MATH 110 based on *Linear Algebra Done Right*

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.

- 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 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 contain 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 notion of a list of vectors. FDVS is defined in terms of spanning list (set), 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, $\mathrm{span}(\emptyset) := \{0\}$.

- Similar to the fact that the sum $U_1 + U_2 + \cdots + U_m$ is the smallest vector space containing all U_i's, $\mathrm{span}(S)$ 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 in $\mathbb{F}$ is the **expression** $a_n z^n + \cdots + a_1 z + a_0$ with all 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 customarily define its degree to be $-\infty$ in the book, or -1 elsewhere.

  - Since we are dealing with z and coefficients in $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ in this book, a polynomial here can be alternatively defined as the **function** $p : \mathbb{F} \to \mathbb{F}$ with $p(z) = a_n z^n + \cdots + a_1 z + a_0$. 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. 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 $T \backslash S$. 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 the linearly independent S but does not belong to $\mathrm{span}(S)$, then the resulting set is still linearly independent.

  - 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, $\operatorname{span}(B) = \operatorname{span}(S) = V$ because all the v's left out are in $\operatorname{span}(B)$.

  ∗ 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, $\operatorname{span}(B) = \operatorname{span}(S)$ 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*. From now on, V is by default 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 remain 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 $\operatorname{span}(L)$. 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, that they are the same space. This is an important result useful in many proofs.

- There are two more things that deserves 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 essentially different, and this cannot be extended 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 $[T_1(T_2 T_3) = (T_1 T_2) T_3]$ and the identity works as usual $(I_W T = T I_V = T)$. The distributive properties $((S_1 + S_2)T = S_1 T + S_2 T$ 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* null T is the preimage of 0_W. The *range/image* 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 surjectivity.
  - A simple theorem to always keep in mind is that the mapping of the basis vectors of V spans range T.

- We are now endowed with all the tools to prove the **rank-nullity theorem**: for a FDVS V and T from V to W, range T is finite-dimensional with

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

  - As always, we need to construct a basis to work with first. Let $u_1, u_2, \ldots, u_n$ be a basis of 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 range T. The fact that the u's are in the nullspace shows that $T(v_1), \ldots, T(v_n)$ spans V. To show the elements are linearly independent, consider that

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

  implies $a_1 v_1 + \cdots + a_n v_n \in$ null T. Yet we also have $b_1 u_1 + \cdots + b_n u_n \in$ 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 injectivity and surjectivity. First of all, there is a criterion that says T is injective iff null $T = 0$ (proof follows straight from the definition of nullspace). The results below follow.

5

- A map to a smaller dimensional space cannot be injective because dim (null T) > 0.

- A map to a large dimensional space cannot be surjective because dim (range T) $\leq$ dim $V <$ dim W.

- V and W are FDVS of equal dimension, then

$$T \text{ is injective } \Longleftrightarrow T \text{ is surjective } \Longleftrightarrow \text{rank } T = \dim V.$$

- 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

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

  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. Injectivity gives us nullity $= 0$ and surjectivity 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 $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

7

- 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 $v_1, \ldots, v_n$, then for any $v \in V$, we define

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

  where $a_1, \ldots, a_n$ 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, $\mathcal{M}(T)_{.,k} = \mathcal{M}(Tv_k)$.

  - 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,

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

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

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

  For a vector $v \in V$, its image $\mathcal{M}(v)$ under the isomorphism $\mathcal{M}$ of V belongs to $\mathbb{F}^{n,1}$. 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.

- 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.)

  (a) $\implies$ (b) by the theorem we proved earlier.

  (b) $\implies$ (c) by checking the rank-nullity theorem. Injectivity means nullity $T = 0$, which implies rank $T = \dim V$.

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