Linear Algebra Notes

Yin-Yun Li*

Fall 2025

These notes are based on the course *Introduction to Linear Algebra (Math4018)* taught by Prof. Kowk-Wing Tsoi at the Department of Mathematics, NTU. The structure follows the material prepared by Prof. Tsoi, with some adjustments and simplifications to aid exercises and comprehension. The details and complete arguments are documented in my handwritten notes.

*Undergraduate student, Department of Economics, National Taiwan University. Please feel free to contact me at b10302331@ntu.edu.tw with any comments or corrections.

Contents

1 Fields and Vector Space

Fields generalize the concepts of scalars. A set F is called a field if we can define addition and multiplication operations to satisfy 11 axioms.

EXERCISE 1.1 .
Write down the 11 axioms (5+5+1). These rules ensure we can add, subtract, multiply and divide by non-zero element in the field.

EXAMPLE 1.1 .
Some non-examples of fields:

1. $\mathbb{Z}$ *is not a field (division).*
2. $M_n(\mathbb{R})$ *is not a field (commutativity under multiplication).*

A set V is called a vector space over a field F if the space endowed with two operations satisfies 10 axioms.

EXERCISE 1.2 .
Write down the 10 axioms (5+5). These rules help us determine whether a given space is a vector space.

EXAMPLE 1.2 .
Let F be a field ($\mathbb{R}, \mathbb{C}, \mathbb{Q}, F_2$). $F^n = \{(x_1, x_2, \cdots x_n)\}$ is a vector space over F. In general, if two fields $F_1 \subseteq F_2$, then F_2 is a vector space over F_1.

EXERCISE 1.3 .
Prove for each $\mathbf{v} \in V$, its additive inverse $-\mathbf{v}$ is unique.

Proof.
Suppose $\exists\, \mathbf{v}' \neq \mathbf{v}''$ be two additive inverse of $\mathbf{v}$. Use $\mathbf{0}$ to relate $\mathbf{v}' = \mathbf{v}''$. ∎

Use the axioms to derive the related property:

PROPERTY 1.1 .
Let V be a vector space over F. Let $\mathbf{v} \in V$ and $a \in F$.

1. $0 \cdot \mathbf{v} = \mathbf{0}$
2. $a \cdot \mathbf{0} = \mathbf{0}$
3. $(-a)\mathbf{v} = -(a\mathbf{v})$
4. $a(-\mathbf{v}) = -(a\mathbf{v})$

Proof.

The first two: let LHS be a vector $\mathbf{w}$, and use $\mathbf{w} + \mathbf{w} = \mathbf{w}$ to show $\mathbf{w} = \mathbf{0}$. The last two: use the uniqueness of additive inverse and the result of 1, 2. ∎

DEFINITION 1.1 .
Let V be a vector space over F. $W \subseteq V$ is a subspace of V if W is also a vector space over F with addition and scalar multiplication inherited from V.

EXERCISE 1.4 .
Write down the sufficient and necessarily condition of a subspace.

THEOREM 1.1 (Subspace of $\mathbb{R}^n$).
Let A be a m by n matrix. Then the solution space of a homogeneous linear system of equations:

$$W = \{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \mathbf{0}\}$$

forms a subspace of $\mathbb{R}^n$.

Proof.

We can prove W is a subspace directly using the above theorem. Alternatively, consider a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$, the matrix A is exactly the matrix representing f. By construction, W is the kernal of f. $ker(f)$ is a subspace of domain $\mathbb{R}^n$ (We will prove this in the future). ∎

THEOREM 1.2 .
Fix a positive integer n. The subset of trace-free matrices:

$$W = \{A \in M_n(F) : tr(A) = 0\}$$

forms a subspace of $M_n(F)$.

Proof.

This proof is similar to the previous one and left as an exercise. ∎

EXAMPLE 1.3 .
Differentiable function space is a subspace of continuous function space. This is an easy exercise.

THEOREM 1.3 .
Intersection of subspaces is also a subspace.

Proof.

Left as an exercise. ∎

EXAMPLE 1.4 .
Consider a vector space $M_n(F)$, W_1 be a symmetric matrices, and W_2 be a upper-triangular matrices. The intersection $W_1 \cap W_2$ is a diagonal matrices, which is a subspace of the $n-$square matrices vector space.

Lastly, we introduce some concepts earlier.

PROPERTY 1.2 .
Reminder:

1. *Union of two subspaces is not necessarily a subspace.*
2. *Intersection of two subspaces is a subspace.*
3. *The sum of two subspaces is the smallest subspace containing both.*
4. *Spanning set is a subspace.*

2 Basis and Dimension

Let V be a (finite-dimension) vector space over a field F. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a finite subset of V.

DEFINITION 2.1 (Linear Combination).
Any vector of the form

$$a_1\mathbf{v_1} + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \ \text{ where } a_i \in F$$

is called a linear combination of S.

DEFINITION 2.2 (Linear Independence).

$$a_1\mathbf{v_1} + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0} \ \text{ implies } \ a_i = 0,\ i = 1, 2, \ldots, n$$

The only way to express the zero vector as a linear combination of S is to take all scalars to zeros.

S is linear dependent if it is not linearly independent.

THEOREM 2.1 .
S is linearly dependent iff either $\mathbf{v}_1 = 0$ or for some r, $\mathbf{v}_r$ is a linear combination of the preceding vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}$.

Proof.
($\Rightarrow$) Let r be the largest integer such that $a_r \neq 0$.
($\Leftarrow$) There exists a linear combination and show that the scalars are not all zeors. ■

DEFINITION 2.3 (Linear Span).
The set of all linear combinations of S is:

$$span(S) = \{a_1\mathbf{v_1} + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n : a_i \in F\}.$$

Note that $\text{span}(\emptyset) = \{\mathbf{0}\}$.

THEOREM 2.2 .
$span(S)$ is a subspace of V.

DEFINITION 2.4 (Spanning / Generating Set).
We say that S spans V / S is a spanning/generating set of V if $span(S) = V$.

For every $\mathbf{v} \in V$ can be expressed as a linear combination of S.

DEFINITION 2.5 (Basis).
A subset S of V is a basis of V if

1. *S is linearly independent.*
2. *S spans V.*

If S is a finite set, then V is finite-dimensional. Otherwise, V is infinite-dimensional. Additionally, a basis is not unique.

> **REMARK .**
> *If S spans V and $\mathbf{w} \in S$, then $\{\mathbf{w}\} \cup S$ is not linearly independent. Conversely, if $\mathbf{w} \notin S$, then $\{\mathbf{w}\} \cup S$ is still linearly independent.*

THEOREM 2.3 .
Suppose S is a basis of V. Then for every $\mathbf{v} \in V$, there exists a unique collection of scalars $a_1, a_2, \ldots, a_n$ such that

$$\mathbf{v} = a_1\mathbf{v_1} + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n.$$

Proof.
When we want to show the uniqueness, we could prove by contradiction. That is, fix a vector which could be expressed as two different linear combinations of S. ■

DEFINITION 2.6 (Coordinates of vector).
Fix an order of elements in a basis S. We call $(a_1, a_2, \ldots a_n)$ is the coordinates of $\mathbf{v}$ w.r.t the basis S.

The coordinates are denoted by $[\mathbf{v}]_S = (a_1, a_2, \ldots, a_n)$, and the same vector could have different coordinates w.r.t different bases or different orders of the same basis.

DEFINITION 2.7 (Dimension).
Suppose V is finite-dimensional vector space. The dimension of V is the size of a /any basis of V.

$$\dim_F(V) = |S|,$$

where S is a basis of V, $|\cdot|$ represents the size / cardinality of a set.

Note that $\text{span}(\emptyset) = \{\mathbf{0}\}$. The empty set spans zero vector space, and the empty set forms a basis of $\{\mathbf{0}\}$, $\dim_F(\{\mathbf{0}\}) = |\emptyset| = 0$. So zero vector space is the only zero-dimensional vector space.

THEOREM 2.4 (Exchange / Replacement Theorem).
Suppose $S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a spanning set of V and $S_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is linearly independent. Then $m \leq n$.

