can be reduced to a finite basis. Furthermore, every linearly independent set of a FDVS V can be extended to a basis.
  - To prove the first claim, we construct our basis B from scratch and try to add all the elements of the spanning set $S = \{v_1, v_2, \ldots, v_n\}$ 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, $\text{span}(B) = \text{span}(S) = V$ because all the v's left out are in $\text{span}(B)$.
    * It is also possible to start B from S 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, $\text{span}(B) = \text{span}(S)$ and none of the vectors in B is in the span of the others. However, this is top-down approach is harder in practice to actually find the basis.
    * This claim shows that every FDVS has a finite basis.
  - To prove the second claim, similarly construct our basis from the given linearly independent set L. We pick a finite basis $B = \{b_1, b_2, \ldots, b_n\}$ of V 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 set or linearly independent set 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 set 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. $\text{span}(B_V \backslash B_U)$ 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 $\{u_1, u_2, \ldots, u_m\}$ 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 $\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2$.

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 $T : V \to W$ such that for all $u, v \in V$ and $\lambda \in \mathbb{F}$ (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 $I_V : V \to V$ 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 $\mathbb{F}$. 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

$(I_W T = TI_V = T)$. 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.

- If $T(x_1, \ldots, x_n) = (a_{11}x_1+\cdots+a_{1n}x_n, \ldots, a_{m1}x_1+\cdots+a_{mn}x_n)$, then T is obviously a linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$. For a linear map $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$, it must be of this form, and all the a_{ij}'s are determined by the mapping of the n canonical basis vectors of $\mathbb{F}^n$.
- The uniqueness of the linear map that takes basis vectors $v_1, v_2, \ldots, v_n$ to corresponding vectors $w_1, w_2, \ldots, w_n$ 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* $\operatorname{null} T$ is the preimage of 0_W. The *range/image* $\operatorname{range} T$ 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 $U = W$, rank $T = \dim W$ implies surjection.
  - An easy theorem to always keep in mind is that the map of the basis vectors of V spans $\operatorname{range} T$. 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, $\operatorname{range} T$ is finite-dimensional with

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

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

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

    implies $a_1 v_1 + \cdots + a_n v_n \in \operatorname{null} T$. Yet we also have $b_1 u_1 + \cdots + b_n u_n \in \operatorname{null} T$, 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 $\operatorname{null} T = 0$ (proof follows straight

from the definition of nullspace). The results below follow.

  - A map to a smaller dimensional space cannot be injective because $\dim(\operatorname{null} T) > 0$.
  - A map to a large dimensional space cannot be surjective because $\dim(\operatorname{range} T) \leq \dim V < \dim W$.
  - V and W are FDVS of equal dimension, then

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

- 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 $x \in \mathbb{F}^n$ to $0 \in \mathbb{F}^m$ 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 $v_1, \ldots, v_n$ and m-dimensional W with basis $w_1, \ldots, w_m$. For $T \in \mathcal{L}(V, W)$, we can find for every v_k, a list of $A_{i,k}$'s such that $Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m = \sum_{j=1}^{m} A_{j,k}w_j$. 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. $\mathcal{M}(T, (v_1, \ldots, v_n), (w_1, \ldots, w_n))$ uniquely determines the linear map $T \in \mathcal{L}(V, W)$.
  - Assume

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

    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, $\mathbb{F}^{m,n}$, the set of matrices with m rows and n columns over $\mathbb{F}$, 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 $c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}$.

3.4 Invertibility and Isomorphic Vector Spaces

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

- An *isomorphism* is an invertible linear map; if an isomorphism exists between two vector spaces, then the vector spaces are *isomorphic*.

- Two FDVS over $\mathbb{F}$ are isomorphic iff they have the same dimension. Injection gives us nullity $= 0$ and surjection gives us rank $T = \dim W$. The rank-nullity theorem tells us $\dim V = \dim W$. On the other hand to prove isomorphism, we show $T(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n$ is both injective and surjective, which is not hard.

  - The theorem above implies n-dimensional VS is always isomorphic to $\mathbb{F}^n$.

- Now we show the isomorphism between a linear map and its corresponding matrix. When the orders of the bases $v_1, \ldots, v_n$ and $w_1, w_2, \ldots, w_m$ are fixed, then $\mathcal{M}$ is a function from $\mathcal{L}(V, W)$ to $\mathbb{F}^{m,n}$. As we have noted earlier $\mathcal{M}$ is linear. Thus, it suffices to show that $\mathcal{M}$ is bijective.

