Linear Algebra with Exercises B

Fall 2017 Kyoto University

Ivan Ip

These notes summarize the definitions, theorems and some examples discussed in class. Please refer to the class notes and reference books for proofs and more in-depth discussions.

Contents

Introduction

Real life problems are **hard**.

Linear Algebra is **easy** (in the mathematical sense).

We make *linear approximations* to real life problems, and reduce the problems to **systems of linear equations** where we can then use the techniques from Linear Algebra to solve for approximate solutions. Linear Algebra also gives new insights and tools to the original problems.

Real Life Problems		Linear Algebra
Optimization Problems	$\longrightarrow$	Tangent Spaces
Economics	$\longrightarrow$	Linear Regression
Stochastic Process	$\longrightarrow$	Transition Matrices
Engineering	$\longrightarrow$	Vector Calculus
Data Science	$\longrightarrow$	Principal Component Analysis
Signals Processing	$\longrightarrow$	Fourier Analysis
Artificial Intelligence	$\longrightarrow$	Deep Learning
Computer Graphics	$\longrightarrow$	Euclidean Geometry

Roughly speaking,

Real Life Problems		Linear Algebra
Data Sets	$\longleftrightarrow$	Vector Spaces
Relationship between data	$\longleftrightarrow$	Linear Transformations

In *Linear Algebra with Exercises A*, we learned techniques to solve linear equations, such as row operations, reduced echelon forms, existence and uniqueness of solutions, basis for null space etc.

In part B of the course, we will focus on the more abstract part of linear algebra, and study the descriptions, structures and properties of **vector spaces** and **linear transformations**.

Mathematical Notations

Numbers:

- $\mathbb{R}$: The set of real numbers
- $\mathbb{R}^n$: n-tuples of real numbers
- $\mathbb{Q}$: Rational numbers
- $\mathbb{C}$: Complex numbers $x + iy$ where $i^2 = -1$
- $\mathbb{N}$: Natural numbers $\{1, 2, 3, ...\}$
- $\mathbb{Z}$: Integers $\{..., -2, -1, 0, 1, 2, ...\}$

Sets:

- $x \in X$: x is an element in the set X
- $S \subset X$: S is a subset of X
- $S \subsetneq X$: S is a subset of X but not equals to X
- $X \times Y$: Cartesian product, the set of pairs $\{(x, y) : x \in X, y \in Y\}$
- $|S|$: Cardinality (size) of the set S
- ϕ : Empty set
- $V \longrightarrow W$: A map from V to W
- $x \mapsto y$: A map sending x to y, ("x maps to y")

Logical symbols:

- $\forall$: "For every"
- $\exists$: "There exists"
- $\exists!$: "There exists unique"
- $:=$: "is defined as"
- $\Longrightarrow$: "implies"
- $\Longleftrightarrow$ or iff : "if and only if". A iff B means
  - if: $B \Longrightarrow A$
  - only if: $A \Longrightarrow B$
- Contrapositive: $A \Longrightarrow B$ is equivalent to $Not(B) \Longrightarrow Not(A)$

CHAPTER 1

Abstract Vector Spaces

1.1 Vector Spaces

Let K be a *field*, i.e. a "number system" where you can add, subtract, multiply and divide.
In this course **we will take K to be $\mathbb{R}, \mathbb{C}$ or $\mathbb{Q}$**.

Definition 1.1. A **vector space** over K is a set V together with two operations: $+$ (*addition*) and $\cdot$ (*scalar multiplication*) subject to the following 10 rules for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c, d \in K$:

(+1) Closure under addition: $\mathbf{u} \in V, \mathbf{v} \in V \Longrightarrow \mathbf{u} + \mathbf{v} \in V$.

(+2) Addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

(+3) Addition is associative: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

(+4) Zero exists: there exists $\mathbf{0} \in V$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.

(+5) Inverse exists: for every $\mathbf{u} \in V$, there exists $\mathbf{u}' \in V$ such that $\mathbf{u} + \mathbf{u}' = \mathbf{0}$. We write $\mathbf{u}' := -\mathbf{u}$.

($\cdot$1) Closure under multiplication: $c \in K, \mathbf{u} \in V \Longrightarrow c \cdot \mathbf{u} \in V$.

($\cdot$2) Multiplication is associative: $(cd) \cdot \mathbf{u} = c \cdot (d \cdot \mathbf{u})$.

($\cdot$3) Multiplication is distributive: $c \cdot (\mathbf{u} + \mathbf{v}) = c \cdot \mathbf{u} + c \cdot \mathbf{v}$.

($\cdot$4) Multiplication is distributive: $(c + d) \cdot \mathbf{u} = c \cdot \mathbf{u} + d \cdot \mathbf{u}$.

($\cdot$5) Unity: $1 \cdot \mathbf{u} = \mathbf{u}$.

The elements of a vector space V are called *vectors*.

Note. We will denote a vector with boldface $\mathbf{u}$ in this note, but you should use $\vec{u}$ for handwriting. Sometimes we will omit the $\cdot$ for scalar multiplication if it is clear from the context.

Note. Unless otherwise specified, all vector spaces in the examples below is over $\mathbb{R}$.

The following facts follow from the definitions

Properties 1.2. For any $u \in V$ and $c \in K$:

- The zero vector $\mathbf{0} \in V$ is unique.
- The negative vector $-\mathbf{u} \in V$ is unique.
- $0 \cdot \mathbf{u} = \mathbf{0}$.
- $c \cdot \mathbf{0} = \mathbf{0}$.
- $-\mathbf{u} = (-1) \cdot \mathbf{u}$.

Examples of vector spaces over $\mathbb{R}$:

Example 1.1. The space $\mathbb{R}^n$, $n \geq 1$ with the usual vector addition and scalar multiplication.

Example 1.2. $\mathbb{C}$ is a vector space over $\mathbb{R}$.

Example 1.3. The subset $\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + y + z = 0\} \subset \mathbb{R}^3$.

Example 1.4. Real-valued functions $f(t)$ defined on $\mathbb{R}$.

Example 1.5. The set of real-valued differentiable functions satisfying the differential equations

$$f + \frac{d^2 f}{dx^2} = 0.$$

Examples of vector spaces over a field K:

Example 1.6. The zero vector space $\{\mathbf{0}\}$.

Example 1.7. Polynomials with coefficients in K:

$$p(t) = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n$$

with $a_i \in K$ for all i.

Example 1.8. The set $M_{m \times n}(K)$ of $m \times n$ matrices with entries in K.

Counter-Examples: these are not vector spaces:

Non-Example 1.9. $\mathbb{R}$ is not a vector space over $\mathbb{C}$.

Non-Example 1.10. The first quadrant $\{\begin{pmatrix} x \\ y \end{pmatrix} : x \geq 0, y \geq 0\} \subset \mathbb{R}^2$.

Non-Example 1.11. The set of all invertible 2×2 matrices.

Non-Example 1.12. Any straight line in $\mathbb{R}^2$ not passing through the origin.

Non-Example 1.13. The set of polynomials of degree exactly n.

Non-Example 1.14. The set of functions satisfying $f(0) = 1$.

1.2 Subspaces

To check whether a subset $H \subset V$ is a vector space, we only need to check zero and closures.

Definition 1.3. A **subspace** of a vector space V is a subset H of V such that

(1) $\mathbf{0} \in H$.

(2) Closure under addition: $\mathbf{u} \in H, \mathbf{v} \in H \Longrightarrow \mathbf{u} + \mathbf{v} \in H$.

(3) Closure under multiplication: $\mathbf{u} \in H, c \in K \Longrightarrow c \cdot \mathbf{u} \in H$.

Example 1.15. Every vector space has a zero subspace $\{\mathbf{0}\}$.

Example 1.16. A plane in $\mathbb{R}^3$ through the origin is a subspace of $\mathbb{R}^3$.

Example 1.17. Polynomials of degree at most n with coefficients in K, **written as** $\mathbb{P}_n(K)$, is a subspace of the vector space of all polynomials with coefficients in K.

Example 1.18. Real-valued functions satisfying $f(0) = 0$ is a subspace of the vector space of all real-valued functions.

Non-Example 1.19. Any straight line in $\mathbb{R}^2$ not passing through the origin is not a vector space.

Non-Example 1.20. $\mathbb{R}^2$ is **not** a subspace of $\mathbb{R}^3$. But $\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x \in \mathbb{R}, y \in \mathbb{R}\} \subset \mathbb{R}^3$, which looks exactly like $\mathbb{R}^2$, is a subspace.

Definition 1.4. Let $S = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ be a set of vectors in V. A **linear combination** of S is a any sum of the form

$$c_1\mathbf{v}_1 + ... + c_p\mathbf{v}_p \in V, \qquad c_1, ..., c_p \in K.$$

The set **spanned** by S is the set of all *linear combinations* of S, denoted by $Span(S)$.

Remark. More generally, if S is an infinite set, we define

$$Span(S) = \left\{ \sum_{i=1}^{N} c_i\mathbf{v}_i : c_i \in K, \mathbf{v}_i \in S \right\}$$

i.e. the set of all linear combinations which are finite sum. It follows that $Span(V) = V$ if V is a vector space.

Theorem 1.5. $Span(S)$ is a subspace of V.

Theorem 1.6. H is a subspace of V if and only if H is non-empty and closed under linear combinations, i.e.

$$c_i \in K, \mathbf{v}_i \in H \Longrightarrow c_1\mathbf{v}_1 + ... + c_p\mathbf{v}_p \in H.$$

Example 1.21. The set $H := \{ \begin{pmatrix} a-3b \\ b-a \\ a \\ b \end{pmatrix} \in \mathbb{R}^4 : a, b \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^4$, since every element of H can be written as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$a \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix} = a\mathbf{v}_1 + b\mathbf{v}_2 \in \mathbb{R}^4.$$

Hence $H = Span(\mathbf{v}_1, \mathbf{v}_2)$ is a subspace by Theorem 1.6.

1.3 Linearly Independent Sets

Definition 1.7. A set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_p\} \subset V$ is **linearly dependent** if

$$c_1\mathbf{v}_1 + ... + c_p\mathbf{v}_p = \mathbf{0}$$

for some $c_i \in K$, not all of them zero.

Linearly independent set are those vectors that are not linearly dependent:

Definition 1.8. A set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_p\} \subset V$ is **linearly independent** if

$$c_1\mathbf{v}_1 + ... + c_p\mathbf{v}_p = \mathbf{0}$$

implies $c_i = 0$ for all i.

Example 1.22. A set of one element $\{\mathbf{v}\}$ is linearly independent iff $\mathbf{v} \neq \mathbf{0}$.

Example 1.23. A set of two nonzero element $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent iff $\mathbf{u}$ is not a multiple of $\mathbf{v}$.

Example 1.24. Any set containing $\mathbf{0}$ is linearly dependent.

Example 1.25. The set of vectors $\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\}$ is linearly independent.

Example 1.26. The set of polynomials $\{t^2, t, 4t - t^2\}$ is linearly dependent.

Example 1.27. The set of functions $\{\sin t, \cos t\}$ is linearly independent. The set $\{\sin 2t, \sin t \cos t\}$ is linearly dependent.

1.4 Bases

Definition 1.9. Let H be a subspace of a vector space V. A set of vectors $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_m\} \subset V$ is a **basis** for H iff

(1) $\mathcal{B}$ is a linearly independent set.

(2) $H = \mathit{Span}(\mathcal{B})$.

Note. The plural of "basis" is **"bases"**.

Example 1.28. The columns of the $n \times n$ identity matrix I_n:

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, ..., \mathbf{e}_n = \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$$

form the **standard basis** for $\mathbb{R}^n$.

Example 1.29. In general, the columns of an invertible matrix $A \in M_{n\times n}(\mathbb{R})$ form a basis of $\mathbb{R}^n$, because $A\mathbf{x} = \mathbf{0}$ only has trivial solution.

Example 1.30. The polynomials $\{1, t, t^2, ..., t^n\}$ from the **standard basis** for $\mathbb{P}_n(\mathbb{R})$.

Theorem 1.10 (Spanning Set Theorem)**.** Let $S = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ be a set in V and let $H = Span(S)$.

(1) If one of the vectors, say $\mathbf{v}_k$, is a linear combination of the remaining vectors in S, then $H = Span(S \setminus \{\mathbf{v}_k\})$.

(2) If $H \neq \{\mathbf{0}\}$, some subset of S is a basis of H.

Theorem 1.11 (Unique Representation Theorem)**.** If $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\} \subset V$ is a basis for V, then for each $x \in V$, there exists **unique** scalars $c_1, ..., c_n \in K$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + ... + c_n\mathbf{b}_n.$$

$c_1, ..., c_n$ are called the **coordinates** of $\mathbf{x}$ relative to the basis $\mathcal{B}$, and

$$[x]_{\mathcal{B}} := \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n$$

is the **coordinate vector** of $\mathbf{x}$ relative to $\mathcal{B}$.

