Math 2370

Huiqiang Jiang

1. Fundamentals of Linear Spaces

A *field* K is a nonempty set together with two operations, usually called addition and multiplication, and denoted by $+$ and $\cdot$ respectively, such that the following axioms hold:
1. Closure of K under addition and multiplication: $a, b \in K \Longrightarrow a + b,\ a \cdot b \in K$;
2. Associativity of addition and multiplication: For any $a, b, c \in K$,

$$a + (b + c) = (a + b) + c,\ a \cdot (b \cdot c) = (a \cdot b) \cdot c;$$

3. Commutativity of addition and multiplication: For any $a, b \in K$,

$$a + b = b + a,\ a \cdot b = b \cdot a;$$

4. Existence of additive and multiplicative identity elements: There exists an element of K, called the additive identity element and denoted by 0, such that for all $a \in K$, $a + 0 = a$. Likewise, there is an element, called the multiplicative identity element and denoted by 1, such that for all $a \in K$, $a \cdot 1 = a$. To exclude the trivial ring, the additive identity and the multiplicative identity are required to be distinct.
5. Existence of additive inverses and multiplicative inverses: For every $a \in K$, there exists an element $-a \in K$, such that

$$a + (-a) = 0.$$

Similarly, for any $a \in K \setminus \{0\}$, there exists an element $a^{-1} \in K$, such that $a \cdot a^{-1} = 1$. We can define subtraction and division operations by

$$a - b = a + (-b) \text{ and } \frac{a}{b} = a \cdot b^{-1} \text{ if } b \neq 0.$$

6. Distributivity of multiplication over addition: For any $a, b, c \in K$,

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Examples of field: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$. In our lecture, K will be either $\mathbb{R}$ or $\mathbb{C}$, the elements in K are called scalars.

A *linearspace* X over a field K is a set in which two operations are defined: Addition, denoted by $+$ such that

$$x, y \in X \Longrightarrow x + y \in X$$

and scalar multiplication such that

$$a \in K \text{ and } x \in X \Longrightarrow ax \in X.$$

These two operations satisfy the following axioms:
1. Associativity of addition:

$$x + (y + z) = (x + y) + z;$$

2. Commutativity of addition:

$$x + y = y + x;$$

3. Identity element of addition: There exists an element $0 \in X$, called the zero vector, such that $x + 0 = x$ for all $x \in X$.
4. Inverse elements of addition: For every $x \in X$, there exists an element $-x \in X$, called the additive inverse of x, such that

$$x + (-x) = 0.$$

5. Compatibility (Associativity) of scalar multiplication with field multiplication: For any $a, b \in K$, $x \in X$,

$$a\,(bx) = (ab)\,x.$$

6. Identity element of scalar multiplication: $1x = x$.
7. Distributivity of scalar multiplication with respect to vector addition:

$$a\,(x + y) = ax + ay.$$

8. Distributivity of scalar multiplication with respect to field addition:

$$(a + b)\,x = ax + bx.$$

The elements in a linear space are called vectors.

Remark 1. *Zero vector is unique.*

Remark 2. $0x = 0$, $(-1)\,x = -x$.

Example 1. $\mathbb{R}^n$, $\mathbb{C}^n$.

Example 2. *Polynomials with real coefficients of order at most* n.

Definition 1. *A one-to-one correspondence between two linear spaces over the same field that maps sums into sums and scalar multiples into scalar multiples is called an isomorphism.*

Example 3. *The linear space of real valued functions on* $\{1, 2, \cdots, n\}$ *is isomorphic to* $\mathbb{R}^n$.

Definition 2. *A subset* Y *of a linear space* X *is called a subspace if sums and scalar multiples of elements of* Y *belong to* Y.

The set $\{0\}$ consisting of the zero element of a linear space X is a subspace of X. It is called the trivial subspace.

Definition 3. *The sum of two subsets Y and Z of a linear space X, is the set defined by*

$$Y + Z = \{y + z \in X : y \in Y, z \in Z\}.$$

The intersection of two subsets Y and Z of a linear space X, is the set defined by

$$Y \cap Z = \{x \in X : x \in Y, x \in Z\}.$$

Proposition 1. *If Y and Z are two linear subspaces of X, then both $Y + Z$ and $Y \cap Z$ are linear subspaces of X.*

Remark 3. *The union of two subspaces may not be a subspace.*

Definition 4. *A linear combination of m vectors $x_1, \cdots, x_m$ of a linear space is a vector of the form*

$$\sum_{j=1}^{m} c_j x_j \text{ where } c_j \in K.$$

Given m vectors $x_1, \cdots, x_m$ of a linear space X, the set of all linear combinations of $x_1, \cdots, x_m$ is a subspace of X, and it is the smallest subspace of X containing $x_1, \cdots, x_m$. This is called the subspace spanned by $x_1, \cdots, x_m$.

Definition 5. *A set of vectors $x_1, \cdots, x_m$ in X spans the whole space X if every x in X can be expressed as a linear combination of $x_1, \cdots, x_m$.*

Definition 6. *The vectors $x_1, \cdots, x_m$ are called linearly dependent if there exist scalars $c_1, \cdots, c_m$, not all of them are zero, such that*

$$\sum_{j=1}^{m} c_j x_j = 0.$$

The vectors $x_1, \cdots, x_m$ are called linearly independent if they are not dependent.

Definition 7. *A finite set of vectors which span X and are linearly independent is called a basis for X.*

Proposition 2. *A linear space which is spanned by a finite set of vectors has a basis.*

Definition 8. *A linear space X is called finite dimensional if it has a basis.*

Theorem 1. *All bases for a finite-dimensional linear space X contain the same number of vectors. This number is called the dimension of X and is denoted as* $\dim X$.

Proof. The theorem follows from the lemma below. □

Lemma 1. *Suppose that the vectors $x_1, \cdots, x_n$ span a linear space X and that the vectors $y_1, \cdots, y_m$ in X are linearly independent. Then $m \leq n$.*

Proof. Since $x_1, \cdots, x_n$ span X, we have

$$y_1 = \sum_{j=1}^{n} c_j x_j.$$

We claim that not all c_j are zero, otherwise $y_1 = 0$ and $y_1, \cdots, y_m$ must be linearly dependent. Suppose $c_k \neq 0$, then x_k can be expressed as a linear combination of y_k and the remaining x_j. So the set consisting of the x_j's, with x_k replaced by y_k span

X. If $m \geq n$, repeat this step $n-1$ more times and conclude that $y_1, \cdots, y_n$ span X. If $m > n$, this contradicts the linear independence of the vectors $y_1, \cdots, y_m$. $\square$

We define the dimension of the trivial space consisting of the single element 0 to be zero.

THEOREM 2. *Every linearly independent set of vectors $y_1, \cdots, y_m$ in a finite dimensional linear space X can be completed to a basis of X.*

THEOREM 3. *Let X be a finite dimensional linear space over K with $\dim X = n$, then X is isomorphic to K^n.*

THEOREM 4. *(a) Every subspace Y of a finite-dimensional linear space X is finite dimensional.*
(b) Every subspace Y has a complement in X, that is, another subspace Z such that every vector x in X can be decomposed uniquely as

$$x = y + z,\ y \in Y,\ z \in Z.$$

Furthermore $\dim X = \dim Y + \dim Z$.