  - For all $A \in \mathbb{F}^{m,n}$, let

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

    Then this is the *unique* (recall the theorem we emphasized 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 $v = a_1 v_1 + \cdots + a_n v_n$. Here $\mathcal{M}$ is an isomorphism from V to $F^{n(,1)}$.

  - Following this definition, $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(Tv_k)$.
  - 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 $v = a_1 v_1 + \cdots + a_n v_n$ and $\mathcal{M}$ is linear,

$$\begin{aligned} \mathcal{M}(Tv) &= a_1 \mathcal{M}(Tv_1) + \cdots + a_n \mathcal{M}(Tv_n) \\ &= a_1 \mathcal{M}(T)_{\cdot,1} + \cdots + a_n \mathcal{M}(T)_{\cdot,n} \\ &= \mathcal{M}(T)\mathcal{M}(v). \end{aligned}$$

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

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, leftrightarrow, "\mathcal{M}"', "\simeq"] & W \arrow[d, leftrightarrow, "\mathcal{M}"', "\simeq"] \\ \mathbb{F}^{n,1} \arrow[r, "\mathcal{M}(T)"] & \mathbb{F}^{m,1} \end{tikzcd}$$

    For a vector $v \in V$, its image $\mathcal{M}(v)$ under the isomorphism $\mathcal{M}$ of V belongs to $\mathbb{F}^{n,1}$. Matrix multiplication with $\mathcal{M}(T)$ gives us $\mathcal{M}(Tv) \in \mathbb{F}^{m,1}$, 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 $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$ and we are using the canonical bases, then the linear map is equivalent to matrix 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 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 $f : S \to S$ (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. Injection means $\operatorname{nullity} T = 0$, which implies $\operatorname{rank} T = \dim V$.

  (c) $\implies$ (a) by showing T is injective, as $\operatorname{nullity} T = \dim V - \operatorname{rank} T$, which is 0.

3.5 Products and Quotients of Vector Spaces

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

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

  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 $\mathbb{F}$ 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 $(v_1, \ldots, v_n)$ 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 $U_1, \ldots, U_n$ of V, define the linear map $\Gamma : U_1 \times \cdots \times U_n \to U_1 + \cdots + U_n$ by

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

  (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 $\operatorname{null} T = 0$ 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 last two theorems lead to the following 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

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

  (a) $\Longrightarrow$ (b) as $v + u = w + ((v - w) + u) \in w + U$ (and the same for the other direction). (b) $\Longrightarrow$ (c) because $v + U$ is not empty, while (c) $\Longrightarrow$ (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 $v-v'$ and $w-w' \in U$, $(v+w)-(v'+w') \in U$, 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* $\pi : V \to V/U$ 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 $\operatorname{null}\pi=U$. 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

  $$a_1v_1+\cdots+a_nv_n+b_1u_1+\cdots+b_mu_m=0,$$

  then necessarily $(a_1v_1+\cdots+a_nv_n) \in U$, which implies that

  $$0+U=(a_1v_1+\cdots+a_nv_n)+U=a_1(v_1+U)+\cdots+a_n(v_n+U).$$

  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 $\tilde{T} : V/(\operatorname{null}T) \to W$ of $T : V \to W$ by

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

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

$$\begin{tikzcd} V \arrow[rr, "T"] \arrow[dr, "\pi"'] & & W \\ & V/\operatorname{null} T \arrow[ur, "\tilde{T}"'] & \end{tikzcd}$$

- 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 $\operatorname{range} \tilde{T} = \operatorname{range} T$. The latter simply follows from the definition, while the first requires us to prove $\operatorname{null} \tilde{T} = \{0 + \operatorname{null} T\}$, the identity. $\tilde{T}(v + \operatorname{null} T) = Tv = 0$ gives us $v - 0 \in \operatorname{null} T$. Hence, $v + \operatorname{null} T = 0 + \operatorname{null} T$, and the conclusion follows.
  - The two claims combined tell us $V/(\operatorname{null} T) \simeq \operatorname{range} T$ under the isomorphism $\tilde{T}$.