Example 1.31. The coordinate vector of the polynomial $\mathbf{p} = t^3 + 2t^2 + 3t + 4 \in \mathbb{P}_3(\mathbb{R})$ relative to the basis $\mathcal{B} = \{1, t, t^2, t^3\}$ is

$$[\mathbf{p}]_{\mathcal{B}} = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^4.$$

We will study the change of basis later.

1.5 Dimensions

Theorem 1.12 (Replacement Theorem)**.** If $V = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$, and $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ is linearly independent set in V, then $m \leq n$.

Proof. (Idea) One can *replace* some $\mathbf{v}_i$ by $\mathbf{u}_1$ so that $\{\mathbf{u}_1, \mathbf{v}_1, ..., \mathbf{v}_n\} \setminus \{\mathbf{v}_i\}$ also spans V. Assume on the contrary that $m > n$. Repeating the process we can replace all $\mathbf{v}$'s by $\mathbf{u}$'s, so that $\{\mathbf{u}_1, ..., \mathbf{u}_n\}$ spans V, hence $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ is linearly dependent. □

Applying this statement to different bases $\mathcal{B}$ and $\mathcal{B}'$, which are both spanning and linearly independent, we get

Theorem 1.13. If a vector space V has a basis of n vectors, then every basis of V must also consists of exactly n vectors.

By this Theorem, the following definition makes sense:

Definition 1.14. If V is spanned by a finite set, then V is said to be **finite dimensional**, $dim(V) < \infty$. The **dimension** of V is the number of vectors in any basis $\mathcal{B}$ of V:

$$\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\} \Longrightarrow \dim V := |\mathcal{B}| = n.$$

If $V = \{\mathbf{0}\}$ is the zero vector space, $\dim V := 0$..

If V is not spanned by a finite set, it is **infinite dimensional**, $dim(V) := \infty$.

Note. If the vector space is over the field K we will write $\dim_K V$. If it is over $\mathbb{R}$ or if the field is not specified (as in the Definition above), we simply write $\dim V$ instead.

Example 1.32. $\dim \mathbb{R}^n = n$.

Example 1.33. $\dim_K \mathbb{P}_n(K) = n + 1$. The space of all polynomials is infinite dimensional.

Example 1.34. $\dim_K M_{m\times n}(K) = mn$.

Example 1.35. Let $V = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$. Then $\dim V = 2$.

Example 1.36. The space of real-valued functions on $\mathbb{R}$ is infinite dimensional.

Example 1.37. $\dim_{\mathbb{R}} \mathbb{C} = 2$ but $\dim_{\mathbb{C}} \mathbb{C} = 1$. $\dim_{\mathbb{R}} \mathbb{R} = 1$ but $\dim_{\mathbb{Q}} \mathbb{R} = \infty$.

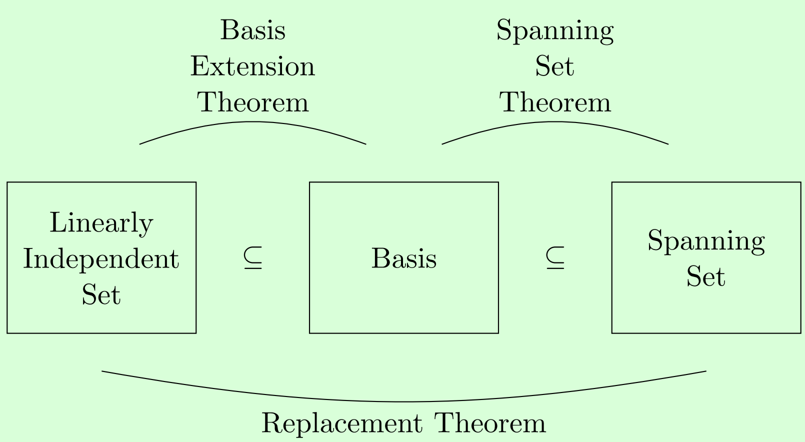
Theorem 1.15 (Basis Extension Theorem)**.** Let H be a subspace of V with $\dim V < \infty$. Any linearly independent set in H can be expanded, if necessary, to a basis for H.
Also, H is finite dimensional and

$$\dim H \leq \dim V.$$

Example 1.38. Subspaces of $\mathbb{R}^3$ are classified as follows:

- 0-dimensional subspaces: only the zero space $\{\mathbf{0}\}$.
- 1-dimensional subspaces: any line passing through origin.
- 2-dimensional subspaces: any plane passing through origin.
- 3-dimensional subspaces: only $\mathbb{R}^3$ itself.

Summarize: For $\dim V < \infty$ and $H \subset V$ a subspace:



If you know the dimension of V, the following Theorem gives a useful criterion to check whether a set is a basis:

Theorem 1.16 (The Basis Theorem)**.** Let V be an n-dimensional vector space, $n \geq 1$, and $S \subset V$ a finite subset with exactly n elements. Then

(1) If S is linearly independent, then S is a basis for V.

(2) If S spans V, then S is a basis for V.

1.6 Intersections, Sums and Direct Sums

We discuss three important construction of vector spaces.

Definition 1.17. Let U, W be subspaces of V.

- $U \cap W$ is the **intersection** of U and W.
- $U + W = \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in U, \mathbf{v} \in W\} \subset V$ is the **sum** of U and W.

Properties 1.18.

- $U \cap W$ and $U + W$ are both vector subspaces of V.

Definition 1.19. Let U, W be subspaces of V. Then V is called the **direct sum** of U and W, written as $V = U \oplus W$ if

(1) $V = U + W$.

(2) $U \cap W = \{\mathbf{0}\}$.

Example 1.39. $\mathbb{R}^3 = \{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R}\} \oplus \{ \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} : z \in \mathbb{R}\}$.

Example 1.40. $\{\text{Space of functions}\} = \left\{ \begin{array}{c} \text{Even functions} \\ f(-t) = f(t) \end{array} \right\} \oplus \left\{ \begin{array}{c} \text{Odd functions} \\ f(-t) = -f(t) \end{array} \right\}$.

Example 1.41. $\{\text{Matrices}\} = \left\{ \begin{array}{c} \text{Symmetric matrices} \\ \mathbf{A}^T = \mathbf{A} \end{array} \right\} \oplus \left\{ \begin{array}{c} \text{Anti-symmetric matrices} \\ \mathbf{A}^T = -\mathbf{A} \end{array} \right\}$.

Example 1.42. $\{\text{Polynomials}\} = \{\text{Constants}\} \oplus \{\mathbf{p}(t) : \mathbf{p}(0) = 0\}$.

Theorem 1.20. $V = U \oplus W$ iff every $\mathbf{v} \in V$ can be written **uniquely** as

$$\mathbf{v} = \mathbf{u} + \mathbf{w}$$

where $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Theorem 1.21 (Dimension formula)**.**

$$\dim U + \dim W = \dim(U + W) + \dim(U \cap W).$$

In particular

$$\dim U + \dim W = \dim(U \oplus W).$$

Example 1.43. If U and W are two *different* planes passing through origin in $\mathbb{R}^3$, then $U \cap W$ must be a line and $U + W = \mathbb{R}^3$. The dimension formula then gives $2 + 2 = 3 + 1$.

CHAPTER 2

Linear Transformations and Matrices

2.1 Linear Transformations

Definition 2.1. A **linear transformation** T from a vector space V to a vector space W is a map

$$T : V \longrightarrow W$$

such that for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalar $c \in K$:

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(2) $T(c \cdot \mathbf{u}) = c \cdot T(\mathbf{u})$

V is called **domain** and W is called **codomain** of T.

The set of all such linear transformations $T : V \longrightarrow W$ is denoted by $L(V, W)$.

Fact. Any $T \in L(V, W)$ is uniquely determined by the image on any basis $\mathcal{B}$ of V.

Example 2.1. The **identity** map $\mathbf{Id} : V \longrightarrow V$ given by $\mathbf{Id}(\mathbf{v}) = \mathbf{v}$.

Example 2.2. Differential operators on the space of real-valued differentiable functions.

Example 2.3. $Tr : M_{3\times 3}(\mathbb{R}) \longrightarrow \mathbb{R}$ on the space of 3×3 matrices with real entries.

Example 2.4. Matrix multiplication: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$.

Theorem 2.2. Let V, W be **finite dimensional** vector spaces with $\dim V = n, \dim W = m$. Then every linear transformation $T \in L(V, W)$ can be represented by an $m \times n$ matrix.

Let $\mathcal{B} \subset V, \mathcal{B}' \subset W$ be their bases. If $T(\mathbf{u}) = \mathbf{w}$, then the matrix is given by

$$\begin{aligned} [T]_{\mathcal{B}'}^{\mathcal{B}} : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ [\mathbf{u}]_{\mathcal{B}} &\mapsto [\mathbf{w}]_{\mathcal{B}'} \end{aligned}$$

It is called the **matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$**.

Definition 2.3.

- The **kernel** or **null space** of T is:

$$\mathrm{Ker}(T) := \{\mathbf{u} \in V : T(\mathbf{u}) = \mathbf{0}\}$$

- The **image** or **range** of T is:

$$\mathrm{Im}(T) := \{\mathbf{w} \in W : \mathbf{w} = T(\mathbf{u}) \text{ for some } \mathbf{u} \in V\}$$

Theorem 2.4. Let $T \in L(V, W)$. Then

- The kernel of T is a subspace of V.
- The image of T is a subspace of W.

Example 2.5. If $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is represented by a matrix $\mathbf{A}$, then

- The kernel of $\mathbf{A}$ is the **null space** Nul$\mathbf{A}$. It is the set of all solutions to $\mathbf{Ax} = \mathbf{0}$ of m homogeneous linear equations in n unknown. It is a subspace of $\mathbb{R}^n$.
- The image of $\mathbf{A}$ is the **column space** Col$\mathbf{A}$. It is the set of all linear combinations of the columns of $\mathbf{A}$. It is a subspace of $\mathbb{R}^m$.

Rule. For a matrix $\mathbf{A}$, the row operations do not affect linear dependence of the columns.
Use the *reduced echelon form* to find the basis of Nul$\mathbf{A}$ and Col$\mathbf{A}$.

Example 2.6. The kernel of $\frac{d}{dx}$ on the space of differentiable functions is the set of all constant functions.

2.2 Injection, Surjection and Isomorphism

Definition 2.5. A linear transformation $T : V \longrightarrow W$ is called

- **one-to-one** or **injective** if $T(\mathbf{u}) = T(\mathbf{v})$ implies $\mathbf{u} = \mathbf{v}$
- **onto** or **surjective** if for every $\mathbf{w} \in W$, there exists $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{w}$
- **isomorphism** if T is one-to-one and onto.

Definition 2.6. If there exists an isomorphism $T \in L(V, W)$, we say V is **isomorphic** to W, written as $V \simeq W$.

Properties 2.7. Let $T \in L(V, W)$.

- If T is injective, then $\mathcal{N}(T) = \{\mathbf{0}\}$, i.e. $T(\mathbf{u}) = \mathbf{0} \Longrightarrow \mathbf{u} = \mathbf{0}$.
- If T is injective, it maps linearly independent set to linearly independent set.
- If T is injective, and $H \subset V$ is a subspace, then $\dim T(H) = \dim H$.
- If T is surjective, then $\mathcal{R}(T) = W$.
- If T is isomorphism, then $\dim V = \dim W$.

Theorem 2.8. If $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ is a basis for a vector space V, then the **coordinate mapping**

$$\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$$

is an isomorphism $V \simeq K^n$.

Example 2.7. $\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R}\} \neq \mathbb{R}^2$ but $\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R}\} \simeq \mathbb{R}^2$

Example 2.8. $\mathbb{P}_n(K) \simeq K^{n+1}$

Example 2.9. $M_{m\times n}(K) \simeq K^{mn}$

Example 2.10. $\mathbb{C} \simeq \mathbb{R}^2$ as vector spaces over $\mathbb{R}$.

2.3 Rank

Definition 2.9.

- The **rank** of T is the dimension of the image of T.
- The **nullity** of T is the dimension of the kernel of T.

Below we have the **Fundamental Theorem of Linear Algebra**, which consists of the Rank-Nullity Theorem and the Theorem of Column Rank = Row Rank :

Theorem 2.10 (Rank–Nullity Theorem)**.** Let $T \in L(V, W)$. Then

$$\dim \mathrm{Im}(T) + \dim \mathrm{Ker}(T) = \dim V$$

Let T be represented by a $m \times n$ matrix $\mathbf{A}$ (i.e. $V = \mathbb{R}^n, W = \mathbb{R}^m$). Then **rank** of $\mathbf{A}$ is the dimension of Col$\mathbf{A}$. The **row space** is the space spanned by the rows of $\mathbf{A}$. It is a subspace of $\mathbb{R}^n$. The **row rank** is the dimension of the row space.