X is said to be the direct sum of two subspaces Y and Z that are complements of each other. More generally X is said to be the direct sum of its subspaces $Y_1, \cdots, Y_m$ if every x in X can be expressed uniquely as

$$x = \sum_{j=1}^{m} y_j \text{ where } y_j \in Y_j.$$

This relation is denoted as

$$X = Y_1 \oplus Y_2 \oplus \cdots \oplus Y_m.$$

If X is finite dimensional and

$$X = Y_1 \oplus Y_2 \oplus \cdots \oplus Y_m,$$

then

$$\dim X = \sum_{j=1}^{m} \dim Y_j.$$

DEFINITION 9. *For X a linear space, Y a subspace, we say that two vectors x_1, x_2 in X are congruent modulo Y, denoted*

$$x_1 \equiv x_2 \ mod\ Y$$

if $x_1 - x_2 \in Y$.

Congruence mod Y is an equivalence relation, that is, it is
(i) symmetric: if $x_1 \equiv x_2$, then $x_2 \equiv x_1$.
(ii) reflexive: $x \equiv x$ for all x in X.
(iii) transitive: if $x_1 \equiv x_2$ and $x_2 \equiv x_3$, then $x_1 \equiv x_3$.

We can divide elements of X into congruence classes mod Y. The congruence class containing the vector x is the set of all vectors congruent with X; we denote it by $\{x\}$.

The set of congruence classes can be made into a linear space by defining addition and multiplication by scalars, as follows:

$$\{x\} + \{y\} = \{x + y\},$$
$$a\{x\} = \{ax\}.$$

The linear space of congruence classes defined above is called the quotient space of X mod Y and is denoted as X/Y.

Remark 4. *X/Y is not a subspace of X.*

Theorem 5. *If Y is a subspace of a finite-dimensional linear space X; then*

$$\dim Y + \dim (X/Y) = \dim X.$$

Proof. Let $x_1, \cdots, x_m$ be a basis for Y, $m = \dim Y$. This set can be completed to form a basis for X by adding $x_{m+1}, \cdots, x_n$, $n = \dim X$. We claim that $\{x_{m+1}\}, \cdots, \{x_n\}$ form a basis for X/Y by verifying that they are linearly independent and span the whole space X/Y. □

Theorem 6. *Suppose X is a finite-dimensional linear space, U and V two subspaces of X. Then we have*

$$\dim (U + V) = \dim U + \dim V - \dim (U \cap V).$$

Proof. If $U \cap V = \{0\}$, then $U + V$ is a direct sum and hence

$$\dim (U + V) = \dim U + \dim V.$$

In general, let $W = U \cap V$, we claim $U/W + V/W = (U + V)/W$ is a direct sum and hence

$$\dim (U/W) + \dim (V/W) = \dim ((U + V)/W).$$

Applying Theorem 5, we have

$$\dim (U + V) = \dim U + \dim V - \dim (U \cap V).$$

□

Definition 10. *The Cartesian sum $X_1 \oplus X_2$ of two linear spaces X_1, X_2 over the same field is the set of pairs (x_1, x_2) where $x_i \in X_i$, $i = 1, 2$. $X_1 \oplus X_2$ is a linear space with addition and multiplication by scalars defined componentwisely.*

Theorem 7.

$$\dim X_1 \oplus X_2 = \dim X_1 + \dim X_2.$$

More generally, we can define the Cartesian sum $\oplus_{k=1}^m X_k$ of m linear spaces X_1, X_2, $\cdots$, X_m, and we have

$$\dim \oplus_{k=1}^m X_k = \sum_{k=1}^m \dim X_k.$$

2. Dual Spaces

Let X be a linear space over a field K. A scalar valued function $l : X \to K$ is called linear if

$$l(x+y) = l(x) + l(y)$$

for all x, y in X, and $l(kx) = kl(x)$ for $\forall x \in X$ and $\forall k \in K$.

The set of linear functions on a linear space X forms a linear space X', the dual space of X, if we define

$$(l+m)(x) = l(x) + m(x)$$

and

$$(kl)(x) = k(l(x)).$$

Theorem 8. *Let X be a linear space of dimension n. Under a chosen basis $x_1, \cdots, x_n$, the elements x of X can be represented as arrays of n scalars:*

$$x = (c_1, \cdots, c_n) = \sum_{k=1}^{n} c_k x_k.$$

Let $a_1, \cdots, a_n$ be any array of n scalars; the function l defined by

$$l(x) = \sum_{k=1}^{n} a_k c_k$$

is a linear function of X. Conversely, every linear function l of X can be so represented.

Theorem 9.

$$\dim X' = \dim X.$$

We write

$$(l, x) \equiv l(x)$$

which is a bilinear function of l and x. The dual of X' is X'', consisting of all linear functions on X'.

Theorem 10. *The bilinear function $(l, x) = l(x)$ gives a natural identification of X with X''. The map $x \mapsto x^{**}$ is an isomorphism where*

$$(x^{**}, l) = (l, x)$$

for any $l \in X^*$.

Definition 11. *Let Y be a subspace of X. The set of linear functions that vanish on Y, that is, satisfy*

$$l(y) = 0 \text{ for all } y \in Y,$$

is called the annihilator of the subspace Y; it is denoted by $Y^\perp$.

Theorem 11. *$Y^\perp$ is a subspace of X' and*

$$\dim Y^\perp + \dim Y = \dim X.$$

Proof. We shall establish a natural isomorphism $T : Y^{\perp} \to (X/Y)'$: For any $l \in Y^{\perp} \subset X'$, we define for any $\{x\} \in X/Y$,

$$(Tl)(\{x\}) = l(x).$$

Then $Tl \in (X/Y)'$ is well defined. One can verify that T is an isomorphism. Hence

$$\dim Y^{\perp} = \dim\left((X/Y)'\right) = \dim(X/Y) = \dim X - \dim Y.$$

□

The dimension of $Y^{\perp}$ is called the co-dimension of Y as a subspace of X.

$$\operatorname{codim} Y + \dim Y = \dim X.$$

Since $Y^{\perp}$ is a subspace of X', its annihilator, denoted by $Y^{\perp\perp}$, is a subspace of X''.

Theorem 12. *Under the natural identification of X'' and X, for every subspace Y of a finite-dimensional space X,*

$$Y^{\perp\perp} = Y.$$

Proof. Under the identification of X'' and X, $Y \subset Y^{\perp\perp}$. Now $\dim Y^{\perp\perp} = \dim Y$ implies $Y^{\perp\perp} = Y$. □

More generally, let S be a subset of X. The annihilator of S is defined by

$$S^{\perp} = \{l \in X' : l(x) = 0 \text{ for any } x \in S\}.$$

Theorem 13. *Let S be a subset of X.*

$$S^{\perp} = (\operatorname{span} S)^{\perp}.$$

Theorem 14. *Let $t_1, t_2, \cdots, t_n$ be n distinct real numbers. For any finite interval I on the real axis, there exist n numbers $m_1, m_2, \cdots, m_n$ such that*

$$\int_I p(t)\,dt = \sum_{k=1}^{n} m_k p(t_k)$$

holds for all polynomials p of degree less than n.

3. Linear Mappings

Let X, U be linear spaces over the same field K. A mapping $T : X \to U$ is called linear if it is additive:

$$T(x+y) = T(x) + T(y), \text{ for any } x, y \in X.$$

and if it is homogeneous:

$$T(kx) = kT(x) \text{ for any } k \in K \text{ and } x \in X.$$

For simplicity, we often write $T(x) = Tx$.