THEOREM 2.5 .
Let V be a vector space over F and W be a subspace of V. Then $dim_F(W) \leq dim_F(V)$.

Proof.

If V is infinite-dimensional, there is nothing to prove. Now suppose V is finite-dimensional. Claim: W must be finite-dimensional. Suppose, by contradiction, there exists a subspace with an infinite basis B_W. B_W is linearly independent in V. Let B_V be a basis of V, which is a finite-dimensional spanning set in V. By replacement theorem, which is a contradiction. Note the above argument uses the axiom of choice. ■

THEOREM 2.6 .
If both S_1 and S_2 are both bases of V, then $|S_1| = |S_2|$.

Proof.
Recall the definition of basis. Apply the exchange theorem to S_1 and S_2 from the 'both' sides. ■

THEOREM 2.7 .
Let V be a vector space of dimension n and $S \subseteq V$ be a subset.

1. *If $|S| > n$, then S is linearly dependent.*
2. *If $|S| < n$, then S does not span V.*

Proof.

1. Suppose, by contradiction, that S is linearly independent. Let B be any basis of V. We have $|B| = n$. By the exchange theorem, $|S| \leq |B| = n$, which is a contradiction. Thus S is linearly dependent.
2. Suppose, by contradiction, that S spans V. Let B be any basis of V. We have $|B| = n$. B is linearly independent and S spans V, by the exchange theorem, $|B| = n \leq |S|$, this contradicts $|S| < n$.

■

The proof from the above theorem provides a refined version of the exchange theorem:

THEOREM 2.8 (Refined Exchange / Replacement Theorem).
Let T be any linearly independent subset of V. Let B be any basis of V. Let S be any spanning set of V. Then $|T| \leq |B| \leq |S|$, where $|B| = \dim_F(V)$.

THEOREM 2.9 .
Let V and W be two finite-dimensional v.s. If $W \subseteq V$ and $\dim_F(W) = \dim_F(V)$, then $V = W$.

Proof.
It suffices to show that $W \supseteq V$. Suppose by contradiction,... Let S be a basis of W. Then we have $W = span(S)$ and $\mathbf{v} \notin span(S)$. Construct the larger size of linearly independent set in V. Apply the replacement theorem, then we have a contradiction. ■

THEOREM 2.10 .
Let V be a v.s. of dimension n. Then any n linearly independent subset of forms a basis of V.

Proof.
Easy proof. Let $W = span(S)$ and show that $W = V$. Recall any spanning set is a subspace of V. ■

EXERCISE 2.1 .
The above theorem shows that a maximal linearly independent set forms a basis of V. In duality, a minimal spanning set also forms a basis of V. Prove the theorem.

THEOREM 2.11 .
Let $S = \{\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n, \mathbf{w}\}$. If $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then $span(S) = span(S \setminus \{\mathbf{w}\})$.

Apply (2.1) iteratively, we have the following sifting method.

THEOREM 2.12 (Sifting method).
Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. There exists $S' \subseteq S$ s.t. S' is linearly independent and $span(S') = span(S)$.

THEOREM 2.13 (Basis Extension Theorem).
Let S be a linearly independent subset of V. Then S can be extended into a basis of V.

Proof.
Let B be any basis of V. Consider $S \cup B$. Apply the sifting method to $S \cup B$. Question: Does S remain in the sifted set? ■

THEOREM 2.14 (Duality of the Basis Extension Theorem).
Given S spans V, then there exists a subset $S' \subseteq S$ forms a basis of V.

Now let us prove replacement theorem

Proof.

Let $S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a spanning set of V and $S_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be linearly independent. Idea is that insert each $\mathbf{w}_i$ in front of S_1 and sift. Think why $\{w_i\} \cup S$ must be linearly dependent, then we get the inequality. ∎

Here we introduce sum of spaces as a subspace.

EXERCISE 2.2 .
Let W_1 and W_2 be two subspaces of a vector space V.

1. *Write down the definition of sum of spaces.*
2. *Prove this is smallest subspace containing W_1 and W_2.*
3. *Write down the definition of the direct sum of spaces.*
4. *Write down the dimension formula when the subspace is a direct sum of two spaces.*

THEOREM 2.15 (Dimension Formula for Sum of Subspaces).
Let V_1 and V_2 be two subspaces of a finite-dimensional vector space V. Then

$$\dim_F(V_1 + V_2) = \dim_F(V_1) + \dim_F(V_2) - \dim_F(V_1 \cap V_2).$$

Proof.
Note that $W_1 \cap W_2 \subseteq W_1, W_2 \subseteq W_1 + W_2$. Start with the smallest set, apply basis extension theorem w.r.t W_1 and W_2. Check the union of the three basis, S, forms the basis in $W1 + W2$, i.e. S is L.I. and spans $W_1 + W_2$. ∎

3 Linear Transformation

3.1 Linear function

Let $f : V \to W$ be a linear transformation, and V is finite dimensional.

DEFINITION 3.1 (linear function/transformation/map/homomorphism).
Check the function f: (i) Distribution on addition (ii) Scalars can be taken out of f.

EXAMPLE 3.1 .
Verify the following examples are linear maps.

1. *Differentiation function: $D : \mathbb{R}[x]_{\leq n} \to \mathbb{R}[x]_{\leq n-1}$*
2. *Trace: $tr : M_n(F) \to F$*
3. *Fix a continuous function $g \in C(\mathbb{R})$. $S_g : C(\mathbb{R}) \to C(\mathbb{R})$ defined by $S_g(f(x)) = f(g(x))$*

THEOREM 3.1 .
If $f : V \to W$ is a linear map. Then $f(\mathbf{0}_V) = \mathbf{0}_W$.

Alternatively, one can verify linearity via a contrapositive argument.

> **REMARK .**
> *Note the following difference:*
>
> 1. *$C(\mathbb{R}) \to C(\mathbb{R}) : f(x) \to f(x-1)$ is about replacing the domain with some function.*
> 2. *$\mathbb{R}^2 \to \mathbb{R}^2 : \{(x,y) : y = f(x)\} \to \{(x,y) : y = f(x-1)\}$ is about shifting the graph of function $f(x)$ to 1 unit to the right.*

THEOREM 3.2 .
Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of V, then f is determined by $\{f(\mathbf{v}_1), f(\mathbf{v}_2), \ldots, f(\mathbf{v}_n)\}$, i.e.,

1. *For every $\mathbf{v} \in V$, $f(\mathbf{v})$ is a linear combination of $\{f(\mathbf{v}_1), f(\mathbf{v}_2), \ldots, f(\mathbf{v}_n)\}$. Actually, this shows $im(f) = span(\{f(\mathbf{v}_1), f(\mathbf{v}_2), \ldots, f(\mathbf{v}_n)\})$.*
2. *Suppose the linear map $g : V \to W$ satisfies $g(\mathbf{v}_i) = f(\mathbf{v}_i)\, \forall i = 1, 2, \cdots, n$. Then $g(\mathbf{v}) = f(\mathbf{v})$ for all $\mathbf{v} \in V$.*

COROLLARY .
If f is an injection ($ker(f) = 0$), then $\{f(\mathbf{v}_1), f(\mathbf{v}_2), \ldots, f(\mathbf{v}_n)\}$ forms a basis in $im(f)$.

The definitions of kernal and image are left as an exercise.

THEOREM 3.3 .
Let $dim(V) = n$ and $dim(W) = k$

1. *$ker(f)$ is a subspace of V.*
2. *$dim(ker(f)) = nullity(f) = 0 \Leftrightarrow ker(f) = \{\mathbf{0}_w\} \Leftrightarrow f$ is injective.*
3. *$im(f)$ is a subspace of W.*
4. *$dim(im(f)) = rank(f) = k \Leftrightarrow im(f) = W \Leftrightarrow f$ is surjective.*

THEOREM 3.4 (Rank-Nullity Theorem).

$$nullity(f) + rank(f) = dim_F(V)$$

Proof.
Start with the smallest set.
Since $ker(f) \subseteq V$, let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of $ker(f)$ and extend it to a basis of V, denoted by $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k\}$.
Show that $\{f(\mathbf{w}_1), f(\mathbf{w}_2), \ldots, f(\mathbf{w}_k)\}$ forms a basis of $im(f)$. ■

THEOREM 3.5 .
If $f : V \to W$ is a linear bijection, then f^{-1} is also linear.