3.6 Duality

- A *linear functional* on V is a linear map belonging to $\mathcal{L}(V, \mathbb{F})$. The *dual space* of V, denoted usually by V^* or V', is the vector space $\mathcal{L}(V, \mathbb{F})$.
- 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 $v_1, \ldots, v_n$, basis of V, is a set of vectors $\varphi_1, \ldots, \varphi_n$ 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 $\varphi_1, \ldots, \varphi_n$ is linearly independent in V'. Consider

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

  Since $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$, all a_j's are 0.
  - Remember that the dual basis vector φ_i of $e_1, \ldots, e_n$ in $\mathbb{F}^n$ has

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

    Namely, the i-th dual basis vector maps $(x_1, \ldots, x_n)$ to its i-th coordinate.
- For $v_1, \ldots, v_n$ basis of V and its associated dual basis $\varphi_1, \ldots, \varphi_n$ 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 they are equal.
  - second, every $v \in V$ can be written as $\sum_{i=1}^n \varphi_i(v)v_i$. For this claim one needs to prove that for $v = a_1v_1 + \cdots + a_nv_n$, every $a_j = \varphi_j(v)$. By linearity, $\varphi_j(v) = \sum_{i=1}^n a_i\varphi_j(v_i) = a_j$.
- 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.

$$\begin{tikzcd} W \arrow[d, "\varphi"'] \arrow[r, Rightarrow, "T'"] & V \arrow[d, "\substack{\varphi\circ T\\ \text{through}\\ W}"] \\ \mathbb{F} & \mathbb{F} \end{tikzcd}$$

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

- The following algebraic properties hold for dual maps:
  * $(S+T)' = S' + T'$,
  * $(\lambda T)' = \lambda T'$,
  * $(ST)' = T'S'$

  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 $U \subseteq V$ as $\{\varphi \in V' \mid \varphi(U) = 0\}$, the subset of all linear functionals on V whose image is 0.
- There is a good example (3.104) in the book about annihilators. For the dual basis $\varphi_1, \ldots, \varphi_5$ in $(\mathbb{F}^5)'$ that corresponds to $e_1, \ldots, e_5$ in $\mathbb{F}^5$, $(\text{span}(e_1, e_2))^0 = \text{span}(\varphi_3, \varphi_4, \varphi_5)$, 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 mentions 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
$$\dim \operatorname{null} i' + \dim \operatorname{range} i' = \dim V' = \dim V.$$

  Note that $i'(\varphi) = \varphi \circ i$ for $\varphi \in V'$, giving us $\operatorname{null} i' = U^0$. 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 $U = \operatorname{range} i'$. 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 $\varphi \in U' \implies \varphi \in \operatorname{range} i'$, 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 $T \in \mathcal{L}(V, W)$, $\operatorname{null} T' = (\operatorname{range} T')^0$:

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

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

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

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

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

  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 $\operatorname{range} T' \supseteq (\operatorname{null} T)^0$ 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 $\varphi(\operatorname{null} T) = 0$, 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:

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

- The *row rank* and *column rank* of matrix A are defined to be the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$ and of the columns of A in $\mathbb{F}^{m,1}$, 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 $(1 \le j \le m)$ of W' and φ_j $(1 \le k \le n)$ 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

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

- First, $\dim \operatorname{range} T =$ column rank of $\mathcal{M}(T)$. From bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, we have $\operatorname{span}(\{Tv_1, \ldots, Tv_n\})$ is isomorphic to $\operatorname{span}(\{\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\})$ under the isomorphism $\mathcal{M}$. Thus, the two VS have the same dimensions, with $\dim \operatorname{span}(\{\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\}) =$ column rank of $\mathcal{M}(T)$. Recall that $\operatorname{span}(\{Tv_1, \ldots, Tv_n\}) = \operatorname{range} T$, thus $\dim \operatorname{range} T =$ column rank of $\mathcal{M}(T)$, as desired.
- For $A \in \mathbb{F}^{m,n}$, 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

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

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" by us back then, but we never had the chance to rigorously prove them.