Example 4. *Isomorphisms are linear mappings.*

Example 5. *Differentiation from $P_n(t)$ tot $P_{n-1}(t)$ is linear.*

Example 6. *Linear functionals are linear mappings.*

Theorem 15. *The image of a subspace of X under a linear map T is a subspace of U. The inverse image of a subspace of U, is a subspace of X.*

Definition 12. *The range of T is the image of X under T; it is denoted as R_T. The null-space of T is the inverse image of $\{0\}$, denoted as N_T. $R_T \subset U$ and $N_T \subset X$ are subspaces.*

Definition 13. $\dim R_T$ *is called the rank of the mapping T and* $\dim N_T$ *is called the nullity of the mapping T.*

Theorem 16 (Rank-Nullity Theorem). *Let $T : X \to U$ be linear. Then*

$$\dim R_T + \dim N_T = \dim X.$$

Proof. Define $\tilde{T} : X/N_T \to R_T$ so that

$$\tilde{T}\{x\} = Tx.$$

Then T is well defined and it is an isomorphism. Hence

$$\dim R_T = \dim X/N_T = \dim X - \dim N_T.$$

□

Corollary 1. *Let $T : X \to U$ be linear.*
(a) Suppose $\dim U < \dim X$*, then $Tx = 0$ for some $x \neq 0$.*
(b) Suppose $\dim U = \dim X$*, and the only vector satisfying $Tx = 0$ is $x = 0$. Then $R_T = U$ and T is an isomorphism.*

Corollary 2. *Suppose $m < n$, then for any real numbers t_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, the system of linear equations*

$$\sum_{j=1}^{n} t_{ij} x_j = 0, 1 \leq i \leq m$$

has a nontrivial solution.

Corollary 3. *Given n^2 real numbers t_{ij}, $1 \leq i, j \leq n$, the inhomogeneous system of linear equations*

$$\sum_{j=1}^{n} t_{ij} x_j = u_i, 1 \leq i \leq n$$

has a unique solution for any u_i, $1 \le i \le n$ *if and only if the homogeneous system*

$$\sum_{j=1}^{n} t_{ij} x_j = 0, 1 \le i \le n$$

has only the trivial solution.

We use $L(X, U)$ to denote the collection of all linear maps from X to U. Suppose that $T, S \in L(X, U)$, we define their sum $T + S$ by

$$(T + S)(x) = Tx + Sx \text{ for any } x \in X$$

and we define, for $k \in K$, kT by

$$(kT)(x) = kTx \text{ for any } x \in X.$$

Then $T + S$, $kT \in L(X, U)$ and $L(X, U)$ is a linear space.

Let $T \in L(X, U)$ and $S \in L(U, V)$, we can define the composition of T with S by

$$S \circ T(x) = S(Tx).$$

Note that composition is associative: if $R \in L(V, Z)$, then

$$R \circ (S \circ T) = (R \circ S) \circ T.$$

Theorem 17. *(i) The composite of linear mappings is also a linear mapping. (ii) Composition is distributive with respect to the addition of linear maps, that is,*

$$(R + S) \circ T = R \circ T + S \circ T$$

whenever the compositions are defined.

Remark 5. *We use* ST *to denote* $S \circ T$*, called the multiplication of* S *and* T*. Note that* $ST \neq TS$ *in general.*

Definition 14. *A linear map is called invertible if it is* 1*-to-*1 *and onto, that is, if it is an isomorphism. The inverse is denoted as* T^{-1}*.*

Theorem 18. *(i) The inverse of an invertible linear map is linear.*
(ii) If S *and* T *are both invertible, and if* $ST = S \circ T$ *is defined, then* ST *also is invertible, and*

$$(ST)^{-1} = T^{-1} S^{-1}.$$

Definition 15. *Let* $T \in L(X, U)$*, the transpose* $T' \in L(U', X')$ *of* T *is defined by*

$$(T'(l))(x) = l(Tx) \text{ for any } l \in U' \text{ and } x \in X.$$

We could use the dual notation to rewrite the above identity as

$$(T'l, x) = (l, Tx).$$

Theorem 19. *Whenever defined, we have*

$$(ST)' = T'S',$$
$$(T + R)' = T' + R',$$
$$\left(T^{-1}\right)' = (T')^{-1}.$$

Example 7. *Let $T:\mathbb{R}^n\to\mathbb{R}^m$ be defined by $y=Tx$ where*

$$y_{i=}\sum_{j=1}^{n}t_{ij}x_j.,1\le i\le m.$$

Identifying $(\mathbb{R}^n)^{'}=\mathbb{R}^n$ and $(\mathbb{R}^m)^{'}=\mathbb{R}^m$, $T':\mathbb{R}^m\to\mathbb{R}^n$ is defined by $v=T'u$ where

$$v_j=\sum_{i=1}^{m}t_{ij}u_i,1\le j\le n.$$

Theorem 20. *Let $T\in L(X,U)$. Identifying $X''=X$ and $U''=U$. We have $T''=T$.*

Theorem 21. *Let $T\in L(X,U)$.*

$$R_T^{\perp}=N_{T'},$$
$$R_T=N_{T'}^{\perp}.$$

Theorem 22. *Let $T\in L(X,U)$.*

$$\dim R_T=\dim R_{T'}.$$

Corollary 4. *Let $T\in L(X,U)$. Suppose that $\dim X=\dim U$, then*

$$\dim N_T=\dim N_{T'}.$$

We now consider $L(X,X)$ which forms an algebra if we define the multiplication as composition.

The set of invertible elements of $L(X,X)$ forms a group under multiplication. This group depends only on the dimension of X, and the field K of scalars. It is denoted as $GL(n,K)$ where $n=\dim X$.

Given an invertible element $S\in L(X,X)$, we assign to each $M\in L(X,X)$ the element $M_S=SMS^{-1}$. This assignment $M\to M_S$ is called a similarity transformation; M is said to be similar to M_S.

Theorem 23. *(a) Every similarity transformation is an automorphism of $L(X,X)$:*

$$(kM)_S=kM_S,$$
$$(M+K)_S=M_S+K_S,$$
$$(MK)_S=M_SK_S.$$

(b) The similarity transformations form a group with

$$(M_S)_T=M_{TS}.$$

Theorem 24. *Similarity is an equivalence relation; that is, it is:*
(i) Reflexive. M is similar to itself.
(ii) Symmetric. If M is similar to K, then K is similar to M.
(iii) Transitive. If M is similar to K, and K is similar to L, then M is similar to L.

Theorem 25. *If either A or B in $L(X,X)$ is invertible, then AB and BA are similar.*

Definition 16. *A linear mapping $P\in L(X,X)$ is called a projection if it satisfies $P^2=P$.*

THEOREM 26. *Let* $P \in L(X, X)$ *be a projection. Then*

$$X = N_P \oplus R_P.$$

And P *restricted on* R_P *is the identity map.*

DEFINITION 17. *The commutator of two mappings* A *and* B *of* X *into* X *is* $AB - BA$. *Two mappings of* X *into* X *commute if their commutator is zero.*

4. Matrices

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $y = Tx$ where

$$y_{i=} \sum_{j=1}^{n} t_{ij} x_j., 1 \leq i \leq m.$$

Then T is a linear map. On the other hand, every map $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ can be written in this form. Actually, t_{ij} is the ith component of Te_j, where $e_j \in \mathbb{R}^n$ has jth component 1, all others 0.

We write

$$T = (t_{ij})_{m\times n} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{pmatrix},$$

which is called an m by n $(m \times n)$ matrix, m being the number of rows, n the number of columns. A matrix is called a square matrix if $m = n$. The numbers t_{ij} are called the entries of the matrix T.

A matrix T can be thought of as a row of column vectors, or a column of row vectors:

$$T = (c_1, \cdots, c_n) = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}$$

where

$$c_j = \begin{pmatrix} t_{1j} \\ \vdots \\ t_{mj} \end{pmatrix} \text{ and } r_i = (t_{i1}, \cdots, t_{in}).$$

Thus

$$Te_j^n = c_j = \sum_{i=1}^{m} t_{ij} e_i^m$$

where we write vectors in $\mathbb{R}^m$ as column vectors.