Note that if f is bijective, then the inverse function of f (f^{-1}) exists.

DEFINITION 3.2 (Linear Isomorphism).
Let V and W be two finite dimension vector spaces over a field F.

1. *$f : V \to W$ is a (linear) isomorphism from V onto W ($V \cong W$) if f is a bijective linear map.*
2. *V and W are said to be isomorphic as vector spaces if there exists a isomorphism $f : V \to W$.*

An trivial but important example is that $id_V : V \to V$ is a isomorphism. So every vector space is isomorphic to itself.

DEFINITION 3.3 (Homeomorphism).
Let (X, d_X) and (Y, d_Y) be two metric spaces.

1. *$f : X \to Y$ is a homeomorphism ($X \cong Y$) if f is bijection, f is continuous*

and f^{-1} is also continuous.

2. *X and Y are said to be homeomorphic as vector spaces if there exists a homeomorphism $f : X \to Y$.*

THEOREM 3.6 .
Fix S as a basis of V, the coordinate mapping $\phi : V_S \to F^n$ defined by $\mathbf{v} \to [\mathbf{v}]_S$, is an isomorphism.

THEOREM 3.7 .
$V \cong W \Leftrightarrow dim(V) = dim(F)$.

Proof.
($\Leftarrow$): $\phi_S : V \to F^n$ and $\phi'_S : W \to F^n$ are isomorphism. So $\phi_{S'}^{-1} \circ \phi_S$ is still an isomorphism. ■

DEFINITION 3.4 .
Let a linear map $f : V \to W$. Fix B_1 and B_2 as bases of V and W, where $dim(V) = n$ and $dim(W) = m$ The matrix is called the matrix representing f w.r.t the bases B_1 and B_2. We denote it as $[f]_{B_1}^{B_2} = (a_{ij})_{i,j}$. Equivalently, given any vector $\mathbf{v}_j \in V$, we have $f(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i$.

Draw the fundamental diagram. The bridge between vector space and actual vector world is linear isomorphism.

THEOREM 3.8 (Linear maps and matrix).

$$[f]_{B_1}^{B_2} \cdot [\mathbf{v}]_{B_1} = [f(\mathbf{v})]_{B_2}$$

COROLLARY .
Rank and nullity are invariant under different choices of bases. $\mathbf{v} \in ker(f) \Leftrightarrow f(\mathbf{v}) = \mathbf{0} \Leftrightarrow [f(\mathbf{v})]_{B_2} = [f]_{B_1}^{B_2} \cdot [\mathbf{v}]_{B_1} = \mathbf{0} \Leftrightarrow [\mathbf{v}]_{B_1} \in ker([f]_{B_1}^{B_2})$. Note that the choice of bases (B_1, B_2) is arbitrary. By dimension formula, we deduce that rank and nullity are preserved.

THEOREM 3.9 (Composition and matrix multiplication).

$$[g \circ f]_{B_1}^{B_3} = [g]_{B_2}^{B_3} \cdot [f]_{B_1}^{B_2}$$

Similarly, we can show the connection regarding sum and scaling (linear homomorphism, exercise), inverse function / matrix (later).

THEOREM 3.10 (Rank inequality).
Let $f : U \to V$ and $g : V \to W$ be two linear maps. Linear map version:

$$rank(g \circ f) \leq \min\{rank(f), rank(g)\}.$$

Matrix version:

$$rank(AB) \leq \min\{rank(A), rank(B)\}.$$

THEOREM 3.11 .
Let $A \in M_{m\times n}(F)$ and $B \in M_{n\times m}(F)$. If $BA = I_n$ and $AB = I_m$, then $n = m = rank(A)$.

This theorem shows that only square matrices can possibly be invertible.

DEFINITION 3.5 .
Let A be a n-square matrix. If there exists B such that $AB = BA = I_n$, then we say A is invertible/non-singular. And B is called the inverse (matrix) of $A(A^{-1})$.

THEOREM 3.12 .

$$([f]_{B_1}^{B_2})^{-1} = [f^{-1}]_{B_2}^{B_1}.$$

Check $[id_V]_{B_1}^{B_1} = I_n$. Use the definition of inverse matrix.

3.2 Matrix Operations

EXERCISE 3.1 .
Row / column operations:

1. *Write down the three types of row operations and define the corresponding elementary row matrix.*

2. *Write down the three types of column operations and define the corresponding elementary column matrix*

THEOREM 3.13 .
Let A be a $m \times n$ matrix.

1. *Performing a row operation on A is equivalent to left-multiplying an elementary row matrix to A.*

2. *Performing a column operation on A is equivalent to right-multiplying an elementary column matrix to A.*

THEOREM 3.14 .
Let A be a $m \times n$ matrix.

1. *If Q is an invertible $m \times m$ matrix, then $rank(QA) = rank(A)$.*
2. *If P is an invertible $n \times n$ matrix, then $rank(AP) = rank(A)$.*

Proof.
Hint: rank inequality. ∎

The theorem shows that row / column operations preserves rank.

EXERCISE 3.2 . 1. *Define the row echelon form (REF) and column echelon form (CEF).*

2. *Prove elementary row matrix and elementary column matrix are invertible (Apply operations reversely).*

THEOREM 3.15 .
Let A be a $m \times n$ matrix.

1. *A can be transformed into a REF by a finite number of row operations. Equivalently, there exists an invertible $m \times m$ matrix Q such that QA is in REF.*
2. *A can be transformed into a CEF by a finite number of column operations. Equivalently, there exists an invertible $n \times n$ matrix P such that AP is in CEF.*

Proof.
An algorithm helps us transform any matrix into echelon form. This is suitable for calculations with concrete examples. Invertiable matrix is the product of elementary matrices. ∎

THEOREM 3.16 .
Let A be a matrix with m columns $\{\mathbf{c}_1, \mathbf{c}_2 \cdots \mathbf{c}_m\}$. Then

1. *$im(A) = span(\{\mathbf{c}_1, \mathbf{c}_2 \cdots \mathbf{c}_m\})$.*
2. *$rank(A)$ equals the maximal number of linearly independent columns of A.*
3. *$rank(A)$ equals the number of non-zero columns when A is transformed into a CEF.*

Proof.

1. Matrix vector multiplication.
2. By sifting method, the sifted subset forms a basis of $im(A)$.

3. Sifting method is equivalent to tranform A into CEF.

■

DEFINITION 3.6 (Smith Normal Form).
A $m \times n$ matrix is said to be in its Smith normal form if the matrix is both in its REF and CEF. Additionally, all non-zero entires of the matrix are 1.

THEOREM 3.17 (QAP Theorem).
Let A be a $m \times n$ matrix. There exsits invertible matrices $Q \in M_m(F)$ and $P \in M_n(F)$ s.t. QAP is in its Smith normal form.

THEOREM 3.18 .
Let A be a $m \times n$ matrix. Then $rank(A) = rank(A^\top)$.

Proof.
$rank(A) = rank(QAP) = rank(I_r) = r = rank(P^\top A^\top Q^\top) = rank(A^\top)$. ■

THEOREM 3.19 .
TFAE:

1. *the rank of A ($dim(im(A))$).*
2. *the maximal number of linearly independent columns of A.*
3. *the number of non-zero columns when A is transformed into a CEF.*
4. *the rank of $A^\top$.*
5. *the maximal number of linearly independent rows of A.*
6. *the number of non-zero rows when A is transformed into a REF.*

Proof.
$rank(A) = rank(A^\top)$
= maximal # of linearly independent columns of $A^\top$
= maximal # of linearly independent rows of A
= # of non-zero columns in CEF of $A^\top$ = # of non-zero rows in REF of A. ■

THEOREM 3.20 .
Every invertible matrix is a product of elementary matrices.

With the above theorem, we can easily find the inverse of an invertible matrix A.