- The division algorithm for polynomials is similar to the one in number theory. For $p, s \in \mathcal{P}(\mathbb{F})$ with $s \neq 0$, then there exist a pair of unique polynomials $q, r \in \mathcal{P}(\mathbb{F})$ 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 $T : \mathcal{P}_{n-m}(\mathbb{F}) \times \mathcal{P}_{m-1}(\mathbb{F}) \to \mathcal{P}_n(\mathbb{F})$ by

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

    Clearly T is a linear map, 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

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

- Each zero of a polynomial corresponds to a degree-1 factor, i.e., for $p \in \mathcal{P}(\mathbb{F})$ and $\lambda \in \mathbb{F}$, $p(\lambda) = 0$ iff $\exists q \in \mathcal{P}(\mathbb{F})$ such that for all $z \in \mathbb{F}$, $p(z) = (z - \lambda)q(z)$. 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 $p \in \mathcal{P}(\mathbb{F})$ 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 $\mathbb{F}$ have a unique representation. Suppose $f, g \in \mathcal{P}(\mathbb{F})$ and $f(z) = g(z)$ for all $z \in \mathbb{F}$, we want to show that $f = g$.

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

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

    $$h(z) = a_0 + a_1 z + \cdots + a_m z^m$$

    is not zero, then $h(z) \neq 0$ for some $z \in \mathbb{F}$. 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 $|a_m|z > |a_0| + \cdots + |a_{m-1}|$,

    $$|a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}| \leq (|a_0| + \cdots + |a_{m-1}|) z^{m-1} < |a_m z^m|.$$

    Therefore $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} \neq -a_m z^m$.

- The proof of the fundamental theorem of algebra is omitted here for the obvious reason. Every nonconstant polynomial on $\mathbb{C}$ 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 $\mathbb{C}$ has a unique factorization (up to the order of factors) of the form

    $$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m), \tag{$\star$}$$

    where $c, \lambda_1, \cdots, \lambda_m \in \mathbb{C}$. 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 $z \in \mathbb{F}$ except for λ_1. Actually

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

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

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M), \tag{1}$$

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 $\overline{\lambda}$, then

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

What we want to show is that such $(x - \lambda)$ appear exactly the same number of times $(x - \overline{\lambda})$ 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 $x \in \mathbb{R}$,

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

Therefore, for all $x \in \mathbb{R}$, $q(x) \in \mathbb{R}$. If we write $q(x) = a_0 + a_1 x + \cdots + a_{n-2}x^{n-2}$ ($n = \deg p$ 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 $x \in \mathbb{R}$. 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 $(x - \lambda_j)(x - \overline{\lambda_j})$. 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 $\mathbb{C}$ (as one of factors over $\mathbb{R}$ being different implies that one of the factors over $\mathbb{C}$ 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 T-*invariant* if $T(U) \subseteq U$, or alternatively, $T|_U$ is an operator on U.
- It can be easily verified that $\{0\}$, V, null V, and 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 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 $T \in \mathcal{L}(V)$, $\lambda \in \mathbb{F}$ 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 space is an eigenvector. If u is an eigenvector of T, then span(u) is invariant under T.
- To move λv to the left side, we consider the linear operator λI that maps v to λv. Therefore, we have $(T - \lambda I)v = 0$. Since $v \neq 0$, $T - \lambda I$ is not injective/surjective/invertible (since it is an operator on V).
- 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 eigenvectors $\lambda_1, \dots, \lambda_n$ 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,

  $$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}, \tag{2}$$

  and by applying T to both sides, we have

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

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

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

  Subtracting the two equations above give us

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

  Since $v_1, \dots, v_{k-1}$ are linearly independent, all the coefficients are 0 and thus all the a'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 idempotent operations. Given $T \in \mathcal{L}(V)$, we define T^m ($m \in \mathbb{Z}^+$) 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 $\mathbb{F}$). For $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ over $\mathbb{F}$, we define $p(T) = a_0 I + a_1 T + \ldots a_n T^n$.
  - If we fix an operator T, the evaluation map $\mathcal{P}(\mathbb{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.
- The product of $pq \in \mathcal{P}(\mathbb{F})$ of polynomials $p, q \in \mathcal{P}(\mathbb{F})$ is defined by

  $$(pq)(z) = p(z)q(z)$$

  for all $z \in \mathbb{F}$. If we expand the polynomials with Sigma notation, one can easily see that $(pq)(z) = p(z)q(z) = q(z)p(z) = (qp)(z)$.
  - It follows directly that $(pq)(T) = p(T)q(T)$ and $p(T)q(T) = q(T)p(T)$, because we only replace the symbol z by 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) 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 $v, Tv, \ldots, T^n v$, which must be linearly dependent. Thus, we have

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

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

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

  over $\mathbb{C}$. 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 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 $\mathcal{M}(T, (v_1, \ldots, v_n))$ is a square matrix. In the j-th column of $\mathcal{M}(T)$, each entry corresponds to each coefficient from $a_1v_1 + \cdots + a_nv_n = Tv_j$. 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 $v_1, \ldots, v_n$ is a basis of V, then the following are equivalent:

  (a) $\mathcal{M}(T)$ is upper triangular;

  (b) $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each j from 1 to n;

  (c) $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each j from 1 to n.