Since matrices represent linear mappings, the algebra of linear mappings induces a corresponding algebra of matrices:

$$T + S = (t_{ij} + s_{ij})_{m\times n},$$
$$kT = (kt_{ij})_{m\times n}.$$

If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $S \in L(\mathbb{R}^m, \mathbb{R}^l)$, then the product $ST = S \circ T \in L(\mathbb{R}^n, \mathbb{R}^l)$. For $e_j \in \mathbb{R}^n$,

$$STe_j^n = St_{ij}e_i^m = t_{ij}Se_i^m = t_{ij}s_{ki}e_k^l = \left(\sum_{i=1}^{m} s_{ki}t_{ij}\right) e_k^l,$$

hence $(ST)_{kj} = \sum_{i=1}^{m} s_{ki}t_{ij}$ which is the product of kth row of S and jth column of T.

Remark 6. *If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $x \in \mathbb{R}^n$, we can also view Tx as the product of two matrices.*

We can write any $n \times n$ matrix A in 2×2 block form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11} is an $k \times k$ matrix and A_{22} is an $(n-k) \times (n-k)$ matrix. Product of block matrices follows the same formula.

The dual of the space $\mathbb{R}^n$ of all column vectors with n components is the space $(\mathbb{R}^n)'$ of all row vectors with n components. Here for $l \in (\mathbb{R}^n)'$ and $x \in \mathbb{R}^n$,

$$lx = \sum_{i=1}^{n} l_i x_i.$$

Let $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $T' \in L\left((\mathbb{R}^m)', (\mathbb{R}^n)'\right)$. Identifying $(\mathbb{R}^n)' = \mathbb{R}^n$ and $(\mathbb{R}^m)' = \mathbb{R}^m$, $T' : \mathbb{R}^m \to \mathbb{R}^n$ has matrix representation T^T, called the transpose of matrix T,

$$\left(T^T\right)_{ij} = T_{ji}.$$

THEOREM 27. *Let $T \in L(\mathbb{R}^n, \mathbb{R}^m)$. The range of T consists of all linear combinations of the columns of the matrix T.*

The $\dim R_T$ is called the column rank of T and $\dim R_{T^T}$ is called the row rank of T. We have $\dim R_T = \dim R_{T^T}$.

Any $T \in L(X, U)$ can be represented by a matrix once we choose bases for X and U. A choice of basis in X defines an isomorphism $B : X \to \mathbb{R}^n$, and similarly we have isomorphism $C : U \to \mathbb{R}^m$. We have $M = CTB^{-1} \in L(\mathbb{R}^n, \mathbb{R}^m)$ which can be represented by a matrix.

If $T \in L(X, X)$, and $B : X \to \mathbb{R}^n$ is an isomorphism, then we have $M = BTB^{-1} \in L(\mathbb{R}^n, \mathbb{R}^n)$ is a square matrix. Let $C : X \to \mathbb{R}^n$ be an isomorphism, then $N = CTC^{-1}$ is another square matrix representing T. Since

$$N = CB^{-1}MBC^{-1},$$

M, N are similar. Similar matrices represents the same mapping under different bases.

DEFINITION 18. *Invertible and singular matrices; Unit matrix I; Upper triangular matrix, lower triangular matrix and diagonal matrix.*

DEFINITION 19. *A square matrix T is called a tridiagonal matrix if $t_{ij} = 0$ whenever $|i - j| > 1$.*

Gaussian elimination can be used to solve linear equations.

5. Determinant and Trace

A simplex in $\mathbb{R}^n$ is a polyhedron with $n+1$ vertices. The simplex is ordered if we have an order for the vertices. We denote the first vertex to be the origin and denote the rest in its order of $a_1, a_2, \cdots, a_n$. Our goal is to define the volume of an ordered simplex.

An ordered simplex S is called degenerate if it lies on an $(n-1)$-dimensional subspace. An ordered nondegenerate simplex

$$S = (0, a_1, a_2, \cdots, a_n)$$

is called positively oriented if it can be deformed continuously and nondegenerately into the standard ordered simplex $(0, e_1, e_2, \cdots, e_n)$, where e_j is the jth unit vector in the standard basis of $\mathbb{R}^n$. Otherwise, we say it is negatively oriented.

For a nondegenerate oriented simplex S we define $O(S) = +1$ (-1) if it is positively (negatively) oriented. For a degenerate simplex S, we set $O(S) = 0$.

The volume of a simplex S is given inductively by the elementary formula

$$\mathrm{Vol}(S) = \frac{1}{n} \mathrm{Vol}(\mathrm{Base}) \times \mathrm{Altitude}.$$

And the signed volume of an ordered simplex S is

$$\Sigma(S) = O(S) \, \mathrm{Vol}(S).$$

We view $\Sigma(S)$ as a function of vectors $(a_1, a_2, \cdots, a_n)$:
1. $\Sigma(S) = 0$ if $a_j = a_k$ for some $k = j$.
2. $\Sigma(S)$ is linear on a_j if we fix other vertices.

Definition 20. *Let $A = (a_1, a_2, \cdots, a_n)$ be a square matrix, where $a_k \in \mathbb{R}^n$, $1 \leq k \leq n$ are column vectors. Its determinant is defined by*

$$\det A = D(a_1, a_2, \cdots, a_n) = n! \Sigma(S)$$

where $S = (0, a_1, a_2, \cdots, a_n)$.

Theorem 28. *(i) $D(a_1, a_2, \cdots, a_n) = 0$ if $a_j = a_k$ for some $k = j$.*
(ii) $D(a_1, a_2, \cdots, a_n)$ is a multilinear function of its arguments.
(iii) Normalization: $D(e_1, e_2, \cdots, e_n) = 1$.
(iv) D is an alternating function of its arguments, in the sense that if a_i and a_j are interchanged, $i \neq j$, the value of D changes by the factor (-1).
(v) If $a_1, a_2, \cdots, a_n$ are linearly dependent, then $D(a_1, a_2, \cdots, a_n) = 0$.

Proof. (iv)

$$\begin{aligned} D(a, b) &= D(a, a) + D(a, b) = D(a, a+b) \\ &= D(a, a+b) - D(a+b, a+b) \\ &= -D(b, a+b) = -D(b, a). \end{aligned}$$

□

Next we introduce the concept of permutation. A permutation is a mapping p of n objects, say the numbers $1, 2, \cdots, n$ onto themselves. Permutations are invertible and they form a group with compositions. These groups, except for $n = 2$, are noncommutative.

Let $x_1, \cdots, x_n$ be n variables; their discriminant is defined to be

$$P(x_1, \cdots, x_n) = \prod_{i<j} (x_i - x_j).$$

Let p be any permutation. Clearly,

$$\prod_{i<j} \left(x_{p(i)} - x_{p(j)}\right)$$

is either $P(x_1, \cdots, x_n)$ or $-P(x_1, \cdots, x_n)$.

Definition 21. *The signature $\sigma(p)$ of a permutation p is defined by*

$$P\left(x_{p(1)}, \cdots, x_{p(n)}\right) = \sigma(p) P(x_1, \cdots, x_n).$$

Hence, $\sigma(p) = \pm 1$.

Theorem 29.

$$\sigma(p_1 \circ p_2) = \sigma(p_1) \sigma(p_2).$$

Proof.

$$\begin{aligned}
\sigma(p_1 \circ p_2) &= \frac{P\left(x_{p_1 p_2(1)}, \cdots, x_{p_1 p_2(n)}\right)}{P(x_1, \cdots, x_n)} \\
&= \frac{P\left(x_{p_1 p_2(1)}, \cdots, x_{p_1 p_2(n)}\right)}{P\left(x_{p_2(1)}, \cdots, x_{p_2(n)}\right)} \cdot \frac{P\left(x_{p_2(1)}, \cdots, x_{p_2(n)}\right)}{P(x_1, \cdots, x_n)} \\
&= \sigma(p_1) \sigma(p_2).
\end{aligned}$$

□

Given any pair of indices, $j \neq k$, we can define a permutation p such that $p(i) = i$ for $i \neq j$ or k, $p(j) = k$ and $p(k) = j$. Such a permutation is called a transposition.