THEOREM 3.21 .
Let A be a n-square matrix. If we conduct row operations to transform the argumented matrix:

$$(A \mid I_n) \xrightarrow{\text{row operation}} (I_n \mid B)$$

Then $B = A^{-1}$.

Proof.

Left multiplying the inverse matrix of A is equivalent to Perform row operations since each invertible matrix is a product of elementary matrices. ∎

EXERCISE 3.3 .
Write down the definition of homogeneous and inhomogeneous linear system of equations. Record the coefficient and argumented matrix.

THEOREM 3.22 .
The solution of a homogenous system $(A \mid \mathbf{0})$ coincides $ker(A)$. Particularly, $dim(ker(A)) =$ (# of columns (domain)) $- rank(A)$.

THEOREM 3.23 .
Let $A \in M_{m\times n}(F)$. If $m < n$, then the homogenous system $(A \mid \mathbf{0})$ has a non-trivial / non-zero solution.

Proof.

Intuitively, there are more variables than equations.
$im(A) \subseteq F^m \to rank(A) < m$. By rank-nullity theorem, $dim(ker(A)) = n - rank(A) \geq n - m \geq 1 (\because n > m)$. ∎

THEOREM 3.24 .
Let $(A \mid \mathbf{b})$ be a general system of equations. The system is consistent (has at least one solution) iff $rank(A) = rank(A \mid \mathbf{b})$.

THEOREM 3.25 (Principle of linearity).
Suppose $\mathbf{v}$ be a solution to the system of equation $(A \mid \mathbf{b})$, then the general solution of $(A \mid \mathbf{b})$ is

$$\mathbf{v} + ker(A) = \{\mathbf{v} + \mathbf{u} : \mathbf{u} \in ker(A)\}.$$

COROLLARY .
The system $(A \mid \mathbf{b})$ has a unique solution iff

1. $rank(A) = rank(A \mid \mathbf{b})$.
2. $ker(A) = \{\mathbf{0}\}$.

THEOREM 3.26 (TFAE Version 1).
Let $T : V \to W$ be a linear transformation such that $dim(V) = dim(W)$. Having fixed bases, T can be represened by a square matrix A. TFAE:

1. *T is injective.*
2. *$ker(T) = \{\mathbf{0}\}$.*
3. *$nullity(T) = 0$.*
4. *$rank(T) = dim(V) = dim(W)$.*
5. *T is surjective.*
6. *$im(T) = W$.*
7. *T is an isomorphism.*
8. *A is invertible (A^{-1} exists).*
9. *$ker(A) = \{\mathbf{0}\} \to nullity(A) = 0$.*
10. *A is full rank ($rank(A) = rank(T)$).*
11. *All of rows of A are linearly independent.*
12. *All of columns of A are linearly independent.*
13. *The system of equations $A\mathbf{x} = \mathbf{v}$ has a unique solution, where $\mathbf{x} = A^{-1}\mathbf{v}$ ($\because rank(A) = rank(A \mid \mathbf{b})$ and $ker(A) = 0$).*
14. *$det(A) = det(A^\top) \neq 0$.*

3.3 Change of coordinate & Linear Operator

DEFINITION 3.7 (Change of coordinate matrix).
A matrix representing $id_V : V \to V$ w.r.t two bases of V. Then the matrix from S_1 basis to S_2 basis is denoted by $[id_V]_{S_1}^{S_2}$.

DEFINITION 3.8 .
A linear transformation $T : V \to V$ is called a linear operator on V or an endomorphism on V.

THEOREM 3.27 .
Let $T : V \to V$ be a linear operator and $Q = [id_V]_{S_1}^{S_2}$ be a change of coordinate matrix. Then:

$$[T]_{S_1}^{S_1} = Q^{-1} \cdot [T]_{S_2}^{S_2} \cdot Q.$$

Proof.

Draw the diagram to derive the equation. ∎

Application: Diagonalization through $B = Q^{-1}AQ$. That is, choose a good basis (Q) appropriately s.t. B is as simple as possible (ideally a diagonal matrix).

THEOREM 3.28 .
Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V. Suppose S and S' are two bases of V. Then:

$$det([T]_S^S) = det([T]_{S'}^{S'}).$$

DEFINITION 3.9 .
Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V. Define the determinant of T by $det(T) := det([T]_S^S)$ for any choice of a basis S on V.

4 Determinants

EXERCISE 4.1 .
Let $A = (a_{ij})_{i,j} \in M_n(F)$.

1. *Write down the definition of determinants $det(A)$ by Laplace expansion (Consider $n = 1$ and $n \geq 2$).*

2. *Write down the definition of $(i, j)-$minor and $(i, j)-$cofactor.*

REMARK .
Let $det : M_n(F) \to F$.

1. *det is not linear.*

2. *det is a multi-linear function.*

EXAMPLE 4.1 .
Let A be an upper triangular matrix. Then $det(A) = \prod_{i=1}^{n} a_{ii}$.

Proof.
Prove by induction.

1. When $n = 1, det(A_1) = 1$.

2. Suppose $n = k$ holds. Then for $n = k + 1$, $det(A_{k+1}) = a_{11}M_{11}$. Note that $M_{1,j} = 0 \ \forall j \neq 1$.

3. So for a matrix A_k, we extend a larger matrix A_{k+1} by the left row and upper column.

■

THEOREM 4.1 (Determinants of elementary matrices).
For elementary matrices, we have:

- $det(E^{I}_{p,q}) = -1$
- $det(E^{II}_{\lambda,p}) = \lambda$
- $det(E^{III}_{\mu,p,q}) = 1$
- $det(E) = det(E^{\top}) \neq 0$

THEOREM 4.2 .
Let $A \in M_n(F)$:

1. *Exchangeing two rows of A multiplies $det(A)$ by 1.*
2. *Multiplying a row of A by a non-zero $\lambda \in F$ multiplies $det(A)$ by λ.*
3. *Adding a multiple of a row to another row does not change $det(A)$.*

Proof.
This proof is very hard.

- For (1), we first prove swapping two rows and generalize to swap any two rows. Express the determinant as a linear combination of some $N_{p,q}$, which are minors deleting $1, 2$ rows and p, q columns.
- For (2) and (3), it suffices to show $det : M_n(F) \to F$ is row-wise linear, or a multi-linear function. That is, determinants is linear on each row with other rows fixed.

■

THEOREM 4.3 .
Let $A \in M_n(F)$.

1. *If $det(A)$ contains a zero row, then $det(A) = 0$.*
2. *Let $\lambda \in F$. Then $det(\lambda A) = \lambda^n det(A)$.*

EXERCISE 4.2 .
Compute $det(-A)$.

THEOREM 4.4 .
A is not invertible if and only if $det(A) = 0$.

Proof.
$(\Rightarrow)$

1. By theorem (3.26), A is not invertible $\Leftrightarrow$ $REF(A)$ contains (at least) one zero row $\Leftrightarrow$ $det(REF(A)) = 0$.
2. Note $det(REF(A)) = det(E_1 E_2 \cdots E_k A) = det(E_1)det(E_2) \cdots det(E_k)det(A)$, where determinants of elementary matrices are non-zero. So $det(A) = 0$.

$(\Leftarrow)$

1. Suppose by contradiction, then $det(A)det(A^{-1}) = det(I)$.

2. Logically equivalent to show that A is invertible then $det(A) = 0$. Since every invertible matrix is a product of elementary matrices, whose determinant are non-zeros.

■

REMARK .
The above theroem is logically equivalent to the (4.6).

THEOREM 4.5 .
Let $A, B \in M_n(F)$, $det(AB) = det(A)det(B)$.

Proof.

If A is invertible, then it is an easy proof. If A is not invertible, we have $det(A) = 0$. Wish $det(AB) = det(A) = 0$. Since $rank(AB) \leq rank(A) < n$, we have $det(AB) = 0$. ■

THEOREM 4.6 .
Let $A \in M_n(F)$.

1. *A is invertible* $\iff det(A) \neq 0$.
2. *If A is invertible, then* $det(A^{-1}) = \frac{1}{det(A)}$

THEOREM 4.7 .
$det(A^\top) = det(A)$.

Proof.