The row space of $\mathbf{A}$ is the column space of $\mathbf{A}^T$, hence the row rank of $\mathbf{A}$ equals the rank of $\mathbf{A}^T$.

Theorem 2.11 (Column Rank = Row Rank)**.** Rank of $\mathbf{A}$ = Rank of $\mathbf{A}^T$.

Theorem 2.12 (Invertible Matrix Theorem)**.** Let $\mathbf{A}$ be an $n \times n$ matrix. Then $\mathbf{A}$ is invertible iff any one of the statements hold:

(1) Columns of $\mathbf{A}$ form a basis of $\mathbb{R}^n$

(2) Col$\mathbf{A} = \mathbb{R}^n$

(3) Rank of $\mathbf{A} = n$

(4) Nul$\mathbf{A} = \{\mathbf{0}\}$

(5) Nullity of $\mathbf{A} = 0$.

2.4 Change of Basis

Recall that for $\mathbf{x} \in V$ and $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ a basis of V,

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n$$

is the $\mathcal{B}$-coordinate vector of $\mathbf{x}$ if

$$\mathbf{x} = c_1\mathbf{b}_1 + ... + c_n\mathbf{b}_n.$$

If $\mathcal{B}' = \{\mathbf{b}'_1, ..., \mathbf{b}'_n\}$ is another basis of V, then

$$[\mathbf{x}]_{\mathcal{B}'} = \begin{pmatrix} c'_1 \\ \vdots \\ c'_n \end{pmatrix} \in \mathbb{R}^n$$

is the $\mathcal{B}'$-coordinate vector of $\mathbf{x}$ if

$$\mathbf{x} = c'_1\mathbf{b}'_1 + ... + c'_n\mathbf{b}'_n.$$

The relationship between the vectors $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{B}'}$ is given by

Theorem 2.13 (Change of Basis formula)**.** There exists an $n \times n$ matrix $P^{\mathcal{B}}_{\mathcal{B}'}$ such that

$$[\mathbf{x}]_{\mathcal{B}'} = P^{\mathcal{B}}_{\mathcal{B}'} \cdot [\mathbf{x}]_{\mathcal{B}}$$

where *column-wise* it is given by

$$P^{\mathcal{B}}_{\mathcal{B}'} = \begin{pmatrix} [\mathbf{b}_1]_{\mathcal{B}'} & [\mathbf{b}_2]_{\mathcal{B}'} & \cdots & [\mathbf{b}_n]_{\mathcal{B}'} \end{pmatrix}$$

$P^{\mathcal{B}}_{\mathcal{B}'}$ is called the **change-of-coordinate matrix** from $\mathcal{B}$ to $\mathcal{B}'$.

In other words, it is the matrix of the identity map $\mathbf{Id}$ with respect to the basis $\mathcal{B}$ and $\mathcal{B}'$ (see Theorem 2.2).

Properties 2.14. The $n \times n$ matrix $\mathcal{P}^{\mathcal{B}}_{\mathcal{B}'}$ is invertible. We have

$$[\mathbf{x}]_{\mathcal{B}} = \left(P^{\mathcal{B}}_{\mathcal{B}'}\right)^{-1} \cdot [\mathbf{x}]_{\mathcal{B}'}$$

Hence

$$P^{\mathcal{B}'}_{\mathcal{B}} = \left(P^{\mathcal{B}}_{\mathcal{B}'}\right)^{-1}$$

Example 2.11. If $\mathcal{B}$ is a basis of $\mathbb{R}^n$ and $\mathcal{E}$ is the standard basis of $\mathbb{R}^n$, then

$$[\mathbf{b}_i]_{\mathcal{E}} = \mathbf{b}_i.$$

Hence we simply have

$$P_{\mathcal{E}}^{\mathcal{B}} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$$

$P^{\mathcal{B}} := P_{\mathcal{E}}^{\mathcal{B}}$ is called the **change-of-coordinate matrix** from $\mathcal{B}$ to the standard basis of $\mathbb{R}^n$.

Properties 2.15. We have

- $P_{\mathcal{B}''}^{\mathcal{B}'} \cdot P_{\mathcal{B}'}^{\mathcal{B}} = P_{\mathcal{B}''}^{\mathcal{B}}$
- $P_{\mathcal{B}'}^{\mathcal{B}} = (P^{\mathcal{B}'})^{-1} \cdot P^{\mathcal{B}}$

Example 2.12. Let $\mathcal{E} = \{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\}$ be the standard basis of $\mathbb{R}^2$. Let

$$\begin{aligned} \mathcal{B} &= \{\mathbf{b}_1 = \begin{pmatrix}1\\1\end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix}1\\-1\end{pmatrix}\}, \\ \mathcal{B}' &= \{\mathbf{b}'_1 = \begin{pmatrix}1\\0\end{pmatrix}, \mathbf{b}'_2 = \begin{pmatrix}1\\1\end{pmatrix}\} \end{aligned}$$

be two other bases of $\mathbb{R}^2$. Then

$$\begin{aligned} P^{\mathcal{B}} &= \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} \\ P^{\mathcal{B}'} &= \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix} \\ P_{\mathcal{B}'}^{\mathcal{B}} &= (P^{\mathcal{B}'})^{-1} \cdot P^{\mathcal{B}} \\ &= \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}^{-1} \cdot \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}0 & 2\\1 & -1\end{pmatrix} \end{aligned}$$

One can check that this obeys the formula from Theorem 2.13.

$$\begin{aligned} \mathbf{b}_1 &= 0 \cdot \mathbf{b}'_1 + 1 \cdot \mathbf{b}'_2 \\ \mathbf{b}_2 &= 2 \cdot \mathbf{b}'_1 + (-1) \cdot \mathbf{b}'_2. \end{aligned}$$

CHAPTER 3

Euclidean Space

We define the geometric concepts of length, distance, angle and perpendicularity for $\mathbb{R}^n$. This gives $\mathbb{R}^n$ the structure of an *Euclidean Space*.

3.1 Inner Product

We write a point $\mathbf{u} \in \mathbb{R}^n$ as a column vector, i.e. $1 \times n$ matrix.

Definition 3.1. The **inner product** of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, i.e.

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

is given by

$$\mathbf{u} \cdot \mathbf{v} := \mathbf{u}^T \mathbf{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^{n} u_i v_i \in \mathbb{R}$$

Note. To avoid confusion, I will omit the dot for scalar multiplication: I use $c\mathbf{u}$ instead of $c \cdot \mathbf{u}$.

Some easily checked properties:

Theorem 3.2. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, c \in \mathbb{R}$

(1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

(3) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ iff $\mathbf{u} = \mathbf{0}$.

More generally:

Definition 3.3. Any vector space V over $\mathbb{R}$ equipped with an inner product $V \times V \longrightarrow \mathbb{R}$ satisfying Theorem 3.2 is called an **inner product space**. When $V = \mathbb{R}^n$ it is called an **Euclidean space**.

Example 3.1 (Optional)**.** An example of inner product space that is *infinite dimensional*: Let $C[a, b]$ be the vector space of real-valued continuous function defined on a closed interval $[a, b] \subset \mathbb{R}$. Then for $f, g \in C[a, b]$,

$$f \cdot g := \int_a^b f(t)g(t)dt$$

gives an inner product on $C[a, b]$.

Remark. All the Definitions and Theorems below applies to inner product spaces.

Remark. When $K = \mathbb{C}$, the inner product involves the complex conjugate

$$\mathbf{u} \cdot \mathbf{v} := \mathbf{u}^* \mathbf{v} := \begin{pmatrix} \overline{u_1} & \cdots & \overline{u_n} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n \overline{u_i} v_i \in \mathbb{C}$$

so that the last property (4) can hold. Also the third property have to be replaced by

(3*) $(\bar{c}\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

Properties 3.4. If $\mathbf{A} = (a_{ij})$ is an $m \times n$ matrix, then the matrix entries are given by

$$a_{ij} = \mathbf{e}'_i \cdot \mathbf{A}\mathbf{e}_j$$

where $\{\mathbf{e}_j\}$ is the standard basis for $\mathbb{R}^n$ and $\{\mathbf{e}'_i\}$ is the standard basis for $\mathbb{R}^m$.

Definition 3.5. The **norm** (or **length**) of $\mathbf{v}$ is the nonnegative scalar

$$\|\mathbf{v}\| := \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \in \mathbb{R}_{\geq 0}$$

For $c \in \mathbb{R}$, we have $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.

Definition 3.6. The vector $\mathbf{u}$ with unit length, i.e. $\|\mathbf{u}\| = 1$ is called a **unit vector**.

Given $\mathbf{v} \neq 0$, $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ has unit length and is called the **normalization** of $\mathbf{v}$

Example 3.2. $\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 2 \\ 0 \end{pmatrix} \in \mathbb{R}^4$ has norm $\|\mathbf{v}\| = \sqrt{1^2 + (-2)^2 + 2^2 + 0^2} = \sqrt{9} = 3$.

$$\frac{1}{3}\mathbf{v} = \begin{pmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{pmatrix}$$

is a unit vector.

Definition 3.7. The distance between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined by

$$dist(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\|$$

Theorem 3.8 (Law of cosine)**.** The angle θ between $\mathbf{u}$ and $\mathbf{v}$ can be calculated by

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

When $\theta = 90°$, we have

Definition 3.9. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** (or **perpendicular**) to each other if $\mathbf{u} \cdot \mathbf{v} = 0$.

Properties 3.10 (Pythagorean Theorem)**.** If $\mathbf{u} \cdot \mathbf{v} = 0$,

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Properties 3.11 (Cauchy-Schwarz Inequality)**.** For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Properties 3.12 (Triangle Inequality)**.** For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Example 3.3. $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ are orthogonal to each other in $\mathbb{R}^2$.

Example 3.4. $\mathbf{0}$ is orthogonal to every vector in $\mathbb{R}^n$.

3.2 Orthogonal Basis

Definition 3.13. Let $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\} \in \mathbb{R}^n$.

- S is called an **orthogonal set** if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$.
- If in addition S is a basis of $W \subset \mathbb{R}^n$, it is called an **orthogonal basis** for W.
- If in addition all vectors in S has unit norm, it is called an **orthonormal basis** for W.

Example 3.5. The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_n\}$ for $\mathbb{R}^n$ is an orthonormal basis.

Example 3.6. The set $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^2$.

Its rescaled version, the set $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^2$.

Theorem 3.14. Let $\mathcal{B} = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ be an orthogonal basis for a subspace $W \subset \mathbb{R}^n$. Then for $\mathbf{x} \in W$ we can solve for the coordinates with respect to $\mathcal{B}$ explicitly as

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

where

$$c_i = \frac{\mathbf{x} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \qquad i = 1, ..., p$$

3.3 Orthogonal Projection

Definition 3.15. Let $W \subset \mathbb{R}^n$ be a subspace. The **orthogonal complement** of W is the set

$$W^\perp := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \text{ for every } \mathbf{w} \in W\}$$

Properties 3.16. We have the following properties:

- $W^\perp$ is a subspace of $\mathbb{R}^n$.
- If $L = W^\perp$, then $W = L^\perp$.
- $\mathbf{x} \in W^\perp$ iff $\mathbf{x}$ is orthogonal to every vector in a spanning set ot W.

Theorem 3.17. Let $A \in M_{m\times n}(\mathbb{R})$. Then

$$(\mathrm{Row}A)^\perp = \mathrm{Nul}A, \qquad (\mathrm{Col}A)^\perp = \mathrm{Nul}A^T$$

Definition 3.18. The orthogonal projection of $\mathbf{b}$ onto $\mathbf{u}$ is given by

$$\mathrm{Proj}_{\mathbf{u}}(\mathbf{b}) := (\mathbf{b} \cdot \mathbf{e})\mathbf{e} = \frac{\mathbf{b} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

where $\mathbf{e} := \frac{\mathbf{u}}{\|\mathbf{u}\|}$ is the normalization of $\mathbf{u}$.