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

- For FDVS V over $\mathbb{C}$ 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 $\dim V$.

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

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

$$Tv_k \in \operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_k)$$

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

  The second proof is nonstandard but quite interesting. It sets $U = \operatorname{span}(v_1)$ 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)

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

  This implies that $Tv_j = \operatorname{span}(v_1, \ldots, v_j)$ for j between 1 and n. Recall in §3.5 we showed that $v_2, \ldots, v_n$ 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 $v_1 \in \operatorname{range} T$. 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 $v_2 \in \operatorname{range} T$. Following this fashion we can conclude that $v_1, \ldots, v_n$ are all in $\operatorname{range} T$. 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 $1 < j \leq$ suppose $\lambda_j = 0$, "looking" at the j-th column above we immediately know that $T|_{\operatorname{span}(v_1,\ldots,v_j)}$ is not injective (it maps into $\operatorname{span}(v_1, \ldots, v_{j-1})$ 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 invertible iff λ is an eigenvalue. Therefore, all the entries λ_i's on the diagonal are exactly all the eigenvalues.

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 $\lambda \in \mathbb{F}$ is given by

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

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, we get what we want. 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 the operator has a diagonal matrix with respect to some basis of T.
- For T on FDVS V with dimension n and distinct eigenvalues $\lambda_1, \ldots, \lambda_m$, 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 $\lambda_1, \ldots, \lambda_n$ 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) $\Longrightarrow$ (b): Let each $U_j = \operatorname{span}(v_j)$, each of which are invariant under T. Since $v_1, \ldots, v_n$ is a basis of V, any $v \in V$ can be written uniquely as a linear combination of $v_1, \ldots, v_n$, 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) $\Longrightarrow$ (a).
- We now show (a) $\Longleftrightarrow$ (c) $\Longleftrightarrow$ (d). (c) $\Longleftrightarrow$ (d) follows from the theorem in §3.5.
  * (a) $\Longrightarrow$ (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) $\Longrightarrow$ (a), choose a basis B_j from each $E(\lambda_j, T)$. Therefore, we get a list of vectors $v_1, \ldots, v_n$ of length $n = \dim V$ by (d). To show the list is linearly independent, consider $0 = u_1 + \cdots + u_n$, 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 $v_1, \ldots, v_n$ 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 $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ (the real n-dimensional space), the *dot product* $x \cdot y \coloneqq x_1y_1 + \cdots + x_ny_n$ is a scalar in $\mathbb{R}$. The norm $\|x\|$ on $\mathbb{R}^n$ is usually given by $\sqrt{x_1^2 + \cdots + x_n^2}$, which is the square root of the dot product of x with itself.

- The usual norm $\|z\| := \sqrt{|z_1|^2 + \cdots + |z_n|^2}$ on $\mathbb{C}^n$ (the complex n-dimensional space) is also a scalar in $\mathbb{R}$.
- Generally on both real and complex n-dimensional spaces, we define the *Euclidean inner product* by

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

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

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

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

  We regard $\mathbb{R}^n$ and $\mathbb{C}^n$ 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 $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) 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 $\mathbb{F}$ 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 $c \in \mathbb{F}$ 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 $|x_1y_1 + \cdots + x_ny_n|^2 \leq (x_1 + \cdots + x_n)^2(y_1 + \cdots + y_n)^2$.
  * One can define an inner product on function spaces such as the vector space V of continuous functions from $[a,b]$ to $\mathbb{F}$. 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 $\mathbb{F} = \mathbb{R}$, 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 $\mathbb{R}$)

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

– 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

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

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 $\mathbb{F}^n$ is orthonormal.

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

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

for all scalars $a_1,\dots,a_m$.