If A is invertible, then we can prove through elementary matrices. If A is not invertible, then we compute $rank(A^\top)$ by QAP and TFAE. ■

This implies column operations on determinants is workable.

EXERCISE 4.3 .
Write down the general form of Laplace expansion and prove it.

DEFINITION 4.1 (Adjugate Matrix).
Let $A = (a_{ij} \in M_n(F))$. *The adjugate matrix of A is defined by* $adj(A) \coloneqq \left((-1)^{i+j}M_{ij}\right)^\top$.

THEOREM 4.8 .
Let $A \in M_n(F)$.*Then we have:*

1. $A \cdot adj(A) = det(A) \cdot I_n$.
2. *If A is invertible, then* $A^{-1} = \frac{1}{det(A)} \cdot adj(A)$.

Proof.

Refer to the handwritten notes. ∎

THEOREM 4.9 (Cramer's rule).

Let $A \in M_n(F)$ be an invertible matrix. Then the system of equations $A\mathbf{x} = \mathbf{b}$ has a unique solution (TFAE). Moreover, if $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$, then $x_i = \frac{det(A_i)}{det(A)}$, where A_i is the matrix obtained by replace the $i-$th column of A by the column vector $\mathbf{b}$.

Proof.

$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{det(A)} adj(A)\mathbf{b}$. Spell the expression out. ∎

Block Matrix

EXERCISE 4.4 .

Write down the definition of 2×2 block matrix. Show the multiplication of two block matrices.

THEOREM 4.10 (Block inversion formula).

Suppose:

1. ***A** and **D** are invertible square matrices.*
2. ***A** and **D** are possibly different sizes.*

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{O} & \mathbf{D}^{-1} \end{pmatrix}.$$

In particular, by determinant formula, $det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix} = det(\mathbf{A})det(\mathbf{D}) \neq 0$.

THEOREM 4.11 (Block determinant formula).

Suppose:

1. ***A** and **D** are square matrices.*
2. ***A** and **D** are possibly different sizes.*

$$det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix} = det(\mathbf{A})det(\mathbf{D}).$$

Proof.

Both formulas need some observations about multiplication of matrices. For the second one, we wish to transform

$$\begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{O} & \mathbf{I} \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{pmatrix}.$$

Then apply the follwoing formula. ∎

THEOREM 4.12 .
Let n be a positive integer and A be a square matrix. Then:

$$det\begin{pmatrix} \mathbf{I_n} & \mathbf{O} \\ \mathbf{O} & \mathbf{A} \end{pmatrix} = det\begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{I_n} \end{pmatrix} = det(A).$$

Proof.

Prove by induction. ∎

EXERCISE 4.5 .
Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same size, prove

$$det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} = det(\mathbf{A}+\mathbf{B})det(\mathbf{A}-\mathbf{B}).$$

THEOREM 4.13 (Schur's factorization).
Suppose

1. *$\boldsymbol{A}$ and $\boldsymbol{D}$ are square matrices.*
2. *$\boldsymbol{A}$ and $\boldsymbol{D}$ are possibly different sizes.*
3. *$\boldsymbol{A}$ is invertible.*

$$det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = det(\mathbf{A})det(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}).$$

Proof.

Observation: To kill $\mathbf{C}$, do row operations with a block of rows at the same time. Consider left multiplying a "block" elementary matrix. ∎

5 Diagonalization

5.1 Similarity

DEFINITION 5.1 .
We say A and B of $M_n(F)$ are similar if there exists an invertible $Q \in M_n(F)$ s.t. $B = Q^{-1}AQ$.

THEOREM 5.1 (Sufficient condition for similarity).
If $A \sim B$ and $A, B \in M_n(\mathbb{F})$, then

1. $rank(A) = rank(B)$.
2. $nullity(A) = nullity(B)$.
3. $det(A) = det(B)$.
4. $tr(A) = tr(B)$.
5. $p_A(x) = p_B(x)$: *the same chracteristic polynoimal.*

If we further assume that F is algebratically closed (e.g. $\mathbb{C}$), then they have the same:

1. *Multi-set of eigenvalues (counted with algebraic multiplicities).*
2. *algebraic multiplicity of each eigenvalue.*
3. *geometric multiplicity of each eigenvalue.*

Proof.

1. By rank inequality, multiplying an invertible matrix preserves the rank of the original matrix.
2. By rank-nullity theorem.
3. Easy proof.
4. First prove $tr(AB) = tr(BA)$, where $A \in M_{m\times n}(F)$ and $B \in M_{n\times m}(F)$.
5. Derive $det(A - xI) = det(B - xI)$.

If we assume $A, B \in M_n(\mathbb{C})$, by the fundamental theorem of algebra,

1. Same chracteristic polynomial implies same set of eigenvalues.
2. Similar argument for algebraic multiplicites of eigenvalues.
3. Show $nullity(A - \lambda I_n) = nullity(B - \lambda I_n)$.

■

REMARK . *1. Multi-set of eigenvalues, which means the repetition of an eigenvalue matters. We will introduce the concept of spectrum later, and the statement is equivalent to say two similar matrix have the same spectrum.*

2. $A \sim B \Leftrightarrow A$ and B represent the same linear operator under different bases.

DEFINITION 5.2 (Diagonalizable operators and matrices). *1. A linear operator $T : V \to V$ is diagonalizable if $\exists$ a basis S of V s.t. the matrix $[T]_S^S$ is diagonal.*

2. A matrix $A \in M_n(F)$ is diagonalizable if A is similar to a diagonal matrix, i.e., $\exists$ an invertible matrix Q s.t. $Q^{-1}AQ$ is diagonal.

DEFINITION 5.3 .
Let $T : V \to V$ be a linear operator.

1. *$\lambda \in F$ is an eigenvalue of T if $T(\mathbf{v}) = \lambda\mathbf{v}$ for some non-zero $\mathbf{v} \in V$.*
2. *$E_\lambda(T) = ker(A - \lambda \cdot id)$ is called the eigenspace w.r.t the eigenvalue λ.*
3. *Any non-zero $\mathbf{v} \in E_\lambda(T)$ is called an eigenvector of T w.r.t the eigenvalue λ. So any eigenvector satisfies $T(\mathbf{v}) = \lambda\mathbf{v}$.*

THEOREM 5.2 .
If $\lambda_1, \lambda_2, \cdots \lambda_n$ be distinct eigenvalues of T and $\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n$ be the corresponding eigenvectors, then $\{\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n\}$ is linearly independent.

Proof.
Prove by induction. Suppose it holds for $n = k - 1$. Then consider $n = k$. Suppose $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_k\mathbf{v}_k = 0$ (Equation 1). Apply T on both sides and use the identity $T(\lambda_i) = \lambda_i\lambda_i \, \forall i = 1, 2, \cdots k$ (Equation 2). Note that $\lambda_i \neq \lambda_j \, \forall i \neq j$. Reduce the $k - th$ term to obtain a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_{k-1}\}$. By induction hypothesis and distinctness of eigenvalues, we have $a_1 = a_2 = \cdots a_{k-1} = 0$. Simplify the equation 1 to get $a_k = 0$. By the principle of mathematical induction, the proof is completed. ■

DEFINITION 5.4 (Chracteristic polynomial).
Two cases:

1. *Let $A \in M_n(F)$. The chracteristic polynoimal of A is defined by $p_A(x) = det(A - xI_n)$.*
2. *Let $T : V \to V$ be a linear operator. The chracteristic polynoimal of T is defined by $p_T(x) = det([T]_s^s - xI_n)$ (For any basis s of V).*

THEOREM 5.3 .
Eigenvalues are the roots of the chracteristic polynomial of A. That is, λ is an eigenvalue of A iff $p_A(\lambda) = det(A - \lambda I_n) = 0$.

REMARK .
The above theorem demonstrates how to find eigenvalues practically.

5.2 Diagonalizability

THEOREM 5.4 (Diagonalizability Criterion I: Eigen-basis).
$A \in M_n(F)$ is diagonalizable iff $\exists$ a basis $\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n$ in F^n s.t. each v_i is an eigenvector of A w.r.t. some eigenvalue λ_i.

Proof.