Theorem 3.19 (Orthogonal Decomposition Theorem)**.** Let $W \subset \mathbb{R}^n$ a subspace. Then each $\mathbf{x} \in \mathbb{R}^n$ can be written *uniquely* in the form

$$\mathbf{x} = \widehat{\mathbf{x}} + \mathbf{z}$$

where $\widehat{\mathbf{x}} \in W$ and $\mathbf{z} \in W^\perp$. Therefore we have

$$\mathbb{R}^n = W \oplus W^\perp$$

We sometimes write $\mathrm{Proj}_W(\mathbf{x}) := \widehat{\mathbf{x}}$. Note that $\mathrm{Proj}_W \in L(\mathbb{R}^n, \mathbb{R}^n)$ with

$$\mathrm{Im}(\mathrm{Proj}_W) = W, \qquad \mathrm{Ker}(\mathrm{Proj}_W) = W^\perp$$

Proof. Explicitly, if $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthogonal basis of W, then

$$\begin{aligned}\widehat{\mathbf{x}} &= \mathrm{Proj}_{\mathbf{u}_1}(\mathbf{x}) + \cdots + \mathrm{Proj}_{\mathbf{u}_p}(\mathbf{x}) \\ &= \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{x} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p\end{aligned}$$

and $\mathbf{z} = \mathbf{x} - \widehat{\mathbf{x}}$. □

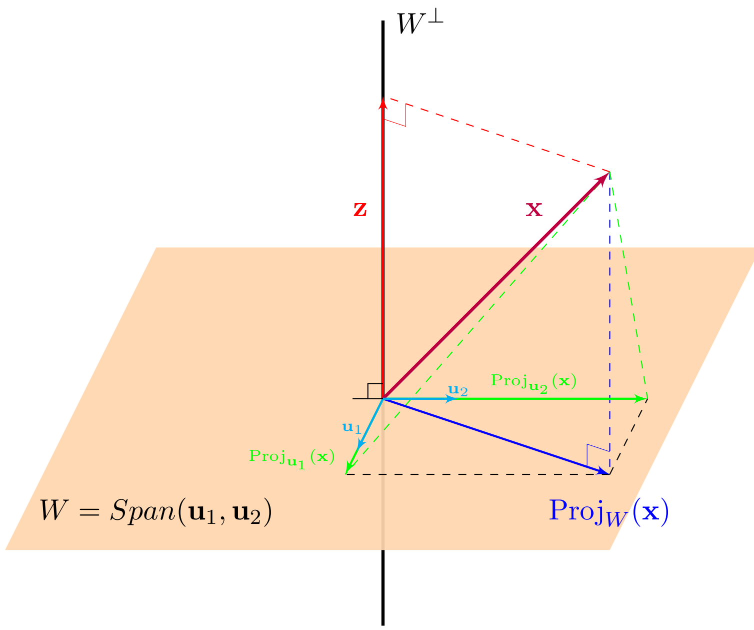


Figure 3.1: Orthogonal Projection

Remark. In particular, the uniqueness statement says that the orthogonal decomposition, i.e. the formula for $\widehat{\mathbf{x}}$, does not depend on the basis used for W in the proof.

Properties 3.20. If $\mathbf{x} \in W$, then $\mathrm{Proj}_W(\mathbf{x}) = \mathbf{x}$.

By using orthonormal basis, we can represent Proj_W as a matrix:

Theorem 3.21. If $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthonormal basis for $W \subset \mathbb{R}^n$, then

$$\mathrm{Proj}_W(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots (\mathbf{x} \cdot \mathbf{u}_p)\mathbf{u}_p$$

Equivalently, if $U = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{pmatrix}$ is an $n \times p$ matrix, then

$$\mathrm{Proj}_W(\mathbf{x}) = UU^T\mathbf{x}$$

The matrix $P := UU^T$ is an $n \times n$ matrix which is called an **orthogonal projection matrix**.

Definition 3.22. A **projection matrix** is an $n \times n$ matrix such that

$$P^2 = P$$

It is an **orthogonal projection** matrix if in addition

$$P^T = P$$

Example 3.7. If $W = Span(\mathbf{v}_1, \mathbf{v}_2)$ where $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for W: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

The normalization

$$\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}$$

is then an orthonormal basis for W. We have

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

and therefore

$$\mathrm{Proj}_W = UU^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{pmatrix}$$

Theorem 3.23 (Best Approximation Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$. Then

$$\|\mathbf{x} - \mathrm{Proj}_W \mathbf{x}\| \leq \|\mathbf{x} - \mathbf{v}\|, \qquad \text{for any } \mathbf{v} \in W$$

i.e. $\mathrm{Proj}_W \mathbf{x} \in W$ is the closest point in W to $\mathbf{x}$.

3.4 Orthogonal Matrix

Definition 3.24. A linear transformation $T \in L(V, W)$ between inner product spaces is called an **isometry** if it preserves the inner product:

$$(T\mathbf{u}) \cdot (T\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

for any vector $\mathbf{u}, \mathbf{v} \in V$.

Theorem 3.25. If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ is a linear isometry which is represented by an $m \times n$ matrix $\mathbf{U}$, then for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

- $\mathbf{U}$ has orthonormal columns
- $\mathbf{U}^T\mathbf{U} = \mathbf{Id}_{n\times n}$
- $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$ (i.e. it preserves length)
- $(\mathbf{U}\mathbf{x}) \cdot (\mathbf{U}\mathbf{y}) = 0$ iff $\mathbf{x} \cdot \mathbf{y} = 0$ (i.e. it preserves right angle)

Definition 3.26. If $n = m$, the square matrix $\mathbf{U}$ corresponding to a linear isometry is called an **orthogonal matrix**. It is invertible with

$$\mathbf{U}^{-1} = \mathbf{U}^T$$

The set of $n \times n$ orthogonal matrices is denoted by $O(n)$.

Properties 3.27. Orthogonal matrices satisfy the following *"group properties"*:

- $\mathbf{Id}_{n\times n} \in O(n)$.
- If $\mathbf{U} \in O(n)$, then $\mathbf{U}^{-1} \in O(n)$.
- If $\mathbf{U}, \mathbf{V} \in O(n)$, then $\mathbf{UV} \in O(n)$.

Example 3.8. In $\mathbb{R}^2$ and $\mathbb{R}^3$, an orthogonal matrix corresponds to combinations of rotations and mirror reflections.

In $\mathbb{R}^2$, all orthogonal matrix is of the form Rotations by angle θ counterclockwise

$$\mathbf{U} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Mirror reflections along the line with slope $\tan\frac{\theta}{2}$ passing through the origin

$$\mathbf{U} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

Example 3.9. The change-of-coordinate matrix $P_{\mathcal{B}'}^{\mathcal{B}}$ between orthonormal bases $\mathcal{B}$ and $\mathcal{B}'$ is an orthogonal matrix.

Non-Example 3.10. Projection Proj_W is in general not an orthogonal matrix: It does not preserve lengths.

3.5 Gram-Schmidt Process

Gram-Schmidt Process gives a simple algorithm to compute an orthogonal basis from an arbitrary basis.

Theorem 3.28 (Gram-Schmidt Process)**.** Let $\{\mathbf{x}_1, ..., \mathbf{x}_p\}$ be a basis for a subspace $W \subset \mathbb{R}^n$. Define

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{x}_1 \\
\mathbf{v}_2 &:= \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_2) \\
\mathbf{v}_3 &:= \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_3) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_3) \\
&\vdots \\
\mathbf{v}_p &:= \mathbf{x}_p - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_p) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_3) - \cdots - \mathrm{Proj}_{\mathbf{v}_{p-1}}(\mathbf{x}_p)
\end{aligned}$$

where

$$\mathrm{Proj}_{\mathbf{v}}(\mathbf{x}) := \frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$$

is the orthogonal projection (see Definition 3.18).

Then $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is an orthogonal basis for W. Furthermore,

$$Span(\mathbf{v}_1, ..., \mathbf{v}_k) = Span(\mathbf{x}_1, ..., \mathbf{x}_k), \qquad \text{for all } 1 \leq k \leq p$$

To obtain an orthonormal basis, just normalize the vectors:

$$\mathbf{v}_i \rightsquigarrow \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$$

The Gram-Schmidt Process implies the following factorization, which is very important in computation algorithms, best linear approximations, and eigenvalues decomposition.

Theorem 3.29 (QR Decomposition)**.** If $\mathbf{A}$ is an $m \times n$ matrix with linearly independent columns, then

$$\mathbf{A} = \mathbf{QR}$$

where

- $\mathbf{Q}$ is a $m \times n$ matrix with orthonormal columns forming a basis for Col$\mathbf{A}$,
- $\mathbf{R}$ is an $n \times n$ upper triangular invertible matrix with positive entries on the diagonal.

Example 3.11. Let

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

Let $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ be the columns of $\mathbf{A}$.

Part 1: Gram-Schmidt Process.

Apply the formula, we obtain

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{v}_1}\mathbf{x}_2 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix} - \frac{3}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \frac{1}{4}\begin{pmatrix}-3\\1\\1\\1\end{pmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_1}\mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_2}\mathbf{x}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{1}{6}\begin{pmatrix}-3\\1\\1\\1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}0\\-2\\1\\1\end{pmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for Col$\mathbf{A}$.

Part 2: Orthonormal basis and Q.

The vectors have lengths

$$\|\mathbf{v}_1\| = 2, \|\mathbf{v}_2\| = \frac{\sqrt{12}}{4}, \|\mathbf{v}_3\| = \frac{\sqrt{6}}{3},$$

hence the corresponding orthonormal basis is

$$\mathbf{v}_1' = \begin{pmatrix}1/2\\1/2\\1/2\\1/2\end{pmatrix}, \mathbf{v}_2' = \begin{pmatrix}-3/\sqrt{12}\\1/\sqrt{12}\\1/\sqrt{12}\\1/\sqrt{12}\end{pmatrix}, \mathbf{v}_3' = \begin{pmatrix}0\\-2/\sqrt{6}\\1/\sqrt{6}\\1/\sqrt{6}\end{pmatrix}$$

Then $\mathbf{Q}$ is formed by $\{\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'\}$, i.e.

$$\mathbf{Q} = \begin{pmatrix}1/2 & -3/\sqrt{12} & 0\\1/2 & 1/\sqrt{12} & -2/\sqrt{6}\\1/2 & 1/\sqrt{12} & 1/\sqrt{6}\\1/2 & 1/\sqrt{12} & 1/\sqrt{6}\end{pmatrix}$$

Part 3: The triangular matrix R.

To find $\mathbf{R}$, recall that $\mathbf{Q}$ is a linear isometry, hence

$$\mathbf{Q}^T\mathbf{A} = \mathbf{Q}^T(\mathbf{QR}) = \mathbf{Id}\cdot\mathbf{R} = \mathbf{R}$$

Therefore

$$\begin{aligned}
\mathbf{R} &= \mathbf{Q}^T\mathbf{A}\\
&= \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2\\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12}\\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0\\ 1 & 1 & 0\\ 1 & 1 & 1\\ 1 & 1 & 1 \end{pmatrix}\\
&= \begin{pmatrix} 2 & 3/2 & 1\\ 0 & 3/\sqrt{12} & 2/\sqrt{12}\\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}
\end{aligned}$$

Note that the diagonal of $\mathbf{R}$ is the same as the length of $\|\mathbf{v}_1\|, \|\mathbf{v}_2\|, \|\mathbf{v}_3\|$ used for the normalization in Part 2.

3.6 Least Square Approximation

If $\mathbf{A}$ is an $m \times n$ matrix with $m > n$, then the system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

is *overdetermined* and we may not have a solution.

If $\mathbf{A}$ has linear independent columns, we can use the QR decomposition to find a best approximation $\widehat{\mathbf{x}}$ to the solution of $\mathbf{A}\widehat{\mathbf{x}} = \mathbf{b}$ such that $\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{b}\|$ is the smallest.

By the Best Approximation Theorem (Theorem 3.23), the closest point $\mathbf{A}\mathbf{x} \in \mathrm{Col}\mathbf{A}$ to $\mathbf{b} \in \mathbb{R}^m$ should be $\mathrm{Proj}_{\mathrm{Col}\mathbf{A}}\mathbf{b}$. But $\mathrm{Col}\mathbf{A}$ has orthonormal basis given by columns of Q, so $\mathrm{Proj}_{\mathrm{Col}\mathbf{A}} = QQ^T$. Hence $\mathbf{A}\widehat{\mathbf{x}} = QQ^T\mathbf{b}$. Using $\mathbf{A} = QR$ we obtain:

Theorem 3.30 (Least Square Approximation)**.** If $\mathbf{A}$ is an $m \times n$ matrix with $m > n$ and has linear independent columns, such that we have the QR decomposition $\mathbf{A} = \mathbf{QR}$, then

$$\widehat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b} \in \mathbb{R}^n$$

is the vector such that for every $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{b}\| \leq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$$

Note. It is easier to solve for $\widehat{\mathbf{x}}$ by

$$R\widehat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$$

instead of finding $\mathbf{R}^{-1}$, because $\mathbf{R}$ is upper triangular, so that we can use backward substitution.