Theorem 30. *The signature of a transposition t is -1. Every permutation p can be written as a composition of transpositions.*

Proof. By induction. □

We have $\sigma(p) = 1$ if p is a composition of even number of transpositions and p is said to be an even permutation. We have $\sigma(p) = -1$ if p is a composition of odd number of transpositions and p is said to be an odd permutation.

Theorem 31. *Assume that for $1 \leq k \leq n$,*

$$a_k = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix} \in \mathbb{R}^n.$$

The determinant

$$D(a_1, \cdots, a_n) = \sum_p \sigma(p) a_{p(1)1} a_{p(2)2} \cdots a_{p(n)n}$$

where the summation is over all permutations.

PROOF.

$$\begin{aligned} D(a_1, \cdots, a_n) &= D\left(\sum_{j=1}^n a_{j1} e_j, \cdots, \sum_{j=1}^n a_{jn} e_j\right) \\ &= \sum_{1 \le j_k \le n, 1 \le k \le n} a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} D(e_{j_1}, \cdots, e_{j_n}) \\ &= \sum_p \sigma(p)\, a_{p(1)1} a_{p(2)2} \cdots a_{p(n)n} \end{aligned}$$

□

REMARK 7. *Determinant is defined by properties 1,2,3 in Theorem 28.*

THEOREM 32.

$$\det A^T = \det A.$$

THEOREM 33. *Let A, B be two $n \times n$ matrices.*

$$\det(BA) = \det A \det B.$$

PROOF. Let $A = (a_1, \cdots, a_n)$.

$$\det(BA) = D(Ba_1, \cdots, Ba_n).$$

Assuming that $\det B \neq 0$, we define

$$C(a_1, \cdots, a_n) = \frac{\det(BA)}{\det B} = \frac{D(Ba_1, \cdots, Ba_n)}{\det B}.$$

We verify that C satisfies properties 1,2,3 in Theorem 28. Hence $C = D$.

When $\det B = 0$, we could do approximation $B(t) = B + tI$. □

COROLLARY 5. *An $n \times n$ matrix A is invertible iff $\det A \neq 0$.*

THEOREM 34 (Laplace expansion). *For any $j = 1, \cdots, n$,*

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij}.$$

Here A_{ij} is the (ij)th minor of A.,i.e., is the $(n-1) \times (n-1)$ matrix obtained by striking out the ith row and jth column of A.

PROOF. The jth column $a_j = \sum a_{ij} e_i$. Hence,

$$\begin{aligned} \det A &= \sum_{i=1}^n a_{ij} D(a_1, \cdots, a_{j-1}, e_i, a_{j+1}, \cdots, a_n) \\ &= \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij} \end{aligned}$$

where we used the lemma below. □

LEMMA 2. *Let A be a matrix whose jth column is e_i. Then*

$$\det A = (-1)^{i+j} \det A_{ij}.$$

PROOF. Suppose $i = j = 1$. We define

$$C(A_{11}) = \det \begin{pmatrix} 1 & 0 \\ 0 & A_{11} \end{pmatrix} = \det A$$

where the second equality follows from the basic properties of det. Then $C = \det A_{11}$ since it satisfies the properties 1,2,3 in Theorem 28. General cases follow similarly or could be deduced from the case $i = j = 1$. □

Let $A_{n\times n}$ be invertible. Then

$$Ax = u$$

has a unique solution. Write $A = (a_1, \cdots, a_n)$ and $x = \sum x_j e_j$, we have

$$u = \sum x_j a_j.$$

We consider

$$A_k = (a_1, \cdots, a_{k-1}, u, a_{k+1}, \cdots, a_n).$$

Then

$$\begin{aligned}\det A_k &= \sum x_j \det(a_1, \cdots, a_{k-1}, a_j, a_{k+1}, \cdots, a_n) \\ &= x_k \det A,\end{aligned}$$

hence

$$x_k = \frac{\det A_k}{\det A}.$$

Since

$$\det A_k = \sum_j (-1)^{j+k} u_j \det A_{jk},$$

we have

$$x_k = \sum_{j=1}^{n} (-1)^{j+k} u_j \frac{\det A_{jk}}{\det A}.$$

Comparing it with $x = A^{-1}u$, we have proved

THEOREM 35. *The inverse matrix A^{-1} of an invertible matrix A has the form*

$$\left(A^{-1}\right)_{kj} = (-1)^{j+k} \frac{\det A_{jk}}{\det A}.$$

DEFINITION 22. *The trace of a square matrix A, denoted as $\operatorname{tr} A$, is the sum of the entries on its diagonal:*

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

THEOREM 36. *(i) Trace is a linear functional on matrices.*
(ii) Trace is commutative: $\operatorname{tr} AB = \operatorname{tr} BA$.

DEFINITION 23.

$$\operatorname{tr} AA^T = \sum_{i,j=1}^{n} (a_{ij})^2$$

The square root of the double sum on the right is called the Euclidean, or Hilbert-Schmidt, norm of the matrix A.

The matrix A is called similar to the matrix B if there is an invertible matrix S such that $A = SBS^{-1}$.

Theorem 37. *Similar matrices have the same determinant and the same trace.*

Remark 8. *The determinant and trace of a linear map T can be defined as the determinant and trace of a matrix representing T.*

If $A = (a_{ij})_{n\times n}$ is an upper triangular square matrix, we have

$$\det A = \prod_{k=1}^{n} a_{kk}.$$

More generally, if A is an upper triangular block matrix, i.e., $A = (A_{ij})_{k\times k}$ where A_{ii} is an $n_i \times n_i$ matrix and $A_{ij} = O$ if $i > j$, then we can show that

$$\det A = \prod_{j=1}^{k} \det A_{jj}.$$

Remark 9. *If A, B, C, D are $n \times n$ matrices, we may not have*

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \det D - \det C \det B.$$

Another guess

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det (AD - CB)$$

is also false in general. For example,

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Theorem 38. *Let A, B, C, D be $n \times n$ matrices and $AC = CA$. Then*

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det (AD - CB).$$

Proof. We first assume that A is invertible. Since

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & O \\ C & D - CA^{-1}B \end{pmatrix},$$

we have

$$\begin{aligned} \det \begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \det A \det \left(D - CA^{-1}B\right) \\ &= \det \left(AD - ACA^{-1}B\right) \\ &= \det (AD - CB). \end{aligned}$$

If A is singular, we consider $A_\varepsilon = A + \varepsilon I$ and then send $\varepsilon \to 0$. □

Theorem 39. *Let A, B be $n \times n$ matrices. Then*

$$\det (I - AB) = \det (I - BA).$$

Remark 10. *In general, it is not true that*

$$\det (A - BC) = \det (A - CB).$$

Any $T \in L(\mathbb{C}^n, \mathbb{C}^m)$ can be represented by an m by n complex matrix$T = (t_{ij})_{m\times n}$. More generally, let X, U be learn spaces over $\mathbb{C}$. Any $T \in L(X, U)$ can be represented by a complex matrix once we choose bases for X and U. The algebra and most properties of real matrices can be extended natually to complex matrices.