- It follows from the theorem above that if we set $a_1e_1+\cdots+a_ne_n=0$, 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

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

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

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

Consider $v = a_1 e_1 + \cdots + a_n e_n$ and take an inner product with e_j on both sides. Then, $\langle v, a_j \rangle = a_j$, as desired. Note that $\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$.

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

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

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

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

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

Then for any $j \le m$, $e_1, \ldots, e_j$ is still orthonormal and has the same span as $v_1, \ldots, v_j$.

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

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

By linear independent of the list of v's, $v_j \notin \operatorname{span}(v_1, \ldots, v_{j-1})$ and thus $v_j \notin \operatorname{span}(e_1, \ldots, e_{j-1})$. Therefore, $v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}$ is nonzero, and we can indeed divide it by its norm to get a new vector e_j with norm 1.

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

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

$v_j \in \operatorname{span}(e_1, \ldots, e_j)$ holds by construction and hence $\operatorname{span}(v_1, \ldots, v_j)$ is a subspace of $\operatorname{span}(e_1, \ldots, e_j)$. 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 $e_1, \ldots, e_m$ and extend the list to a basis $e_1, \ldots, e_m, v_{m+1}, \ldots, v_n$ 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 $v_{m+1}, \ldots, v_n$ will be orthonormalized to $f_{m+1}, \ldots, f_n$. 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 $\mathbb{F} = \mathbb{C}$, 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 $\operatorname{span}(v_1, \ldots, v_j)$ being invariant under for each $1 \leq j \leq n$. This 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 $\varphi_u(v) = \langle v, u \rangle$ 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 $e_1, \ldots, e_n$. Then for any v,

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

  Therefore $\overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n$ 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 two inner products are the same for any common first slot, then the second slot must be equal as well; we will use this frequently in Chapter 7.)

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

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

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

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

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

6.3 Orthogonal Complement 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 $\mathbb{R}^3$: if U is a line in $\mathbb{R}^3$ 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 $\mathbb{R}^3$ through the origin, then its orthogonal complement is the line through the origin perpendicular to the plane U.
- It is easy to verify and understand the following properties of orthogonal complements:
  (a) For $U \subseteq V$, $U^\perp \subseteq V$.

(b) $0^\perp = V$.

(c) $V^\perp = \{0\}$.

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

(e) 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 $e_1, \ldots, e_m$ be the orthonormal basis of U. Any $u \in U$ can be written as

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

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

  Therefore, w is orthogonal to $\operatorname{span}(e_1, \ldots, e_m) = U$, 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 U is the map $P_U : V \to U$ 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, P_U is surjective onto U.

- Also, for orthonormal basis $e_1, \ldots, e_m$, $P_U(v) = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$, 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 V and then use the expression above the 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) $\operatorname{range} P_U = U$; [for $\operatorname{range} P_U \subseteq U$ by definition, and (a) above]

  (d) $v - P_U(v) \in U^\perp$; [true because $w = v - P_U(v)$ is unique in $U^\perp$]

  (e) $\operatorname{null} P_U = U^\perp$; [$U^\perp \subseteq \operatorname{null} P_U$ 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; [by (b)]

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

- 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 $u = P_U(v)$, as expected. Formally, we want to show that for all $u \in U$,
$$\|v - P_U(v)\| \leq \|v - u\|,$$
which achieve 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 $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

7.1 Self-Adjoint and Normal Operators

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

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

The second slot must be the same for all v and thus $(T^*)^* = T$.

As you may have noticed from (a), (b), & (c), the map $^* : \mathcal{L}(V,W) \to \mathcal{L}(W,V)$ 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 $\operatorname{null} T$ and $\operatorname{range} T$. For $T \in \mathcal{L}(V,W)$, we have

  (a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

  (b) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$; [take the orthogonal complement of both sides in (a)]

  (c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$; [replace T^* in (a) by $(T^*)^* = T$]

  (d) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$. [take the orthogonal complement of both sides in (d)]

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

- With respect to an orthonormal basis $e_1, \ldots, e_n$ of V and an orthonormal basis $f_1, \ldots, f_m$ of W, $\mathcal{M}(T^*, (f_1, \ldots, f_m), (e_1, \ldots, e_n))$ is just the conjugate transpose of $\mathcal{M}(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m))$. (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

  $$\langle Te_k, f_1, f\rangle_1 + \cdots + \langle Te_k, f_m, f\rangle_m = Te_k.$$

  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 $1 \le j \le m$ and $1 \le k \le n$. It is clear that $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ are conjugate transposes of one another.