Example 3.12. Continue our example, if $\mathbf{b} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$, then the closest point $\widehat{\mathbf{x}}$ such that $\|\mathbf{Ax}-\mathbf{b}\|$ is smallest is given by

$$\mathbf{R}\widehat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$$

$$\begin{pmatrix}2 & 3/2 & 1\\ 0 & 3/\sqrt{12} & 2/\sqrt{12}\\ 0 & 0 & 2/\sqrt{6}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}1/2 & 1/2 & 1/2 & 1/2\\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12}\\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6}\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$$

$$= \begin{pmatrix}5\\ \sqrt{3}\\ 3/\sqrt{6}\end{pmatrix}$$

This is a very simple system of linear equations, and we can solve for $\widehat{\mathbf{x}}$ to get

$$\widehat{\mathbf{x}} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}2/3\\1\\3/2\end{pmatrix}$$

Therefore

$$\mathbf{A}\widehat{\mathbf{x}} = \begin{pmatrix}1 & 0 & 0\\1 & 1 & 0\\1 & 1 & 1\\1 & 1 & 1\end{pmatrix}\begin{pmatrix}2/3\\1\\3/2\end{pmatrix} = \begin{pmatrix}2/3\\5/3\\19/6\\19/6\end{pmatrix}$$

is the closest approximation in Col$\mathbf{A}$ to $\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$.

Note. It is called *"least square"* because $\|\mathbf{u}\|$ is computed by summing the squares of the coordinates. We want to find the smallest $\|\mathbf{Ax}-\mathbf{b}\|$. This is very useful in regression problems in statistics, where we want to fit the data onto a linear model as closely as possible.

CHAPTER 4

Eigenvectors and Eigenvalues

In this Chapter we will learn the important notion of eigenvectors and eigenvalues of matrices and linear transformation in general.

4.1 Eigenvectors

Definition 4.1. An **eigenvector** of a linear transformation $T \in L(V,V)$ is a *nonzero* vector $\mathbf{u} \in V$ such that $T\mathbf{u} = \lambda\mathbf{u}$ for some scalar $\lambda \in K$.

λ is called the **eigenvalue** of the eigenvector $\mathbf{u}$.

The space $V_\lambda := \{\mathbf{u} : T\mathbf{u} = \lambda\mathbf{u}\} \subset V$ is called the **eigenspace** of the eigenvalue λ.

We have the same definition if $T \in L(\mathbb{R}^n, \mathbb{R}^n)$ is represented by an $n \times n$ matrix $\mathbf{A}$.

Note. V_λ contains $\mathbf{0}$, although $\mathbf{0}$ is not an eigenvector by definition.

Properties 4.2. The eigenspace $V_\lambda = \mathrm{Nul}(\mathbf{A} - \lambda\mathbf{Id})$ is a vector space.

In particular, any linear combinations of eigenvectors with eigenvalue λ is again an eigenvector with eigenvalue λ if it is nonzero.

Properties 4.3. λ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\mathbf{A} - \lambda\mathbf{Id}) = 0$.

General strategy:

Step 1. Find the eigenvalues λ using determinant.

Step 2. For each eigenvalue λ, find the eigenvectors by solving the linear equations $(\mathbf{A}-\lambda\mathbf{Id})\mathbf{x} = \mathbf{0}$, which have non-trivial solutions.

Example 4.1. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$. To find eigenvalues,

$$\det \begin{pmatrix} 1-\lambda & 1 \\ 4 & 1-\lambda \end{pmatrix} = (1-\lambda)(1-\lambda) - 4 = \lambda^2 - 2\lambda - 3 = (\lambda-3)(\lambda+1) = 0,$$

hence $\lambda = 3$ or $\lambda = -1$.

For $\lambda = 3$, we have $\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \Longrightarrow \begin{pmatrix} t \\ 2t \end{pmatrix}$ are eigenvectors for all $t \in \mathbb{R}$.

For $\lambda = -1$, we have $\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \Longrightarrow \begin{pmatrix} t \\ -2t \end{pmatrix}$ are eigenvectors for all $t \in \mathbb{R}$.

Some useful theorems:

Theorem 4.4. The eigenvalues of a triangular matrix (in particular diagonal matrix) are the entries on its main diagonal.

Theorem 4.5 (Invertible Matrix Theorem)**.** $\mathbf{A}$ is invertible if and only if 0 is not an eigenvalue of $\mathbf{A}$.

Example 4.2. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$ has eigenvalues $\lambda = 1, 2, 3$.

Example 4.3. $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ has eigenvalues $\lambda = 1, 3$ only. V_3 is 1-dimensional spanned by $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, while V_1 is 2-dimensional spanned by $\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\}$.

Example 4.4. $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ has eigenvalues $\lambda = 1, 3$ only. V_3 is 1-dimensional spanned by $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, but V_1 is only 1-dimensional spanned by $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Example 4.5. Existence of eigenvalues depend on the field K. For example, $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has no eigenvalues in $\mathbb{R}$, but it has two complex eigenvalues $\lambda = i$ with eigenvectors $\begin{pmatrix} i \\ 1 \end{pmatrix}$, and $\lambda = -i$ with eigenvectors $\begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Theorem 4.6. If $\mathbf{v}_1, ..., \mathbf{v}_r$ are eigenvectors corresponding to *distinct* eigenvalues $\lambda_1, ..., \lambda_r$, then the set $\{\mathbf{v}_1, ..., \mathbf{v}_r\}$ is linearly independent.

4.2 Determinants

To find eigenvalues, we need to solve the **characteristic equation**

$$\det(\mathbf{A} - \lambda \mathbf{Id}) = 0$$

Let us recall the definition of Determinants of a matrix.

Definition 4.7. The **determinant** is the *unique* function

$$\det : M_{n\times n} \longrightarrow K$$

such that it satisfies the following properties:

- Determinant of identity matrix is 1: $\det(\mathbf{Id}) = 1$
- It is **skew-symmetric**: Interchange two rows gives a sign:

$$\det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_i & - \\ - & \mathbf{r}_j & - \\ & \vdots & \end{pmatrix} = -\det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_j & - \\ - & \mathbf{r}_i & - \\ & \vdots & \end{pmatrix}$$

  In particular, if the matrix has two same rows, $\det = 0$.
- It is **multilinear**, i.e. it is linear respective to rows:
  - Addition:

$$\det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}+\mathbf{s} & - \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ - & \mathbf{r} & - \\ & \vdots & \end{pmatrix} + \det \begin{pmatrix} & \vdots & \\ - & \mathbf{s} & - \\ & \vdots & \end{pmatrix}$$

– Scalar multiplication: for $k \in K$,

$$\det \begin{pmatrix} & \vdots & \\ - & k \cdot \mathbf{r} & - \\ & \vdots & \end{pmatrix} = k \det \begin{pmatrix} & \vdots & \\ - & \mathbf{r} & - \\ & \vdots & \end{pmatrix}$$

– zero:

$$\det \begin{pmatrix} & \vdots & \\ - & \mathbf{0} & - \\ & \vdots & \end{pmatrix} = 0$$

Remark. The definition of the determinant means that $\det \mathbf{A}$ is the *signed volume* of the "parallelepiped" spanned by the columns of $\mathbf{A}$ (and $|\det \mathbf{A}|$ is the volume).

From this definition, one can calculate the determinant of $\mathbf{A}$ using the *reduced row echelon form*, because it says that row operations do not change the determinant:

$$\det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_i + c\mathbf{r}_j & - \\ - & \mathbf{r}_j & - \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_i & - \\ - & \mathbf{r}_j & - \\ & \vdots & \end{pmatrix} + c \det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_j & - \\ - & \mathbf{r}_j & - \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ - & \mathbf{r}_i & - \\ - & \mathbf{r}_j & - \\ & \vdots & \end{pmatrix}$$

This also implies that

Properties 4.8. If $\mathbf{A}$ is triangular (in particular diagonal), then $\det \mathbf{A}$ is the product of the entries on the main diagonal of $\mathbf{A}$.

Example 4.6.

$$\begin{aligned} \det \begin{pmatrix} 0 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix} &\overset{R_1 \longleftrightarrow R_2}{=} -\det \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 3 \\ 0 & 3 & 2 \end{pmatrix} \\ &\overset{\frac{1}{2}R_2}{=} -2\det \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 3/2 \\ 0 & 3 & 2 \end{pmatrix} \\ &\overset{R_3 - 3R_2}{=} -2\det \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 3/2 \\ 0 & 0 & -5/2 \end{pmatrix} \\ &= -2 \cdot 1 \cdot 1 \cdot (-5/2) = 5. \end{aligned}$$

Properties 4.9. The properties of determinant under row operations means that det change as follows if you multiply your matrix by the **elementary matrices** on the left:

- $\mathbf{E} = \begin{array}{cc} \begin{array}{cccc} & i & j & \end{array} & \\ \begin{pmatrix} \ddots & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & \ddots \end{pmatrix} & \begin{array}{c} \\ i \\ j \\ \\ \end{array} \end{array}$: Interchanging two rows:

$$\det(\mathbf{EA}) = -\det(\mathbf{A})$$

- $\mathbf{E} = \begin{array}{cc} \begin{array}{ccc} & i & \end{array} & \\ \begin{pmatrix} \ddots & & \\ & k & \\ & & \ddots \end{pmatrix} & \begin{array}{c} \\ i \\ \\ \end{array} \end{array}$: Scalar multiplying i-th row by k:

$$\det(\mathbf{EA}) = k\det(\mathbf{A})$$

- $\mathbf{E} = \begin{array}{cc} \begin{array}{cccc} & i & j & \end{array} & \\ \begin{pmatrix} \ddots & & & \\ & 1 & c & \\ & 0 & 1 & \\ & & & \ddots \end{pmatrix} & \begin{array}{c} \\ i \\ j \\ \\ \end{array} \end{array}$: Adding multiples of j-th row to i-th row:

$$\det(\mathbf{EA}) = \det(\mathbf{A})$$

Here the $\ddots$ means it is 1 on the diagonal and 0 otherwise outside the part shown.

The determinant has the following very useful properties

Theorem 4.10. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices

- $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$
- $\det \mathbf{AB} = (\det \mathbf{A})(\det \mathbf{B})$
- $\det \mathbf{A}^T = \det \mathbf{A}$
- $\det k\mathbf{A} = k^n \det \mathbf{A}$

Corollary 4.11. If $\mathbf{A}$ is invertible, $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$.

If $\mathbf{Q}$ is orthogonal matrix, then $\det(\mathbf{Q}) = \pm 1$

Alternative way to compute determinant is

Theorem 4.12 (The Laplace Expansion Theorem)**.** If $\mathbf{A}$ is $n \times n$ matrix, $\det \mathbf{A}$ is computed by

$$\det \mathbf{A} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

or

$$\det \mathbf{A} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

where $C_{ij} = (-1)^{i+j} \det \mathbf{A}_{ij}$ and $\mathbf{A}_{ij}$ is the submatrix obtained by deleting the i-th row and j-th column.

Example 4.7. Using the same example above, using the first row,

$$\det \begin{pmatrix} 0 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix} = 0 \cdot \det \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = 0 - 2 \cdot 2 + 3 \cdot 3 = 5$$

or using the second column,

$$\det \begin{pmatrix} 0 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix} = -2 \cdot \det \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} + 0 \cdot \det \begin{pmatrix} 0 & 3 \\ 0 & 2 \end{pmatrix} - 3 \cdot \det \begin{pmatrix} 0 & 3 \\ 1 & 1 \end{pmatrix} = -2 \cdot 2 + 0 - 3 \cdot (-3) = 5$$

of course, the smart way is to choose the first column, because there are 2 zeros, making computation easier :)

4.3 Characteristic polynomial

Definition 4.13. If $\mathbf{A}$ is $n \times n$ matrix,

$$p(\lambda) := \det(\mathbf{A} - \lambda \mathbf{Id})$$

is a polynomial in λ of degree n, called the **characteristic polynomial**.

Therefore eigenvalues are the roots of the characteristic equations $p(\lambda) = 0$.

Definition 4.14. We define the notion of multiplicity:

- The dimension of the eigenspace V_λ is called the **geometric multiplicity**.
- The multiplicity of the root λ of $p(\lambda) = 0$ is called the **algebraic multiplicity**.

From the fundamental theorem of algebra, we note that since any polynomial of degree n has n roots (with repeats),

The algebraic multiplicities add up to n

Example 4.8. If $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, then the characteristic polynomial is

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{Id}) = \det \begin{pmatrix} 3-\lambda & 0 & 0 \\ 0 & 1-\lambda & 1 \\ 0 & 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2(3-\lambda)$$

Note that $\lambda = 1$ is a multiple root, hence the algebraic multiplicity of $\lambda = 1$ is 2.