If $A = (a_1, \cdots, a_n)$ is an $n \times n$ complex matrix, we could define the determinant

$$\det A = D(a_1, \cdots, a_n)$$

by extending D linearly to complex vectors and the formula

$$\det A == \sum_p \sigma(p)\, a_{p(1)1} a_{p(2)2} \cdots a_{p(n)n}$$

is still valid.

Theorem 40. *Let A, B be real matrices. Then A, B are similar as real matrices is equivalent to they are similar as complex matrices.*

6. Determinants of Special Matrices

Definition 24. *Let $n \geq 2$. Given n scalars $a_1, \cdots, a_n$, the Vandermonde matrix $V(a_1, \cdots, a_n)$ is a square matrix whose columns form a geometric progression:*

$$V(a_1, \cdots, a_n) = \begin{pmatrix} 1 & \cdots & 1 \\ a_1 & & a_n \\ \vdots & & \vdots \\ a_1^{n-1} & \cdots & a_n^{n-1} \end{pmatrix}.$$

Theorem 41.

$$\det V(a_1, \cdots, a_n) = \prod_{j>i} (a_j - a_i).$$

Proof. $\det V$ is a polynomial in the a_i of degree less than or equal to $n(n-1)/2$. Whenever $a_i = a_j$ for some $i \neq j$, we have

$$\det V = 0,$$

Hence, $a_i - a_j$ is a factor of $\det V$. Hence $\det V$ is divisible by the product

$$\prod_{j>i} (a_j - a_i)$$

which is also a polynomial in the a_i of degree equal to $n(n-1)/2$. Hence for some constant c_n.

$$\det V(a_1, \cdots, a_n) = c_n \prod_{j>i} (a_j - a_i).$$

We consider the coefficient of a_n^{n-1}, we have

$$c_n \prod_{n>j>i} (a_j - a_i) = \det V(a_1, \cdots, a_{n-1}) = c_{n-1} \prod_{n>j>i} (a_j - a_i)$$

hence $c_n = c_{n-1}$. Since $c_2 = 1$, we have $c_n = 1$. □

Example 8. *Let*

$$p(s) = x_1 + x_2 s + \cdots x_n s^{n-1}$$

be a polynomial in s. Let $a_1, \cdots, a_n$ be n distinct numbers, and let $p_1, \cdots, p_n$ be n arbitrary complex numbers; we wish to choose the coefficients $x_1, \cdots, x_n$, so that

$$p(a_j) = p_j,\ 1 \leq j \leq n.$$

Then we have

$$V(a_1, \cdots, a_n) x = p.$$

Definition 25. *Given $2n$ scalars $a_1, \cdots, a_n$, $b_1, \cdots, b_n$. The Cauchy matrix*

$$C(a_1, \cdots, a_n; b_1, \cdots, b_n)$$

is the $n \times n$ matrix whose ij-th element is

$$\frac{1}{a_i + b_j}.$$

Theorem 42.

$$\det C(a_1, \cdots, a_n; b_1, \cdots, b_n) = \frac{\prod_{j>i} (a_j - a_i) \prod_{j>i} (b_j - b_i)}{\prod_{i,j} (a_i + b_j)}.$$

PROOF. We have

$$\begin{aligned}&\prod_{i,j}(a_i+b_j)\det C(a_1,\cdots,a_n;b_1,\cdots,b_n)\\&=\det T_{ij}\end{aligned}$$

where

$$T_{ij}=\prod_{k\neq j}(a_i+b_k).$$

One can show that for some constant c_n,

$$\prod_{i,j}(a_i+b_j)\det C(a_1,\cdots,a_n;b_1,\cdots,b_n)=c_n\prod_{j>i}(a_j-a_i)\prod_{j>i}(b_j-b_i).$$

Let $a_n=b_n=d$, the coefficient of d^{2n-2} is

$$c_n\prod_{n>j>i}(a_j-a_i)\prod_{n>j>i}(b_j-b_i).$$

On the other hand, since for $1\le ij<n$,

$$\begin{aligned}T_{nn}&=\prod_{k\neq n}(d+b_k)=d^{n-1}+\cdots,\\T_{nj}&=\prod_{k\neq j}(d+b_k)=2d^{n-1}+\cdots,\\T_{in}&=\prod_{k\neq n}(a_i+b_k),\\T_{ij}&=\prod_{k\neq j}(a_i+b_k)=(a_i+d)\prod_{k\neq j,k\neq n}(a_i+b_k),\end{aligned}$$

we see the coefficient of d^{2n-2} is decided by

$$\begin{aligned}&d^{n-1}\det(T_{ij})_{(n-1)\times(n-1)}\\&=d^{2(n-1)}\prod_{i,j\le n-1}(a_i+b_j)\det C(a_1,\cdots,a_{n-1};b_1,\cdots,b_{n-1})\\&=d^{2(n-1)}c_{n-1}\prod_{j>i}(a_j-a_i)\prod_{j>i}(b_j-b_i).\end{aligned}$$

Hence, we conclude $c_n=c_{n-1}$. Since $c_2=1$ we conclude $c_n=1$. □

EXAMPLE 9. *If we consider*

$$\begin{aligned}a_1=1,a_2=2,a_3=3,\\b_1=1,b_2=2,b_3=3.\end{aligned}$$

We have

$$A=\begin{pmatrix}\frac12&\frac13&\frac14\\\frac13&\frac14&\frac15\\\frac14&\frac15&\frac16\end{pmatrix}.$$

And

$$\det A=\frac{(2)^2}{2\times3\times4\times3\times4\times5\times4\times5\times6}=\frac{1}{43200}.$$

EXAMPLE 10. *Consider*

$$T = \left(\frac{1}{1 + a_i a_j} \right)$$

for n *given scalars* $a_1, \cdots, a_n$.

7. Spectral Theory

Let A be an $n \times n$ matrix.

Definition 26. *Suppose that for a nonzero vector v and a scalar number λ,*

$$Av = \lambda v.$$

Then λ is called an eigenvalue of A and v is called an eigenvector of A corresponding to eigenvalue λ.

Let v be an eigenvector of A corresponding to eigenvalue λ. We have for any positive integer k,

$$A^k v = \lambda^k v.$$

And more generally, for any polynomial p,

$$p(A)\, v = p(\lambda)\, v.$$

Theorem 43. *λ is an eigenvalue of A if and only if*

$$\det(\lambda I - A) = 0.$$

The polynomial

$$p_A(\lambda) = \det(\lambda I - A)$$

is called the characteristic polynomial of the matrix A.

Theorem 44. *Eigenvectors of a matrix A corresponding to distinct eigenvalues are linearly independent.*

Proof. Let λ_k, $1 \le k \le K$ be distinct eigenvalues of A and v_k, $1 \le k \le K$ be corresponding eigenvectors. We prove by induction in K. The case $K = 1$ is trivial. Suppose the result holds for $K = N$. Now for $K = N + 1$, suppose for constants c_k, $1 \le k \le N + 1$, we have

$$\sum_{k=1}^{N+1} c_k v_k = 0. \tag{7.1}$$

Applying A, we have

$$\sum_{k=1}^{N+1} c_k \lambda_k v_k = 0$$

which implies

$$\sum_{k=1}^{N} c_k (\lambda_k - \lambda_{N+1})\, v_k = 0.$$

Since v_k, $1 \le k \le N$ are linearly independent and $\lambda_k - \lambda_{N+1} \ne 0$ for $1 \le k \le N$, we have $c_k = 0$, $1 \le k \le N$, and (7.1) implies $c_{N+1} = 0$. Hence the result holds for $K = N + 1$ too. □

Corollary 6. *If the characteristic polynomial p_A of the $n \times n$ matrix A has n distinct roots, then A has n linearly independent eigenvectors which forms a basis.*