- 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 $\mathbb{C}$ 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

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

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

- (†) 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 $u, w \in V$,

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

  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 $v \in V$, $\langle Tv, v\rangle \in \mathbb{R}$.

  Let us look at the equation

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

  If $\langle Tv, v\rangle \in \mathbb{R}$ 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 real inner product space V. 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 $\operatorname{null} T = \operatorname{null} T^*$.

- 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 $\overline{\lambda}$ is an eigenvalue of T^*:

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

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

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

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

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

- 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, $T^*v=\overline{\beta}v$. 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:

  (a) T is normal.

  (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.

  The equivalency between (b) and (c) is easy (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) $\Longrightarrow$ (c): Given T is normal, by Schur's theorem (since $\mathbb{F} = \mathbb{C}$) we can find an orthonormal basis $e_1, \ldots, e_n$ 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

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

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

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

Schur's theorem gives us a shortcut to prove (a) $\Longrightarrow$ (c) by comparing $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$. By contrast, in the spectral theorem for T on real inner product spaces, (a) $\Longrightarrow$ (c) is not easy to prove because in the first place we do not know over $\mathbb{R}$ (unlike over $\mathbb{C}$) 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 $\mathbb{R}$.

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 if T is self-adjoint, then T must have an eigenvalue, from which we can prove that (a) $\Longrightarrow$ (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 $\mathbb{R}$ instead of $\mathbb{C}$.
  - We can show that for self-adjoint $T \in \mathcal{L}(V)$, $T^2 + bT + CI$ is injective under $b, c \in \mathbb{R}$, $b^2 - 4c < 0$. For an arbitrary nonzero $v \in V$,

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

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

  Once we show $T^2 + bT + CI$ is injective, for any $v \in V$,

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

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 the 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 T 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. $U \coloneqq \operatorname{span}(u)$ 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 $e_1, \ldots, e_n$ corresponding to nonnegative eigenvectors $\lambda_1, \ldots, \lambda_n$. Construct $R \in \mathcal{L}(V)$ such that $Re_j = \sqrt{\lambda_j}e_j$ for every e_j. Then $R^2 e_j = \lambda e_j = Te_j$, so that $R^2 = T$. Show that R is a positive operator 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. 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 theorem says 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 $e_1, \ldots, e_n$ consisting of eigenvectors of the positive operator R. Let $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ be the corresponding eigenvalues. If we now consider

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

  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 $e_1, \ldots, e_n$ is a basis, we have $a_j(\lambda - \lambda_j) = 0$ for all j. Thus,

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

- 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 isometry. 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 $u, v \in V$;

(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) $\Longrightarrow$ (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) $\Longrightarrow$ (c): replace u and v by e_j and e_k

(c) $\Longrightarrow$ (d): trivial

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

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

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

– 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.

- For scalars $\lambda_1, \dots, \lambda_n$ all with absolute value 1 satisfying $Se_j = \lambda_j e_j$ for orthonormal basis $e_1, \dots, e_n$ of V and $S \in \mathcal{L}(V)$, S must be an isometry.

  First, $Sv = \langle v, e_1\rangle Se_1 + \cdots + \langle v, e_n\rangle Se_n = \lambda_1\langle v, e_1\rangle e_1 + \cdots + \lambda_n\langle v, e_n\rangle$. Since the λ's all have absolute value 1, $\|Sv\| = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 = \|v\|^2$.

- 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 $\mathbb{C}$, 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 $\mathbb{C}$, S is an isometry is equivalent to saying that there is an orthonormal basis of V consisting of eigenvalues of S with corresponding eigenvalues all of absolute value 1.

7.4 Polar Decomposition and Singular Value Decomposition

- 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 $z \in \mathbb{C}$, 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 $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ 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 $\operatorname{range} T$ to S on the entire V.