On the other hand, the eigenspace V_1 is spanned by $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ only (see Exercise 4.4), so $\lambda = 1$ has geometric multiplicity $= 1$ only.

In other words:

Geometric multiplicity is in general not the same as algebraic multiplicity.

Theorem 4.15. The characteristic polynomial $p(\lambda)$ has the following properties:

- The top term is λ^n with coefficient $(-1)^n$
- The coefficient of λ^{n-1} is $(-1)^{n-1}\mathrm{Tr}\mathbf{A}$
- The constant term is $\det \mathbf{A}$.

Finally, we have the following interesting result, which can be used to calculate inverse of matrix!

Theorem 4.16 (Cayley-Hamilton Theorem)**.** If $p(\lambda)$ is the characteristic polynomial of $\mathbf{A}$, then

$$p(\mathbf{A}) = \mathbf{O}$$

where $\mathbf{O}$ is the zero matrix.

Example 4.9 (cont'd)**.** Since $(1-\lambda)^2(3-\lambda) = 3 - 7\lambda + 5\lambda^2 - \lambda^3$, the matrix $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ satisfies

$$3\mathbf{Id} - 7\mathbf{A} + 5\mathbf{A}^2 - \mathbf{A}^3 = \mathbf{O}$$

Therefore

$$3\mathbf{Id} = 7\mathbf{A} - 5\mathbf{A}^2 + \mathbf{A}^3$$

Multiplying both sides by $\mathbf{A}^{-1}$ we obtain

$$3\mathbf{A}^{-1} = 7\mathbf{Id} - 5\mathbf{A} + \mathbf{A}^2$$

This gives the inverse of $\mathbf{A}$ easily by usual matrix multiplication only.

4.4 Similarity

Definition 4.17. If $\mathbf{A}, \mathbf{B}$ are $n \times n$ matrices, then **$\mathbf{A}$ is similar to $\mathbf{B}$** if there is an invertible matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$$

Since if $\mathbf{Q} = \mathbf{P}^{-1}$, then also $\mathbf{B} = \mathbf{Q}\mathbf{A}\mathbf{Q}^{-1}$ and $\mathbf{B}$ is similar to $\mathbf{A}$. Therefore we can just say $\mathbf{A}$ and $\mathbf{B}$ are similar. We usually write

$$\mathbf{A} \sim \mathbf{B}$$

We have the following properties:

Theorem 4.18. If $\mathbf{A}$ and $\mathbf{B}$ are similar, then

- They have the same determinant

- They have the same characteristic polynomial
- They have the same eigenvalues
- They have the same algebraic and geometric multiplicities.

The polynomials of symmetric matrices are also related, which is very useful if the matrix is similar to a diagonal matrix (see next Chapter)

Theorem 4.19. If
$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$$
then for any integer n,
$$\mathbf{A}^n = \mathbf{P}\mathbf{B}^n\mathbf{P}^{-1}$$
In particular, for any polynomial $p(x)$,
$$p(\mathbf{A}) = \mathbf{P}p(\mathbf{B})\mathbf{P}^{-1}$$

Recall that for a linear transformation $T \in L(V,V)$, it can be represented as an $n \times n$ matrix $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ with respect to a basis $\mathcal{B}$ of V:
$$[T]_{\mathcal{B}} : [\mathbf{x}]_{\mathcal{B}} \mapsto [T\mathbf{x}]_{\mathcal{B}}$$
Also recall the change of basis matrix
$$P_{\mathcal{B}'}^{\mathcal{B}} : [\mathbf{x}]_{\mathcal{B}} \mapsto [\mathbf{x}]_{\mathcal{B}'}$$
Therefore we have the following interpretation of similar matrix:

Theorem 4.20. Let $T \in L(V,V)$ such that $[T]_{\mathcal{B}} = \mathbf{A}, [T]_{\mathcal{B}'} = \mathbf{B}$ and $\mathbf{P} = P_{\mathcal{B}'}^{\mathcal{B}}$.
Then $\mathbf{A}$ is similar to $\mathbf{B}$:
$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$$
In other words, **similar matrix represents the same linear transformation with respect to different basis!**

Pictorially, we have
$$\begin{tikzcd}
[\mathbf{x}]_{\mathcal{B}} \arrow[r, "\mathbf{A}"] \arrow[d, "\mathbf{P}"'] & [T\mathbf{x}]_{\mathcal{B}} \\
[\mathbf{x}]_{\mathcal{B}'} \arrow[r, "\mathbf{B}"] & [T\mathbf{x}]_{\mathcal{B}'} \arrow[u, "\mathbf{P}^{-1}"']
\end{tikzcd}$$

This means that for the linear transformation $T : V \longrightarrow V$, if we choose a "nice basis" $\mathcal{B}$ of B, the matrix $[T]_{\mathcal{B}}$ can be very nice! One choice of "nice basis" is given by **diagonalization**, which means under this basis, the linear transformation is represented by a diagonal matrix. We will study this in more detail in the next Chapter.

Example 4.10. In $\mathbb{R}^2$, let $\mathcal{E} = \{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\}$ be the standard basis, and $\mathcal{B} = \{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}-2\\1\end{pmatrix}\}$ be another basis. Let T be the linear transformation represented in the standard basis $\mathcal{E}$ by $\mathbf{A} = \begin{pmatrix}14/5 & 2/5\\2/5 & 11/5\end{pmatrix}$. Then we can diagonalize $\mathbf{A}$ as follows:

$$\begin{pmatrix}1 & -2\\2 & 1\end{pmatrix}^{-1}\begin{pmatrix}14/5 & 2/5\\2/5 & 11/5\end{pmatrix}\begin{pmatrix}1 & -2\\2 & 1\end{pmatrix} = \begin{pmatrix}2 & 0\\0 & 3\end{pmatrix}$$

See the picture below for the geometric meaning.

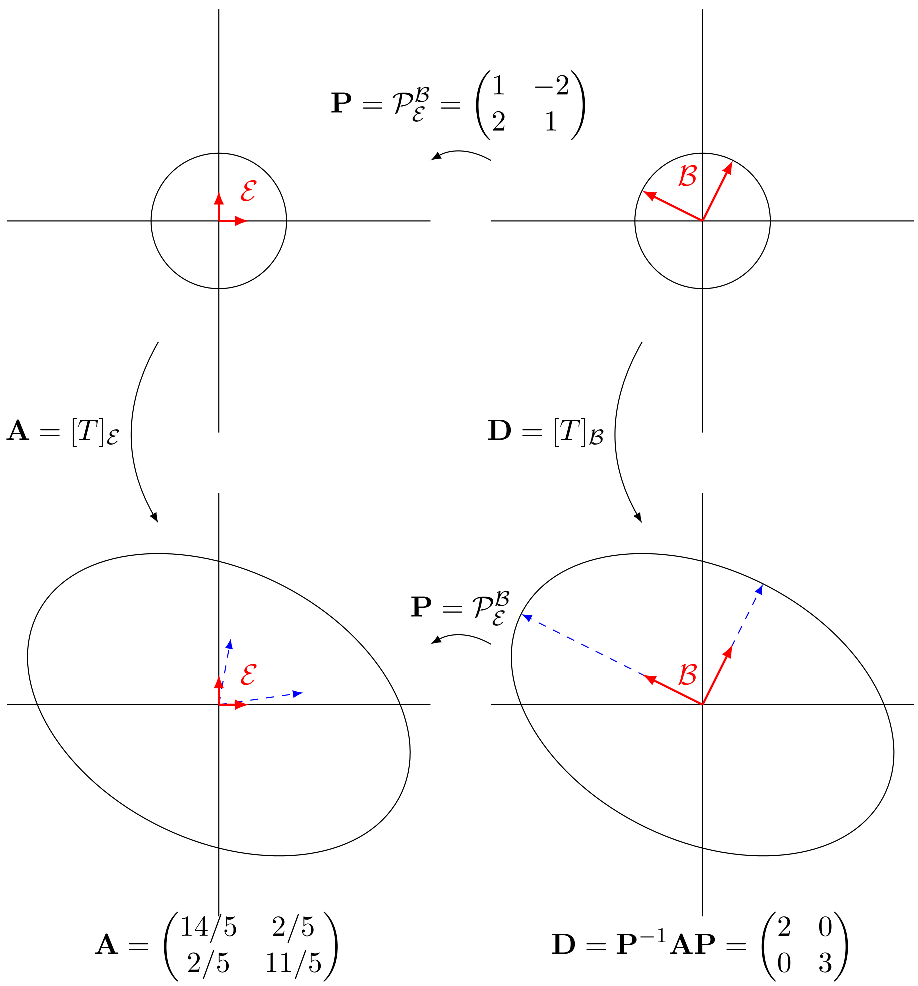


Figure 4.1: Pictorial explanation of similar matrix.

CHAPTER 5

Diagonalization

In this Chapter, we will learn how to diagonalize a matrix, when we can do it, and what else we can do if we fail to do it.

5.1 Diagonalization

Definition 5.1. A square $n \times n$ matrix $\mathbf{A}$ is **diagonalizable** if $\mathbf{A}$ is similar to a diagonal matrix, i.e.

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

for a diagonal matrix $\mathbf{D}$ and an invertible matrix $\mathbf{P}$.

Diagonalization let us simplify many matrix calculations and prove algebraic theorems. The most important application is the following. If $\mathbf{A}$ is diagonalizable, then it is easy to compute its powers:

Properties 5.2. If $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, then $\mathbf{A}^n = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}$.

Example 5.1. Let $\mathbf{A} = \begin{pmatrix} 4 & -3 \\ 2 & -1 \end{pmatrix}$. Then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where

$$\mathbf{P} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{P}^{-1} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix}$$

Then for example

$$\mathbf{D}^8 = \begin{pmatrix} 2^8 & 0 \\ 0 & 1^8 \end{pmatrix} = \begin{pmatrix} 256 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{A}^8 &= \mathbf{P}\mathbf{D}^8\mathbf{P}^{-1} \\ &= \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 256 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 766 & -765 \\ 510 & -509 \end{pmatrix} \end{aligned}$$

The Main Theorem of the Chapter is the following

Theorem 5.3 (The Diagonalization Theorem)**.** An $n \times n$ matrix $\mathbf{A}$ is diagonalizable

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

if and only if $\mathbf{A}$ has n linearly independent eigenvectors.
(Equivalently, $\mathbb{R}^n$ has a basis formed by eigenvectors of $\mathbf{A}$)

- The columns of $\mathbf{P}$ consists of eigenvectors of $\mathbf{A}$
- $\mathbf{D}$ is a diagonal matrix consists of the corresponding eigenvalues.

Proof. Since the columns of $\mathbf{P}$ is linearly independent, $\mathbf{P}$ is invertible. We have

$$\mathbf{A}\mathbf{P} = \mathbf{A} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \mathbf{P}\mathbf{D}$$

$\square$

Example 5.2. Let us diagonalize $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$

Step 1: Find Eigenvalues. Characteristic equation is

$$\mathbf{p}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{Id}) = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2) = 0$$

Hence the eigenvalues are $\lambda = 7$ and $\lambda = -2$.

Step 2: Find Eigenvectors. We find by usual procedure the linearly independent eigenvectors:

$$\lambda = 7 : \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \qquad \lambda = -2 : \mathbf{v}_3 = \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$$

Step 3: P constructed from eigenvectors. Putting them in columns,

$$\mathbf{P} = \begin{pmatrix} 1 & -1 & -2 \\ 0 & 2 & -1 \\ 1 & 0 & 2 \end{pmatrix}$$

Step 4: D consists of the eigenvalues. Putting the eigenvalues according to $\mathbf{v}_i$:

$$\mathbf{D} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

and we have

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

We have seen in last chapter (Theorem 4.6) that if the eigenvectors have different eigenvalues, then they are linearly independent. Therefore by the Diagonalization Theorem

Corollary 5.4. If $\mathbf{A}$ is an $n \times n$ matrix with n different eigenvalues, then it is diagonalizable.

Example 5.3. The matrix $\mathbf{A} = \begin{pmatrix} 3 & 4 & 5 \\ 0 & 0 & 7 \\ 0 & 0 & 6 \end{pmatrix}$ is triangular, hence the eigenvalues are the diagonal entries $\lambda = 3, \lambda = 0$ and $\lambda = 6$. Since they are all different, $\mathbf{A}$ is diagonalizable.

Non-Example 5.4. We have seen from Example 4.4 that the matrix $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ has 2 eigenvalues $\lambda = 1, 3$ only, so we cannot apply the Corollary. In fact, each of the eigenvalue has only 1-dimensional eigenvectors. Hence $\mathbb{R}^3$ does not have a basis formed by eigenvectors and so it is not diagonalizable by the Diagonalization Theorem.