Suppose A has n linearly independent eigenvectors v_k, $1 \le k \le n$ corresponding to eigenvalues λ_k, $1 \le k \le n$. Then A is similar to the diagonal matrix $\Lambda = [\lambda_1, \cdots, \lambda_n]$. Actually, let $S = (v_1, \cdots, v_n)$, we have

$$A = S \Lambda S^{-1}.$$

Example 11. *The Fibonacci sequence* $f_0, f_1, \cdots$ *is defined by the recurrence relation*

$$f_{n+1} = f_n + f_{n-1}$$

with the starting data $f_0 = 0$, $f_1 = 1$. *Let*

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

we have

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = A \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix}.$$

Simple computation yields

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2},$$
$$v_1 = \begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}.$$

Since

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} v_1 - \frac{1}{\sqrt{5}} v_2,$$

we have

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = A^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{\lambda_1^n}{\sqrt{5}} v_1 - \frac{\lambda_2^n}{\sqrt{5}} v_2$$

which implies

$$f_n = \frac{\lambda_1^n}{\sqrt{5}} - \frac{\lambda_2^n}{\sqrt{5}}.$$

Theorem 45. *Denote by* λ_k, $1 \le k \le n$, *the eigenvalues of* A, *with the same multiplicity they have as roots of the characteristic equation of* A. *Then*

$$\sum_{k=1}^{n} \lambda_k = \operatorname{tr} A \text{ and } \prod_{k=1}^{n} \lambda_k = \det A.$$

Proof. The first identity follows from the expansion

$$\begin{aligned} p_A(\lambda) = \det(\lambda I - A) &= \sum_p \sigma(p) \prod_{k=1}^{n} \left(\lambda \delta_{p_k k} - a_{p_k k}\right) \\ &= \lambda^n - (\operatorname{tr} A) \lambda^{n-1} + \cdots. \end{aligned}$$

The second identity follows from

$$(-1)^n \prod_{k=1}^{n} \lambda_k = p_A(0) = \det(-A) = (-1)^n \det(A).$$

□

Theorem 46 (Spectral Mapping Theorem). *(a) Let* q *be any polynomial,* A *a square matrix,* λ *an eigenvalue of* A. *Then* $q(\lambda)$ *is an eigenvalue of* $q(A)$.
(b) Every eigenvalue of $q(A)$ *is of the form* $q(\lambda)$, *where* λ *is an eigenvalue of* A.

Proof. (a) Let v be the eigenvector of A corresponding to λ. Then

$$q(A)v = q(\lambda)v.$$

(b) Let μ be an eigenvalue of $q(A)$, then

$$\det(\mu I - q(A)) = 0.$$

Suppose the roots of $q(\lambda) - \mu = 0$ is given by λ_i, then

$$q(\lambda) - \mu = c\prod(\lambda - \lambda_i),$$

which implies

$$\prod \det(\lambda_i I - A) = 0.$$

Hence, for some λ_i, $\det(\lambda_i I - A) = 0$. Hence $\mu = q(\lambda_i)$ where λ_i is an eigenvalue of A. $\square$

If in particular we take $q = p_A$, we conclude that all eigenvalues of $p_A(A)$ are zero. In fact a little more is true.

Theorem 47 (Cayley-Hamilton). *Every matrix A satisfies its own characteristic equation*

$$p_A(A) = 0.$$

Proof. Let

$$Q(s) = sI - A$$

and $P(s)$ defined as the matrix of cofactors of $Q(s)$, i.e.

$$P_{ij}(s) = (-1)^{i+j} D_{ji}(s)$$

where D_{ij} is the determinant of the ij-th minor of $Q(s)$. Then we have

$$P(s)Q(s) = (\det Q(s))I = p_A(s)I.$$

Since the coefficients of Q commutes with A, we have

$$P(A)Q(A) = p_A(A)I,$$

hence $p_A(A) = 0$. $\square$

Lemma 3. *Let*

$$P(s) = \sum P_k s^k, P(s) = \sum Q_k s^k, R(s) = \sum R_k s^k$$

be polynomials in s where the coefficients P_k, Q_k and R_k are $n \times n$ matrices. Suppose that

$$P(s)Q(s) = R(s)$$

and matrix A commutes with each Q_k, then we have

$$P(A)Q(A) = R(A).$$

Definition 27. *A nonzero vector u is said to be a generalized eigenvector of A corresponding to eigenvalue λ, if*

$$(A - \lambda I)^m u = 0$$

for some positive integer m.

Theorem 48 (Spectral Theorem). *Let A be an $n \times n$ matrix with complex entries. Every vector in $\mathbb{C}^n$ can be written as a sum of eigenvectors of A, genuine or generalized.*

Proof. Let x be any vector; the $n+1$ vectors $x, Ax, A^2x, \cdots, A^n x$ must be linearly dependent; therefore there is a polynomial p of degree less than or equal to n such that

$$p(A)x = 0$$

We factor p and rewrite this as

$$\prod_j (A - r_j I)^{m_j} x = 0.$$

All invertible factors can be removed. The remaining r_j are all eigenvalues of A. Applying Lemma 4 to $p_j(s) = (s - r_j)^m$, we have a decomposition of x as a sum of generalized eigenvectors. □

Lemma 4. *Let p and q be a pair of polynomials with complex coefficients and assume that p and q have no common zero.*
(a) There are two polynomials a and b such that

$$ap + bq = 1.$$

(b) Let A be a square matrix with complex entries. Then

$$N_{pq} = N_p \oplus N_q.$$

Here N_p, N_q, and N_{pq} are the null spaces of $p(A)$, $q(A)$, and $p(A)q(A)$.
(c) Let p_k $k = 1, 2, \cdots, m$ be polynomials with complex coefficients and assume that p_k have no common zero. Then

$$N_{p_1 \cdots p_m} = \oplus_{k=1}^m N_{p_k}.$$

Here N_{p_k} is the null space of $p_k(A)$.

Proof. (a) Denote by $\mathcal{P}$ all polynomials of the form $ap + bq$. Among them there is one, nonzero, of lowest degree; call it d. We claim that d divides both p and q; for suppose not; then the division algorithm yields a remainder r, say

$$r = p - md.$$

Since p and d belong to $\mathcal{P}$ so does r; since r has lower degree than d, this is a contradiction. We claim that d has degree zero; for if it had degree greater than zero, it would, by the fundamental theorem of algebra, have a root. Since d divides p and q, and p and q have no common zero, d is a nonzero constant. Hence $1 \in \mathcal{P}$.

(b) From (a), There are two polynomials a and b such that

$$a(A)p(A) + b(A)q(A) = I.$$

For any x, we have

$$x = a(A)p(A)x + b(A)q(A)x \overset{\triangle}{=} x_2 + x_1.$$

Here it is easy to verify that if $x \in N_{pq}$ then $x_1 \in N_q$, $x_2 \in N_p$. Now suppose $x \in N_p \cap N_q$, the above formula implies

$$x = a(A)p(A)x + b(A)q(A)x = 0.$$

Hence $N_{pq} = N_p \oplus N_q$. □

We denote by $\mathcal{P}_A$ the set of all polynomials p which satisfy $p(A) = 0$. $\mathcal{P}_A$ forms a linear space. Denote by $m = m_A$ a nonzero polynomial of smallest degree in $\mathcal{P}_A$; we claim that all p in $\mathcal{P}_A$ are multiples of m. Except for a constant factor, which we fix so that the leading coefficient of m_A is 1, $m = m_A$ is unique. This polynomial is called the minimal polynomial of A.