This is the definition of diagonalizability itself. Associate A with a linear operator $f_A : F^n \to F^n$. Let B be the standard basis.

1. Think why $A = [f_A]_B^B$.

2. f_A is diagonalizable iff $\exists$ a basis S of F^n s.t. $[f_A]_S^S$ is diagonal. $\Leftrightarrow$ A is diagonalizable iff $\exists$ an invertible $Q = [id_V]_S^B$.

3. Think: Q is essentially formed by putting each eignevectors in S as columns.

4. Think: Why S is a set of eigenvectors? Why the diagonal entries in $[f_A]_S^S$ are the corresponding eigenvalues? ($f_A(\mathbf{v}_i) = A\mathbf{v}_i = \lambda\mathbf{v}_i$).

■

THEOREM 5.5 .
If $A \in M_n(F)$ has n distinct eigenvalues, then A is diagonalizable.

Proof.

1. Distinct eigenvalues shows that ...?

2. By replacement theorem, we find exactly a set of eigen-basis, hence A is diagonalizable.

■

DEFINITION 5.5 .
Let $A \in M_n(\mathbb{C})$. By the fundamental theorem of algebra, $p_A(x)$ can be factorized as $(-1)^n(x-\lambda_1)^{r_1}(x-\lambda_2)^{r_2}\cdots(x-\lambda_k)^{r_k}$, where $\lambda_1, \lambda_2, \cdots \lambda_k$ are distinct. For each eigenvalue λ_i,

1. *Algebraic multiplicity of λ_i: # of copies of $(x-\lambda_i)$ in $p_A(x)$, denoted by $a(\lambda_i) = r_i$.*

2. *Geometric multiplicity of λ_i, denoted by $g(\lambda_i) = nullity(A - \lambda_i I_n)$, which measures the dimension of the eigenspace of λ_i.*

THEOREM 5.6 .
Let $A \in M_n(\mathbb{C})$. For each λ of A:

$$1 \leq g(\lambda) \leq a(\lambda).$$

Proof.

1. $1 \leq g(\lambda) : det(A - \lambda I_n) = 0$ and use TFAE.

2. $g(\lambda) \leq a(\lambda)$: Take a basis of $(A - \lambda I_n)$ (assume dimension equals p). By basis extension theorem, we can obtain a basis S of $\mathbb{C}^n$. Apply associated linear operator T and get a "quasi-diagonla" matrix. Note that $p_{[T]_S^S}(x) = (-1)^n(x-\lambda)^p \times r(x)$. $r(x)$ may or may not contain $(x-\lambda)$.

■

THEOREM 5.7 (Diagonalizability Criterion II: Multiplicity).
Let $A \in M_n(\mathbb{C})$. TFAE:

1. *A is diagonalizable.*
2. *$\exists$ a basis $\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n$ in $\mathbb{C}^n$ s.t. each v_i is an eigenvector of A w.r.t. some eigenvalue λ_i.*
3. *For each λ of A, we have $g(\lambda) = a(\lambda)$.*

Proof.
It suffices to show:

1. (1) implies (3): Let D be a diagonal matrix, $A \sim D$ implies they have the same geometric multiplicities.
2. (3) implies (2): For each eigenspace, choose S_i to be a basis of $E_{\lambda_i}(A)$. Claim that $S = S_1 \cup S_2 \cup \cdots \cup S_k$ forms an eigen basis of A. Detail: for each λ_i, we may choose more than one eigenvector, why is the whole set S linearly independent?

■

EXERCISE 5.1 .
Take a polynomial $f \in F[x]$. For $A \in M_n(F)$, define $f(A)$.

THEOREM 5.8 (Cayley-Hamilton Theorem).
Let $A \in M_n(F)$ and its chracteristic polynoimal $p_A(x)$. Then $p_A(A) = \mathbf{O}$ (the zero matrix).

Proof.
Use $(A - xI)adj(A - xI) = det(A - xI)I = p_A(x)I$. Compare coefficients. ■

DEFINITION 5.6 (Minimal Polynomial).
Let $A \in M_n(\mathbb{C})$. The minimal polynoimal is the monic polynoimal of minimal degree s.t. $m_A(A) = \mathbf{O}$.

Monic means the leading coefficient is 1.

THEOREM 5.9 .
Let $A \in M_n(\mathbb{C})$, $p_A(x)$ and $m_A(x)$ be its chracteristic polynomial and minial polynomial respectively.

1. *For any $f \in \mathbb{C}[x]$ satisfying $f(A) = \mathbf{O}$, then $f(x)$ is divisible by $m_A(x)$/ $m_A(x)$ divides $f(x)$.*

2. *$p_A(x)$ is divisible by $m_A(x)$/ $m_A(x)$ divides $p_A(x)$.*

3. *Each eigenvalue λ of A is a root of $m_A(x)$/ $m_A(\lambda) = 0$.*

Proof.

1. Conduct division of $f(x)$ by $m_A(x)$. Aim to argue the remainder $r(x)$ must be zero polynomial. Suppose not, then $r(x)$ is a samller degree polynomial satisfying $r(A) = 0$, which yields a contradiction.

2. By Cayley-Hamilton and the property (1).

3. Use (5.10), we have: $m_A(A)\mathbf{v} = m_A(\lambda)\mathbf{v}$. This forces $m_A(\lambda) = 0$.

■

THEOREM 5.10 .
Let f be a non-constant polynomial. If λ is an eigenvalue of A, then $f(\lambda)$ is an eigenvalue of $f(A)$.

Proof.
Key observation: $A^k\mathbf{v} = \lambda^k\mathbf{v}$ for any $k \in \mathbb{N}$. ■

LEMMA .
Every matrix $A \in M_n(\mathbb{C})$ has a unique minimal polynomial.

Proof.
Use the property (1). ■

LEMMA (Candidate of minial polynomial).
Let $A \in M_n(\mathbb{C})$ and $\lambda_1, \lambda_2, \cdots \lambda_k$ be all distinct eigenvalues of A. If $p_A(x) = (-1)^n \cdot (x-\lambda_1)^{r_1}(x-\lambda_2)^{r_2}\cdots(x-\lambda_k)^{r_k}$. Then $m_A(x) = (x-\lambda_1)^{s_1}(x-\lambda_2)^{s_2}\cdots(x-\lambda_k)^{s_k}$, where $1 \le s_i \le r_i$ for each $i = 1, 2, \cdots, k$.

Proof.
Property (2) and (3) imply this lemma. ■

THEOREM 5.11 (Diagonalizability Criterion III: Minimal Polynomial).
Let $A \in M_n(\mathbb{C})$. TFAE:

1. *A is diagonalizable.*

2. *$\exists$ a basis $\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n$ in $\mathbb{C}^n$ s.t. each v_i is an eigenvector of A w.r.t. some eigenvalue λ_i.*

3. *For each λ of A, we have $g(\lambda) = a(\lambda)$.*

4. *The minial polynoimal $m_A(x)$ has no repeated factor.*

Proof.

1. (2) implies (4): Suppose (2) is true. Let $p_A(x) = (-1)^n \cdot (x-\lambda_1)^{r_1}(x-\lambda_2)^{r_2}\cdots(x-\lambda_k)^{r_k}$ and $q(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$, where $\lambda_i \neq \lambda_j \forall i \neq j$. Aim to show $q(x) = m_A(x)$. By property (3) in (5.9), we have $q(x) \mid m_A(x)$. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n\}$ be a basis of $\mathbb{C}^n$ and eigenvectors of A. For each $\mathbf{v} \in S$, $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda = \lambda_i$. By (5.10), $q(A)\mathbf{v} = q(\lambda)\mathbf{v}$ and $q(\lambda) = 0$. Note that $S \subseteq ker(q(A)) \to |S| \leq nullity(q(A)) \leftrightarrow rank(q(A)) = 0 \leftrightarrow q(A) = \mathbf{O}$. So by property (1) in (5.9), $m_A(x) \mid q(A)$.