From this Non-Example, we can also deduce that

Theorem 5.5. A square matrix $\mathbf{A}$ is diagonalizable if and only if for each eigenvalue λ, the algebraic multiplicity equals the geometric multiplicity.

5.2 Symmetric Matrices

A wide class of diagonalizable matrices are given by symmetric matrices, and the diagonalization has very nice properties.

Definition 5.6. A linear operator $T \in L(V, V)$ on an inner product space is called **symmetric** if

$$T\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot T\mathbf{v}$$

If T is represented by an $n \times n$ square matrix $\mathbf{A}$ on $V = \mathbb{R}^n$, then a matrix is called symmetric if

$$\mathbf{A}^T = \mathbf{A}$$

The first important property of symmetric matrix is the orthogonality between eigenspaces.

Theorem 5.7. If $\mathbf{A}$ is symmetric, then two eigenvectors from different eigenspaces are orthogonal.

Proof. If $\mathbf{v}_1 \in V_{\lambda_1}, \mathbf{v}_2 \in V_{\lambda_2}$ are eigenvectors with eigenvalues λ_1, λ_2 such that $\lambda_1 \neq \lambda_2$, then

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{A}\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{A}\mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$$

and so we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Therefore, if we normalize the eigenvectors, then the matrix $\mathbf{P}$ formed from the eigenvectors will consist of orthonormal columns, i.e. $\mathbf{P}$ is an orthogonal matrix.

Definition 5.8. A matrix $\mathbf{A}$ is **orthogonally diagonalizable** if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Theorem 5.9. An $n \times n$ matrix $\mathbf{A}$ is symmetric if and only if it is orthogonally diagonalizable.

In particular, $\mathbf{A}$ is diagonalizable means that each eigenvalue λ has the same algebraic and geometric multiplicity. That is, dimension of the eigenspace V_λ = the number of linearly independent eigenvectors with eigenvalue λ = multiplicity of the root λ of $p(\lambda) = 0$.

Example 5.5 (Exercise 5.2 cont'd)**.** We have diagonalize the matrix $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$ before. But the matrix $\mathbf{P}$ we found is not an orthogonal matrix.

We have found before (**Step 1, Step 2.**)

$$\lambda = 7 : \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \qquad \lambda = -2 : \mathbf{v}_3 = \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$$

Since $\mathbf{A}$ is symmetric, different eigenspaces are orthogonal to each other. So for example we see that

$$\mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_2 \cdot \mathbf{v}_3 = 0$$

So we just need to find an orthogonal basis for the eigenspace V_7.

Step 2a: Use the Gram-Schmidt process on V_7:

$$\mathbf{b}_1 = \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{b}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{b}_1 \cdot \mathbf{v}_2}{\mathbf{b}_1 \cdot \mathbf{b}_1} \right) \mathbf{b}_1 = \frac{1}{2} \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$$

Therefore $\{\mathbf{b}_1, \mathbf{b}_2\}$ is an orthogonal basis for V_7, and $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{v}_3\}$ is an orthogonal eigenvector basis for $\mathbb{R}^3$.

Step 2b: Normalize.

$$\mathbf{b}_1' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{b}_2' = \frac{1}{\sqrt{18}} \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3' = \frac{1}{3} \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$$

Step 3, Step 4: Construct P and D

Putting together the eigenvectors, we have

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \end{pmatrix}$$

and $\mathbf{D} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$, consisting of the eigenvalues, is the same as before.

Theorem 5.10. If $\mathbf{A}$ is a symmetric $n \times n$ matrix, then it has n real eigenvalues (counted with multiplicity) i.e. the characteristic polynomial $p(\lambda)$ has n real roots (counted with repeated roots).

The collection of Theorems 5.7, 5.9, and 5.10 in this Section are known as the **Spectral Theorem for Symmetric Matrices**.

5.3 Minimal Polynomials

By the Cayley-Hamilton Theorem, if $p(\lambda)$ is the characteristic polynomial of a square matrix $\mathbf{A}$, then

$$p(\mathbf{A}) = \mathbf{O}$$

Although this polynomial tells us about the eigenvalues (and their multiplicities), it is sometimes too "big" to tell us information about the structure of the matrix.

Definition 5.11. The **minimal polynomial** $m(\lambda)$ is the *unique* polynomial such that

$$m(\mathbf{A}) = \mathbf{O}$$

with leading coefficient 1, and has the smallest degree among such polynomials.

To see it is unique: If we have different minimal polynomials m, m', then $m(\mathbf{A}) - m'(\mathbf{A}) = \mathbf{O}$, but since m, m' have the same degree with the same leading coefficient, $m - m'$ is a polynomial with smaller degree, contradicting the fact that m has smallest degree.

Since it has the smallest degree, in particular we have

$$\deg(m) \leq \deg(p) = n$$

Example 5.6. The diagonal matrix $\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ has characteristic polynomial

$$p(\lambda) = (2 - \lambda)^3$$

but obviously $\mathbf{A} - 2\mathbf{Id} = \mathbf{O}$, hence the minimal polynomial of $\mathbf{A}$ is just

$$m(\lambda) = \lambda - 2$$

In particular,

The minimal polynomial $m(\lambda)$ of $\mathbf{A}$ has degree 1 if and only if $\mathbf{A}$ is a multiple of $\mathbf{Id}$.

Example 5.7. The diagonal matrix $\mathbf{A}=\begin{pmatrix}1&0&0\\0&2&0\\0&0&2\end{pmatrix}$ has characteristic polynomial

$$p(\lambda)=(1-\lambda)(2-\lambda)^2$$

Since $\mathbf{A}$ is not a multiple of $\mathbf{Id}$, $m(\lambda)$ has degree at least 2. Since $(\mathbf{A}-\mathbf{Id})(\mathbf{A}-2\mathbf{Id})=\mathbf{O}$, the polynomial

$$m(\lambda)=(\lambda-1)(\lambda-2)$$

having degree 2 is the minimal polynomial.

Example 5.8. The matrix $\mathbf{A}=\begin{pmatrix}1&1&0\\0&1&1\\0&0&1\end{pmatrix}$ has characteristic polynomial

$$p(\lambda)=(1-\lambda)^3$$

and it turns out that the minimal polynomial is the same also (up to a sign):

$$m(\lambda)=(\lambda-1)^3$$

From the above examples, we also observe that

Theorem 5.12. $p(\lambda)=m(\lambda)q(\lambda)$ for some polynomial $q(\lambda)$. That is $m(\lambda)$ **divides** $p(\lambda)$.

Proof. We can do a polynomial division

$$p(\lambda)=m(\lambda)q(\lambda)+r(\lambda)$$

where $r(\lambda)$ is the remainder with $\deg(r)<\deg(m)$. Since $p(\mathbf{A})=\mathbf{O}$ and $m(\mathbf{A})=\mathbf{O}$, we must have $r(\mathbf{A})=\mathbf{O}$. But since $\deg(r)<\deg(m)$ and m is minimal, r must be the zero polynomial. $\square$

Theorem 5.13. Let $\lambda_1,...,\lambda_k$ be the eigenvalues of $\mathbf{A}$ (i.e. roots of $p(\lambda)$) Then

$$m(\lambda)=(\lambda-\lambda_1)^{s_1}\cdots(\lambda-\lambda_k)^{s_k}$$

where $1\le s_i\le m_i$ where m_i is the algebraic multiplicity of λ_i.

Proof. To see $s_i \geq 1$, note that if $\mathbf{v}_i$ is an eigenvector for the eigenvalue λ_i, then since $m(\mathbf{A}) = \mathbf{O}$,

$$\mathbf{0} = m(\mathbf{A})\mathbf{v}_i = m(\lambda_i)\mathbf{v}_i$$

But since $\mathbf{v}_i \neq 0$, we have $m(\lambda_i) = 0$, so λ_i is a root of $m(\lambda)$. $\square$

Finally, the most useful criterion is the following result:

Theorem 5.14. An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if each $s_i = 1$. That is, $m(\lambda)$ only has linear factors.

Using this result, minimal polynomials can let us determine whether a matrix is diagonalizable or not without even calculating the eigenspaces!

Example 5.9. The matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ -4 & 3 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = (\lambda - 1)^2$. Since $m(\lambda) \neq \lambda - 1$ because $\mathbf{A} \neq \mathbf{Id}$, we must have $m(\lambda) = (\lambda - 1)^2$, hence $\mathbf{A}$ is not diagonalizable.

Example 5.10. The matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = \lambda(\lambda - 1)^2$, hence it has eigenvalues $\lambda = 1$ and $\lambda = 0$. The minimal polynomial can only be $\lambda(\lambda-1)$ or $\lambda(\lambda-1)^2$. Since

$$\mathbf{A}(\mathbf{A} - \mathbf{Id}) \neq \mathbf{O}$$

the minimal polynomial must be $m(\lambda) = \lambda(\lambda - 1)^2$, hence $\mathbf{A}$ is not diagonalizable.

Example 5.11. The matrix $\mathbf{A} = \begin{pmatrix} 2 & -2 & 2 \\ 0 & -2 & 4 \\ 0 & -2 & 4 \end{pmatrix}$ has characteristic polynomial $p(\lambda) = \lambda(\lambda - 2)^2$, hence it has eigenvalues $\lambda = 2$ and $\lambda = 0$. The minimal polynomial can only be $\lambda(\lambda-2)$ or $\lambda(\lambda-2)^2$. Since

$$\mathbf{A}(\mathbf{A} - 2\mathbf{Id}) = \mathbf{O}$$

the minimal polynomial is $m(\lambda) = \lambda(\lambda - 2)$, hence $\mathbf{A}$ is diagonalizable.

5.4 Jordan Canonical Form

Finally we arrive at the most powerful tool in Linear Algebra, called the Jordan Canonical Form. This completely determines the structure of a given matrix. It is also the best approximation to diagonalization if the matrix is not diagonalizable.

The result below works **as long as $p(\lambda)$ has n roots (counted with multiplicity)**, and so it is always available if the field is $K = \mathbb{C}$, so that the characteristic polynomial $p(\lambda)$ always has n roots by the Fundamental Theorem of Algebra.

Definition 5.15. Let $\mathbf{A} = \mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_m$ denote the $n \times n$ matrix in *block form*:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & \mathbf{A}_m \end{pmatrix}$$

such that $\mathbf{A}_i$ are square matrices of size $d_i \times d_i$, and $\mathbf{O}$ are zero matrices of the appropriate sizes. In particular $n = d_1 + d_2 + \cdots + d_m$.

For any $d \geq 1$, and $\lambda \in \mathbb{C}$, let $\mathbf{J}_\lambda^{(d)}$ be the **Jordan block** denote the $d \times d$ matrix

$$\mathbf{J}_\lambda^{(d)} = \begin{pmatrix} \lambda & 1 & & & & & \\ & \lambda & 1 & & & & \\ & & \lambda & 1 & & & \\ & & & \ddots & \ddots & & \\ & & & & \lambda & 1 & \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{pmatrix}$$

where all the unmarked entries are 0.

Note. When $d = 1$, we have $\mathbf{J}_\lambda^{(1)} = \big(\lambda\big)$.

With these notations, we can now state the Main Big Theorem

Theorem 5.16 (Jordan Canonical Form)**.** Let $\mathbf{A} \in M_{n\times n}(\mathbb{C})$. Then $\mathbf{A}$ is similar to

$$\mathbf{J} := \mathbf{J}_{\lambda_1}^{(d_1)} \oplus \cdots \mathbf{J}_{\lambda_m}^{(d_m)}$$

where λ_i belongs to the eigenvalues of $\mathbf{A}$. (λ_i **with different index may be the same!**). This decomposition is *unique* up to permuting the order of the Jordan blocks.

Since eigenvalues, characteristic polynomials, minimal polynomials, and multiplicity etc. are all the same for similar matrices, if **we can determine the Jordan block from these data, we can determine the Jordan Canonical Form of a matrix A.**

Let us first consider a single block.

Properties 5.17. The Jordan block $\mathbf{J}_\lambda^{(d)}$ has

- only one eigenvalue λ
- characteristic polynomial $(t-\lambda)^d$
- minimal polynomial $(t-\lambda)^d$
- geometric multiplicity of λ is 1.

Now let us combine several blocks of the same eigenvalues:

Properties 5.18. The matrix $\mathbf{J}_\lambda^{(d_1)} \oplus \cdots \oplus \mathbf{J}_\lambda^{(d_k)}$ has

- only one eigenvalue λ
- characteristic polynomial $(t-\lambda)^{d_1+\cdots d_k}$
- minimal polynomial $(t-\lambda)^{\max(d_1,\ldots,d_k)}$
- geometric multiplicity of λ is k.