To describe precisely the minimal polynomial we return to generalized eigenvector. We denote by $N_m = N_m(\lambda)$ the nullspace of $(A - \lambda I)^m$. The subspaces N_m, consist of generalized eigenvectors; they are indexed increasingly, that is,

$$N_1 \subset N_2 \subset N_3 \subset \cdots .$$

Since these are subspaces of a finite-dimensional space, they must be equal from a certain index on. We denote by $d = d(\lambda)$ the smallest such index, that is,

$$\begin{aligned} N_d &= N_{d+k} \text{ for any } k \geq 1, \\ N_{d-1} &\neq N_d. \end{aligned}$$

$d(\lambda)$ is called the index of the eigenvalue λ.

Remark 11. *A maps N_d into itself. i.e., N_d is an invariant subspace under A.*

Theorem 49. *Let $\lambda_1, \cdots, \lambda_k$ be distinct eigenvalues of A with index $d_j = d(\lambda_j)$, $1 \leq j \leq k$. Then*
(1)

$$\mathbb{C}^n = \oplus_{j=1}^k N_{d_j}(\lambda_j),$$

(2) and the minimal polynomial

$$m_A = \prod_{j=1}^k (s - \lambda_j)^{d_j}.$$

Proof. (1). It follows from the spectral theorem and Lemma 4.
(2) Since

$$x = \sum_{j=1}^k x_j$$

where $x_j \in N_{d_j}(\lambda_j)$, we have

$$\prod_{j=1}^k (A - \lambda_j I)^{d_j} x = \sum_{j=1}^k \left(\prod_{i=1}^k (A - \lambda_i I)^{d_i} \right) x_j = 0.$$

Hence

$$\prod_{j=1}^k (A - \lambda_j I)^{d_j} = 0.$$

On the other hand, if

$$\prod_{j=1}^k (A - \lambda_j I)^{e_j} = 0$$

where $e_1 < d_1$ and $e_j \leq d_j$, we conclude

$$\mathbb{C}^n = \oplus_{j=1}^k N_{e_j}(\lambda_j)$$

which is impossible. □

THEOREM 50. *Suppose the pair of matrices A and B are similar. Then A and B have the same eigenvalues: $\lambda_1, \cdots, \lambda_k$. Furthermore, the nullspaces*

$$N_m(\lambda_j) = \text{null space of } (A - \lambda_j I)^m,$$
$$M_m(\lambda_j) = \text{null space of } (B - \lambda_j I)^m$$

have for all j and m the same dimensions.

PROOF. Let S be invertible such that

$$A = SBS^{-1},$$

then we have for any m and λ,

$$(A - \lambda I)^m = S(B - \lambda I)^m S^{-1}$$

□

Since $A - \lambda I$ maps N_{i+1} into N_i, $A - \lambda I$ defines a map from N_{i+1}/N_i into N_i/N_{i-1} for any $i \geq 1$ where $N_0 = \{0\}$.

LEMMA 5. *The map*

$$A - \lambda I : N_{i+1}/N_i \to N_i/N_{i-1}$$

is one-to-one. Hence

$$\dim(N_{i+1}/N_i) \leq \dim(N_i/N_{i-1})$$

PROOF. Let $B = A - \lambda I$. If $\left\{B\{x\}_{N_{i+1}/N_i}\right\}_{N_i/N_{i-1}} = \{0\}$, then $Bx \in N_{i-1}$, hence $x \in N_i$ and $\{x\}_{N_{i+1}/N_i} = \{0\}_{N_{i+1}/N_i}$. □

Next, we construct Jordan Canonical form of a matrix.

Let 0 be an eigenvalue of A. We want to construct a basis of $N_d = N_d(0)$.

Step I. Let $l_0 = \dim(N_d/N_{d-1})$, we construct $x_1, x_2, \cdots, x_{l_0}$ such that $\{x_1\}, \{x_2\}, \cdots, \{x_{l_0}\}$ form a basis of N_d/N_{d-1}.

Step II. Let $l_1 = \dim(N_{d-1}/N_{d-2}) \geq l_0$, we construct $Ax_1, Ax_2, \cdots, Ax_{l_0}, x_{l_0+1}, \cdots, x_{l_1}$ such that their quotient classes form a basis of N_{d-1}/N_{d-2}.

Step III. We continue this process until we reach N_1. We thus constructed a basis of N_d.

Step IV. It is illuminating to arrange the basis elements in a table:

$$\begin{array}{ccccc}
x_1 & Ax_1 & \cdots & & A^{d-1}x_1 \\
\vdots & & & & \vdots \\
x_{l_0} & Ax_{l_0} & \cdots & & A^{d-1}x_{l_0} \\
x_{l_0+1} & Ax_{l_0+1} & \cdots & A^{d-2}x_{l_0+1} & \\
\vdots & & & & \\
x_{l_1} & Ax_{l_1} & \cdots & A^{d-2}x_{l_1} & \\
\vdots & & & & \\
x_{l_{d-2}+1} & & & & \\
\vdots & & & & \\
x_{l_{d-1}} & & & &
\end{array}$$

Noticing

$$\begin{aligned}\dim N_d =& \sum_{k=0}^{d-1} l_k \\ =& dl_0 + (d-1)(l_1 - l_0) + (d-2)(l_2 - l_1) + \cdots 1 \times (l_{d-1} - l_{d-2}).\end{aligned}$$

Here in the above table, the span of the vectors in each row is invariant under A. And A restricted to each row has matrix representation of the form

$$J_m = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

which is called a Jordan block where

$$J_m(i,j) = 1 \text{ if } j = i+1 \text{ and } 0 \text{ otherwise.}$$

THEOREM 51. *Any matrix A is similar to its Jordan canonical form which consists diagonal blocks of the form*

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$

where λ is the eigenvalue of A.

THEOREM 52. *If A and B have the same eigenvalues $\{\lambda_j\}$, and if*

$$\dim N_m(\lambda_j) = \dim M_m(\lambda_j)$$

holds for any m, j, then A and B are similar.

REMARK 12. *The dimension of $N_{d(\lambda)}(\lambda)$ equals the multiplicity of λ as the root of the characteristic equation of A.*

THEOREM 53. *Suppose*

$$AB = BA.$$

Then there is a basis in $\mathbb{C}^n$ which consists of eigenvectors and generalized eigenvectors of both A and B. Theorem remains true if A, B are replaced by any number of pairwise commuting linear maps.

PROOF. Let $\{\lambda_j\}_{j=1}^k$ be distinct eigenvalues of A, $d_j = d(\lambda_j)$ be the index and $N_j = N_{d_j}(\lambda_j)$ the null space of $(A - \lambda_j I)^{d_j}$. Then

$$\mathbb{C}^n = \oplus N_j.$$

For any x, we have

$$B(A - \lambda_j I)^{d_j} x = (A - \lambda_j I)^{d_j} Bx.$$

Hence $B : N_j \to N_j$. Applying the spectral decomposition theorem to $B : N_j \to N_j$, we can find a basis of N_j consists of eigenvectors and generalized eigenvectors of B. $\square$

THEOREM 54. *Every square matrix A is similar to its transpose A^T.*

Proof. Recall that

$$\dim N_A = \dim N_{A^T}.$$

Since the transpose of $A - \lambda I$ is $A^T - \lambda I$ it follows that A and A^T have the same eigenvalues. The transpose of $(A - \lambda I)^m$ is $\left(A^T - \lambda I\right)^m$; therefore their nullspaces have the same dimension. Hence A and A^T are similar. $\square$

Theorem 55. *Let λ, μ be two distinct eigenvalues of A. Suppose u is an eigenvector of A w.r.t. λ and suppose v is an eigenvector of A^T w.r.t. μ. Then $u^T v = 0$.*

Proof. $v^T A u = u^T A^T v = \lambda v^T u = \mu u^T v$. $\square$