2. Suppose $m_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$, where $\lambda_i \neq \lambda_j \forall i \neq j$. By definition, $m_A(A) = \mathbf{O}$. Use (1), we have:

$$nullity(\prod_{i=1}^{k}(A-\lambda_i)) \leq \sum_{i=1}^{k} nullity(A-\lambda_i I) \leq \sum_{i=1}^{k} g(\lambda_i) = n.$$

   However, $nullity(m_A(A)) = nullity(\mathbf{O}) = n$. This forces the equality.

■

PROPOSITION 1 .
Let $B, C \in M_n(F)$, then $nullity(BC) \leq nullity(B) + nullity(C)$.

Proof.

First, show that $ker(C) \subseteq ker(BC)$. Extend a basis of $ker(C)$, $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n\}$ into a basis of $ker(BC)$, $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_p\}$.
Next, show $\{C\mathbf{w}_1, C\mathbf{w}_2, \cdots, C\mathbf{w}_p\} \subseteq ker(B)$. In particular, $\{C\mathbf{w}_1, C\mathbf{w}_2, \cdots, C\mathbf{w}_p\}$ is linearly independent. (Show the linear combination lies in $ker(C)$, and use the linear independence of the basis in $ker(BC)$). ■

THEOREM 5.12 (Schur Triangulation).
Every square matrix $A \in M_n(\mathbb{C})$ is similar to an upper triangular matrix.

Proof.

Prove by induction. Left as an exercise. ■

DEFINITION 5.7 (Spectrum).
The spectrum of a linear operator T is the multi-set contains the all eigenvalues of T, denoted by $spec(T) = \{\lambda_1, \lambda_2, \cdots \lambda_n\}$. Note that these elements might repeat.

THEOREM 5.13 (Spectral mapping theorem).
Let $A \in M_n(\mathbb{C})$ and $f \in \mathbb{C}[x]$ be a non-constant polynomial.

1. *If λ is an eigenvalue of A, then $f(\lambda)$ is an eigenvalue of $f(A)$.*
2. *If μ is an eigenvalue of $f(A)$, then $\exists$ an eigenvalue λ of A s.t. $\mu = f(\lambda)$.*

That is, if $spec(A) = \{\lambda_1, \lambda_2, \cdots \lambda_n\}$, then $spec(f(A)) = \{f(\lambda_1), f(\lambda_2), \cdots f(\lambda_n)\}$. So spectrum mapping is a bijection between multi-sets.

Proof.
Key result is that diagonal entries of upper triangular matrix is eigenvalues. Use Schur triangulation to deduce the spectrum mapping. ∎

5.3 Application of Diagonalization

DEFINITION 5.8 (T-invariant subspace).
Let $T : V \to V$ be a linear operator. A subspace $W \subseteq V$ is said to be a T-invariant subspace *if $T(W) \subseteq W$. Moreover, the restriction of T on W is defined by $T|_W : W \to W$, where $T|_W(\mathbf{w}) = T(\mathbf{w}) \, \forall \mathbf{w} \in W$.*

THEOREM 5.14 .
Let $W \subseteq V$ be an T-invariant subspace and let $T|_W : W \to W$ be the restriction. Then $m_{T|_W}(x)$ divdides $m_T(x)$.

Proof.
Key observation: $m_T(T) = \mathbf{0}$ implies $m_T(T|_W) = \mathbf{0}$. Use the property (1) in (5.9). ∎

EXERCISE 5.2 .
Verify $ker(T|_W) = ker(T) \cap W$.

THEOREM 5.15 (Simultaneous Eigenvector).
Let T_1, $T_2 : V \to V$ be two commutable linear operators of a complex finite-dimensional V, then there exists $\mathbf{v}$ s.t. $\mathbf{v}$ is simultaneously an eigenvector of both T_1 and T_2.

Proof.
Let $E_\lambda(T_1)$ be an eigenspace of T_1.

1. Claim that $E_\lambda(T_1)$ is T_2 invariant.

2. Consider the restriction $T_2|_{E_\lambda(T_1)} : E_\lambda(T_1) \to E_\lambda(T_1)$. Take an eigenvector $\mathbf{v}$ of $T_2|_{E_\lambda(T_1)}$ w.r.t. some eigenvalue λ'. This vector is simultaneously an eigenvector of $E_\lambda(T_1)$.

■

THEOREM 5.16 (Simultaneous Diagonalizability).
Let $T_1, T_2 : V \to V$ be two commutable and diagonalizable linear operators of a complex finite-dimensional V, then there exists a simultaneous eigen-basis S s.t. $[T_1]_S^S$ and $[T_2]_S^S$ are both diagonal and hence simultaneously diagonalizable.

Proof.
This is just an stronger verison of the previous theorem. ■

DEFINITION 5.9 (Exponential of Matrix).
Let $A \in M_n(\mathbb{C})$. The exponential of A is defined by:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

THEOREM 5.17 .
If $D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$ *is a diagonal matrix, then* $e^D = \begin{pmatrix} e^{d_1} & 0 & \cdots & 0 \\ 0 & e^{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{d_n} \end{pmatrix}$

THEOREM 5.18 .
Let $A, B \in M_n(\mathbb{C})$, then:

1. *e^A converges absolutely.*
2. *If $AB = BA$, then $e^{A+B} = e^A e^B$.*

Proof.
The above theorems are stated without proof, as they may require knowledge of normed spaces, which is beyond the current scope. ■

THEOREM 5.19 .
Let $A \in M_n(\mathbb{C})$.

1. *If A B, then e^A e^B.*
2. *Consider a function $f : \mathbb{C} \to M_n(\mathbb{C})$ defined by $f(t) = e^{At}$, then*

$$f'(t) = \frac{d}{dt} e^{At} = Ae^{At}.$$

COROLLARY .
Give a system of differential equations:

$$\frac{d}{dt}\mathbf{v}(t) = A\mathbf{v}(t),$$

where $A \in M_n(\mathbb{C})$, then the solution is given by:

$$\mathbf{v}(t) = e^{At}\mathbf{v}(0).$$

REMARK .
To compute e^{At}, we need to adress higher powers of A, then diagonalization or Cayley-Hamilton is useful here.

EXERCISE 5.3 .
Give a system of difference equations:

$$\mathbf{v}_n = A\mathbf{v}_{n-1},$$

where $A \in M_n(\mathbb{C})$, then the solution is given by:

$$\mathbf{v}_n = A^n\mathbf{v}_0.$$

6 Linear Duality

Let V and W be two finite-dimensional vector spaces over a field $\mathcal{F}$ s.t. $dim(V) = n$ and $dim(W) = m$. Recall our *dream diagram*:

$$\begin{array}{ccc} V & \xrightarrow{\mathcal{L}(V,W)} & W \\ \Big\downarrow{\scriptstyle \phi_S} & & \Big\downarrow{\scriptstyle \phi_{S'}} \\ \mathcal{F}^n & \xrightarrow[\mathcal{M}_{m\times n}(\mathcal{F})]{} & \mathcal{F}^m \end{array}$$

Fix S as a basis of V, the coordinate mapping $\phi : V \to F^n$ defined by $\mathbf{v} \mapsto [\mathbf{v}]_S$ is an isomorphism. So does the coordinate mapping $\phi_{S'}$.

Now consider the vector space of all linear functions from V to W, denoted by $\mathcal{L}(V, W) = \{f : V \to W : f \text{ is linear}\}$, then we have:

THEOREM 6.1 .
Fix bases S and S' of V and W respectively, then there exists an isomorphism of vector spaces:

$$\Phi_S^{S'} : \mathcal{L}(V, W) \to \mathcal{M}_{m\times n}(\mathcal{F}),$$

defined by $f \mapsto [f]_S^{S'}$.

COROLLARY .
If both V, W are finite dimensional over $\mathcal{F}$, then

$$dim_{\mathcal{F}}(\mathcal{L}(V, W)) = dim_{\mathcal{F}}(V) dim_{\mathcal{F}}(W).$$

Next, we apply the above result to the special case when $W = \mathcal{F}$.

DEFINITION 6.1 (Dual Space).
Let V be a vector space over a field $\mathcal{F}$. The dual space of V, denoted by V^*, is defined by:

$$V^* = \mathcal{L}(V, \mathcal{F}) = \{f : V \to \mathcal{F} : f \text{ is linear}\}.$$

Take any element $f \in V^*$, we call $f : V \to \mathcal{F}$ is linear functional on V.