    Because $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ is now bijective, $\dim\sqrt{T^*T} = \dim\operatorname{range} T$, from which we know $\dim(\operatorname{range}\sqrt{T^*T})^\perp = \dim(\operatorname{range} T)^\perp$ because both T^*T and T are linear operators on V. Thus we have an orthonormal basis $e_1, \ldots, e_m$ of $(\operatorname{range}\sqrt{T^*T})^\perp$ and another orthonormal basis $f_1, \ldots, f_m$ of $(\operatorname{range} T)^\perp$. If we let $S_2(e_1) = f_1, \ldots, S_2(e_m) = f_m$, 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 unique written into as the sum of $u \in \operatorname{range}\sqrt{T^*T}$ and $w \in (\operatorname{range}\sqrt{T^*T})^\perp$, 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 value decomposition theorem says the following: suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$, then there exists two orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ of V such that
  $$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$
  for every $v \in V$.

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, $v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$, 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 $f_1 \coloneqq Se_1, \dots, f_n \coloneqq Se_n$ 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.

- 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 theorem 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 showed earlier in proving every positive operator has a unique positive square root. Note that this establishes a one-to-one correspondence between every λ_j and every $\sqrt{\lambda_j}$, and hence $\dim E(\lambda, T^*T) = \dim E(\lambda, \sqrt{T^*T})$.

8 Operators on Complex Vector Spaces

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

8.1 Generalized Eigenvectors and Nilpotent Operators

Remark. In this section n means $\dim V$.

- $\{0\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \operatorname{null} T^2 \cdots \subseteq \operatorname{null} T^k \subseteq \operatorname{null} T^{k+1} \subseteq \cdots$ holds with a quite obvious proof.
- Furthermore, once $\operatorname{null} T^m = \operatorname{null} T^{m+1}$ for some $m \geq 0$, $\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \cdots$ and so on. Correspondingly we need to show that $\operatorname{null} T^{m+k} \supseteq \operatorname{null} T^{m+k+1}$, which is not hard either.

- As a corollary, if $\operatorname{null} T = 0$, then all powers of T are injective.
- When do we hit the first m such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$? This m always exists, and $m \le \dim V$.

- For T on V with dimension n, we have $\operatorname{null} T^n = \operatorname{null} T^{n+1} = \cdots$.

  For the proof, assume that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$, then by the previous two claims we have

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

  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 $\operatorname{null} T^n$ is at least of dimension n, from which we can increase no further.

  From now on we will call

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

  the *nullspace power chain*.

- Say T is on V with dimension n, $V = \operatorname{null} T^n \oplus \operatorname{range} T^n$. (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 $\operatorname{null} T^n \cap \operatorname{range} T^n = \{0\}$. Second, we need to show that the direct sum is actually V.

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

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

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

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

- We now generalizes 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 λ is a nonzero vector $v \in V$ such that $(T - \lambda I)^j v = 0$ for some $j \in \mathbb{Z}^+$. (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 $E(\lambda, T) \subseteq G(\lambda, T)$. When λ is not an eigenvalue, $G(\lambda, T) = \{0\}$.

- We now state that $G(\lambda, T) = \operatorname{null}(T - \lambda I)^n$, where $n = \dim V$. (In particular, $G(0, T) = \operatorname{null} T^n$) 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 $\lambda_1, \ldots, \lambda_m$ be the list of eigenvalues and $v_1, \ldots, v_m$ be the list of corresponding eigenvectors we are considering. Suppose $0 = a_1 v_1 + \cdots + a_m v_m$.

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

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

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 is *nilpotent* if $N^k = 0$ for some $k \in \mathbb{Z}^+$. A nilpotent operator cannot be injective/surjective. [Recall the composition of injective (resp. surjective/bijective) functions must be injective (resp. surjective/bijective).]
  - The only eigenvalue of a nilpotent operator is 0. Proof is obvious.
- Given nilpotent N, $N^{\dim V} = 0$, again by the nullspace power chain.
- $\mathcal{M}(N)$ 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 $\operatorname{null} N$ to a basis of $\operatorname{null} N^2$, which can be extended to $\operatorname{null} N^3$ and so on. The first few columns concerning $N(\text{basis vector of } \operatorname{null} N)$ must be all 0. The next few columns concerning $N(\text{basis vectors of } \operatorname{null} N^2)$ have nonzero entries strictly above the diagonal because

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

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

8.2 Decomposition of an Operator

8.3 Characteristic and Minimal Polynomials

8.4 Jordan Form

9 Operators on Real Vector Spaces

9.1 Complexification

9.2 Operators on Real Inner Product Spaces

10 Trace and Determinant

10.1 Trace

10.2 Determinant