THEOREM 6.2 .
Fix a basis S of V and $B = \{1\}$ be a standard basis of $\mathcal{F}$, then there exists an isomorphism of vector spaces:

$$\Phi_S^B : \mathcal{L}(V, \mathcal{F}) \to \mathcal{M}_{1\times n}(\mathcal{F}),$$

defined by $f \mapsto [f]_S^B$.

COROLLARY .
If both V is finite dimensional over $\mathcal{F}$, then

$$dim_{\mathcal{F}}(V^*) = dim_{\mathcal{F}}(V).$$

DEFINITION 6.2 (Dual Basis).
Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n\}$ be a basis of V. $S^* = \{\mathbf{v}_1^*, \mathbf{v}_2^*, \cdots \mathbf{v}_n^*\}$ is said to be the dual basis of S in V^* if $\phi_S^B(\mathbf{v}_k^*) = e_k$ for each $k = 1, 2, \cdots n$.

LEMMA .
Fix any basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n\}$ be a basis of V, then its dual basis $S^* = \{\mathbf{v}_1^*, \mathbf{v}_2^*, \cdots \mathbf{v}_n^*\}$ satisfies, for each i:

$$\begin{cases} \mathbf{v}_i^*(\mathbf{v}_j) = 1, \text{ if } i = j, \\ \mathbf{v}_i^*(\mathbf{v}_j) = 0, \text{ if } i \neq j. \end{cases}$$

Proof.

For $i = 1$, consider $\mathbf{v}_1^* : V \to \mathcal{F}$. By definition, $\Phi_S^B(\mathbf{v}_1^*) = e_1$, i.e., $[v_1^*]_S^B = (1, 0, \cdots, 0)$. This implies that $\mathbf{v}_1^*(\mathbf{v}_1) = 1 \cdot 1$ and $\mathbf{v}_1^*(\mathbf{v}_j) = 0 \cdot 1$ for $j \neq 1$. Now repeat the above argument for $i = 2, 3, \cdots n$. ∎

DEFINITION 6.3 (Dual linear maps).
Let $T : V \to W$ be a linear map btw two vector spaces V and W. The dual of T is a linear map btw the dual spaces of V and W:

$$T^* : W^* \to V^* \text{ given by } w^* \mapsto w^* \circ T.$$

THEOREM 6.3 .
Let $T : V \to W$ be a linear map btw two finite-dimensional vector spaces V and W. Fix bases S and S' of V and W respectively, then

$$[T^*]_{S'^*}^{S^*} = ([T]_S^{S'})^T.$$

Proof.

Let $S = \{\mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n\}$ be a basis of V. Let $S' = \{\mathbf{w}_1, \mathbf{w}_2 \cdots \mathbf{w}_m\}$ be a basis of W. Suppose $[T]_S^{S'} = A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$, we want to show $[T^*]_{S'^*}^{S^*} = A^\top$.

Let $\mathbf{v} = \sum_{k=1}^n c_k \mathbf{v}_k$. Consider $T^*(\mathbf{w}_1^*)(\mathbf{v}) = \mathbf{w}_1^* \circ T(\mathbf{v}) = \mathbf{w}_1^*(\sum_{k=1}^n c_k T(\mathbf{v}_k)) = \mathbf{w}_1^*(\sum_{k=1}^n c_k \sum_{j=1}^m a_{jk}\mathbf{w}_j) = \sum_{k=1}^n c_k a_{1k} \cdot 1$. Note that $c_k = v_k^*(\mathbf{v})$, so we get $T^*(\mathbf{w}_1^*) = b_1^* \circ T(\mathbf{v}) = \sum_{k=1}^n a_{1k} v_k^*(\mathbf{v})$. ∎

DEFINITION 6.4 (Annihilator).
Let V be a vector space and $U \subseteq V$ be a subspace. The annihilator of U is a subspace of V^* defined by:

$$U^0 = \{f \in V^* : f(\mathbf{u}) = 0 \ \forall \mathbf{u} \in U\}.$$

THEOREM 6.4 (Linear Duality Theorem (I)).
Let V be a finite dimensional vector space and $U \subseteq V$, then:

$$dim(V) = dim(U) + dim(U^0).$$

Proof.

Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_k\}$ be a basis of U. By basis extension theorem, we can extend it into $\{\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_k, \mathbf{u}_{k+1}, \cdots \mathbf{u}_n\}$. So $\{\mathbf{u}_1^*, \mathbf{u}_2^*, \cdots \mathbf{u}_k^*, \mathbf{u}_{k+1}^*, \cdots \mathbf{u}_n^*\}$ is the dual basis in V^*. Next, we want to show $\{\mathbf{u}_{k+1}^*, \cdots \mathbf{u}_n^*\}$ forms a basis of U^0.

1. For each $k+1 \leq j \leq n$, $\mathbf{u}_j^*(\mathbf{u}_i) = 0$ for $1 \leq i \leq k$. So $\mathbf{u}_j^* \in U^0$.
2. Let $\theta \in U^0 \subseteq V^*$, so $\theta = \sum_{i=1}^n a_i \mathbf{u}_i^*$, and we wish $a_1 = a_2 = \cdots a_k = 0$. Let $\mathbf{u} \in U$, so $\mathbf{u} = \sum_{i=1}^k \mathbf{u}_i$. By definition, $\theta(\mathbf{u_j}) = 0$ forces $\sum_{i=1}^k a_i \mathbf{u}_i^*(\mathbf{u}_j) = 0$ for $1 \leq j \leq k$.
3. Linear independence is trivial.

By the claim, we have $dim(U^0) = n - k = dim(V) - dim(U)$. ∎

THEOREM 6.5 (Linear Duality Theorem (II)).
Let $T : V \to W$ be a linear map btw finite dimensional vector spaces. Then we have:

1. $ker(T^*) = (im(T))^0$.
2. $rank(T) = rank(T^*)$.
3. $im(T^*) = (ker(T))^0$.

Proof.

Left as an exercise. ∎

COROLLARY .
(6.5) *implies that*

1. *T^* is injective iff T is surjective.*
2. $rank(A) = rank(A^\top)$.

3. *T^* is surjective iff T is injective.*

Proof.

1. $im(T) = W \leftrightarrow dim((im(T)^0)) = 0 = dim(ker(T^*))$.
2. Recall theorem (6).
3. $ker(T) = \{\mathbf{0}\} \leftrightarrow dim((ker(T))^0) = dim(im(T^*)) = dim(V) = dim(V^*)$

■

DEFINITION 6.5 (Double Dual).
The dual of a dual space is defined by:

$$(V^*)^* = \{g : V^* \to F : g \text{ is linear}\}.$$

THEOREM 6.6 (Linear Duality Theorem (III)).
Let $ev : V \to (V^*)^*$ defined by $\mathbf{v} \mapsto ev(\mathbf{v})$. That is, for each $\mathbf{v} \in V$, $ev(\mathbf{v}) : V^* \to \mathcal{F}$ is a linear functional on V^*. So the ev function satisfies that for $f \in V^*$, $ev(\mathbf{v})(f) = f(\mathbf{v})$. Under this setting, we have:

1. *ev is an injective linear map.*
2. *If V is finite dimensional, then ev is an isomorphism btw V and $(V^*)^*$.*

Proof.

1. Take $\mathbf{v} \in ker(ev)$, we have $ev(\mathbf{v}) = \mathbf{0}$. That is, $ev(\mathbf{v})(f) = f(\mathbf{v}) = 0$ for $f \in V^*$. Since f is aribitrary in the dual space V^*, this forces $\mathbf{v} = \mathbf{0}$ and $ker(ev) = \{\mathbf{0}\}$. This proves the injectiveness.
2. Note that $dim(V) = dim(V^*) = dim((V^*)^*)$, ev:$V \to (V^*)^*$ is injective linear map from 1, by (3.26), ev is also a surjective map and hence an isomorphism.

■

REMARK .
If V is infinite dimensional, then ev *might not be surjective in general, which is beyond the current scope.*