Now we can do the same analysis by combining different Jordan blocks. We arrive at the following structure:

Theorem 5.19. Given a matrix $\mathbf{A}$ in the Jordan canonical form:

- The eigenvalues $\lambda_1, ..., \lambda_k$ are the entries of the diagonal.
- The characteristic polynomial is
$$p(\lambda) = (\lambda-\lambda_1)^{r_1}\cdots(\lambda-\lambda_k)^{r_k}$$
where r_i is the number of occurrences of λ_i on the diagonal.
- The minimal polynomial is
$$m(\lambda) = (\lambda-\lambda_1)^{s_1}\cdots(\lambda-\lambda_k)^{s_k}$$
where s_i is the size of the largest λ_i-block in $\mathbf{A}$
- The geometric multiplicity of λ_i is the number of λ_i-blocks in $\mathbf{A}$.

Example 5.12. Assume $\mathbf{A}$ is a 6×6 matrix with characteristic polynomial

$$p(\lambda) = (\lambda - 2)^4 (\lambda - 3)^2$$

and minimal polynomial

$$m(\lambda) = (\lambda - 2)^2 (\lambda - 3)^2,$$

with eigenspaces $\dim V_2 = 3, \dim V_3 = 1$. Then it must have 3 blocks of $\lambda = 2$, with maximum block-size of 2 so that the $\lambda = 2$ blocks add up to 4 rows. It also has 1 block of $\lambda = 3$ with block-size 2. Hence

$$\mathbf{A} \sim \mathbf{J}_2^{(2)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_3^{(2)}$$

$$\mathbf{A} \sim \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

The uniqueness of Jordan Canonical Form says that $\mathbf{A}$ is also similar to the matrix where the Jordan blocks are in different order. For example we can have:

$$\mathbf{A} \sim \begin{pmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Example 5.13. Another example, let $\mathbf{A}$ be a matrix such that it has characteristic polynomial

$$p(\lambda) = \lambda^4 (\lambda - 1)^3 (\lambda - 2)^3$$

and minimal polynomial

$$m(\lambda) = \lambda^3 (\lambda - 1)^2 (\lambda - 2)$$

With this information only, we can determine

$$\mathbf{A} \sim \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(1)} \oplus \mathbf{J}_1^{(2)} \oplus \mathbf{J}_1^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)} \oplus \mathbf{J}_2^{(1)}$$

$$\mathbf{A} \sim \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

It turns out that when the matrix is bigger than 6×6, sometimes we **cannot determine** the Jordan Canonical Form just by knowing $p(\lambda), m(\lambda)$ and the dimension of the eigenspaces only:

Example 5.14. Consider a 7×7 matrix $\mathbf{A}$. Let $p(\lambda) = \lambda^7$, $m(\lambda) = \lambda^3$, and $\dim V_0 = 3$. Then $\mathbf{A}$ has 3 blocks and the largest block has size 3. So it may be similar to

$$\mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(1)} \qquad \text{or} \qquad \mathbf{J}_0^{(3)} \oplus \mathbf{J}_0^{(2)} \oplus \mathbf{J}_0^{(2)}$$

However, by the uniqueness of Jordan Canonical Form, we know that these two are not similar to each other, but we cannot tell which one is similar to $\mathbf{A}$ just from the given information.

To determine which one is the Jordan Canonical Form of $\mathbf{A}$, we need more techniques. In the Homework, we will discuss how one can determine exactly the size of the Jordan blocks, as well as the transformation matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{PJP}^{-1}$.

5.5 Positive definite matrix (Optional)

One application of the diagonalization of symmetric matrix allows us to analyses quadratic functions, and define "square root" and "absolute value" of a matrix, which is useful in advanced linear algebra and optimization problems.

Definition 5.20. Let $\mathbf{A}$ be a symmetric matrix. The quadratic function

$$\mathbf{Q}(\mathbf{x}) := \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x} \cdot \mathbf{A}\mathbf{x}$$

is called the **quadratic form** associated to $\mathbf{A}$.

Definition 5.21. A quadratic form $\mathbf{Q}$ (or symmetric matrix $\mathbf{A}$) is called **positive definite** if

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A}\mathbf{x} > 0, \qquad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n$$

It is called **positive semidefinite** if

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A}\mathbf{x} \geq 0, \qquad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n$$

Example 5.15. $\mathbf{Q}(\mathbf{x}) := \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 9x^2 + y^2$ is positive definite.

Example 5.16. Let $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$. Then $\mathbf{Q}(\mathbf{x}) := \mathbf{x}^T\mathbf{A}\mathbf{x} = 5x^2 + 8xy + 5y^2$ is positive definite. We can see that it represents ellipses as follows: We can diagonalize the matrix by $\mathbf{A} = \mathbf{PDP}^T$ where

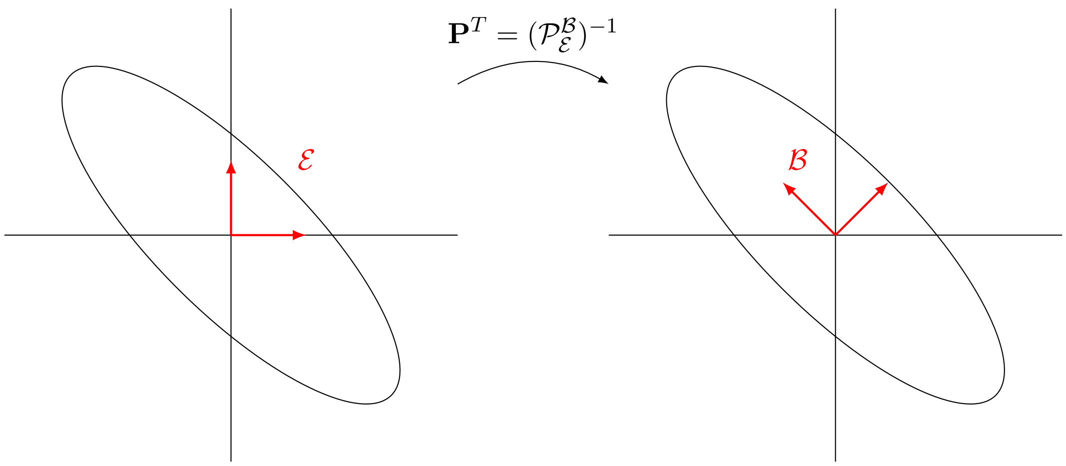
$\mathbf{D} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{P} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Then

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{D} \mathbf{P}^T \mathbf{x} = (\mathbf{P}^T \mathbf{x})^T \mathbf{D} (\mathbf{P}^T \mathbf{x})$$

Therefore if we let $\widehat{\mathbf{x}} = \begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} = \mathbf{P}^T \mathbf{x}$, i.e. rotating the basis by $\mathbf{P}^{-1}$, then

$$\mathbf{Q}(\widehat{\mathbf{x}}) = 9\widehat{x}^2 + \widehat{y}^2$$

and it is represented by an ellipse.



$\mathbf{Q}(\mathbf{x}) = 5x^2 + 8xy + 5y^2 = 1$ $\qquad$ $\mathbf{Q}(\widehat{\mathbf{x}}) = 9\widehat{x}^2 + \widehat{y}^2 = 1$

Figure 5.1: Pictorial explanation of similar matrix.

Theorem 5.22. A quadratic form $\mathbf{Q}$ associated to a symmetric matrix $\mathbf{A}$ is positive (semi)definite if and only if $\lambda_i > 0$ ($\lambda_i \geq 0$) for all the eigenvalues of $\mathbf{A}$.

Proof. Substitute $\mathbf{x} \rightsquigarrow \mathbf{P}\mathbf{x}$ where $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ is the diagonalization. $\square$

Remark. If all eigenvalues are $\lambda_i < 0$ ($\lambda_i \leq 0$), we call the quadratic form **negative (semi)definite**. Otherwise if some are positive and some are negative, it is called **indefinite**.

We can always find a "square root" of $\mathbf{A}$ if it is positive (semi)definite.

Theorem 5.23. If $\mathbf{A}$ is positive (semi)definite, then there exists exactly one positive (semi)definite matrix $\mathbf{B}$ such that

$$\mathbf{B}^2 = \mathbf{A}$$

We call $\mathbf{B}$ the **square root** of $\mathbf{A}$ and denote it by $\sqrt{\mathbf{A}}$. It is given by

$$\sqrt{\mathbf{A}} = \mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}^T$$

where $\mathbf{D}^{\frac{1}{2}}$ is the diagonal matrix where we take the square root of the entries of $\mathbf{D}$.

We can also construct the "absolute value" of *any* matrix $\mathbf{A}$:

Theorem 5.24. Let $\mathbf{A}$ be any $m \times n$ matrix. Then $\mathbf{A}^T\mathbf{A}$ is a positive semidefinite matrix, and

$$|\mathbf{A}| := \sqrt{\mathbf{A}^T\mathbf{A}}$$

is called the **absolute value** of $\mathbf{A}$.

Proof. $\mathbf{A}^T\mathbf{A}$ is symmetric, and $\mathbf{x} \cdot \mathbf{A}^T\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. □

This is used in the construction of Singular Value Decomposition in the next section.

Example 5.17. Let $\mathbf{A} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{P}^T$ where $\mathbf{P} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Then

$$\sqrt{\mathbf{A}} = \mathbf{P}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{P}^T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Let $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ -2/5 & -11/5 \end{pmatrix}$. Then $\mathbf{B}^T\mathbf{B} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$, therefore by above $|\mathbf{B}| = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

5.6 Singular Value Decomposition (Optional)

We know that not all matrix can be diagonalized. One solution to this is to use Jordan Canonical Form, which give us an approximation. Another approach is the **Singular Value Decomposition**, and this can even be applied to rectangular matrix! This method is also extremely important in data analysis.

Recall that if $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ is the eigenvector, the effect is "stretching by λ" along the direction of $\mathbf{v}$. We want to consider all such directions if possible, even for rectangular matrix.

Definition 5.25. Let $\mathbf{A}$ be $m \times n$ matrix. The **singular values** of $\mathbf{A}$ is the eigenvalues σ_i of $|\mathbf{A}| = \sqrt{\mathbf{A}^T\mathbf{A}}$.

If $\mathbf{A}$ has rank r, then we have r nonzero singular values. We arrange them as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Since $\mathbf{A}^T\mathbf{A}$ is a positive definite symmetric matrix, it has an orthonormal set of eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ with positive eigenvalues $\{\lambda_1, ..., \lambda_n\}$. Then

$$\|\mathbf{A}\mathbf{v}_i\|^2 = \mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mathbf{v}_i^T\lambda_i\mathbf{v}_i = \lambda_i$$

therefore the singular values $\sigma_i = \sqrt{\lambda_i} = \|\mathbf{A}\mathbf{v}_i\|$ of $\mathbf{A}$ is precisely the length of the vector $\mathbf{A}\mathbf{v}_i$.

Let us denote a "quasi-diagonal" matrix of size $m \times n$ and rank $r \leq m, n$:

$$\boldsymbol{\Sigma} = \begin{array}{cc} \begin{array}{cc} r & n-r \end{array} & \\ \begin{pmatrix} \mathbf{D} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} & \begin{array}{c} r \\ m-r \end{array} \end{array}$$

where $\mathbf{D}$ is a diagonal matrix. (When $r = m$ or n, we omit the rows or columns of zeros).

Theorem 5.26 (Singular value decomposition)**.** Let $\mathbf{A}$ be an $m \times n$ matrix with rank r. Then we have the factorization

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

where

- $\boldsymbol{\Sigma}$ is as above with $\mathbf{D}$ consists of the first r singular values of $\mathbf{A}$
- $\mathbf{U}$ is an $m \times m$ orthogonal matrix
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix

$\mathbf{V} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}$ where the columns are the orthonormal eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of $\mathbf{A}^T\mathbf{A}$.

For $\mathbf{U}$, extend the orthogonal set $\{\mathbf{A}\mathbf{v}_1, ..., \mathbf{A}\mathbf{v}_r\}$ to a basis of $\mathbb{R}^m$, and normalize to obtain an orthonormal basis $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$. Then $\mathbf{U} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix}$.

Example 5.18. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$. Then $\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and it has eigenvalues $\lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 0$ with orthonormal eigenvectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

Therefore

$$\mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Also $\sigma_1 = \sqrt{\lambda_1} = 2, \sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$. Therefore

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Finally

$$\mathbf{u}_1 = \frac{\mathbf{A}\mathbf{v}_1}{\|\mathbf{A}\mathbf{v}_1\|} = \frac{\mathbf{A}\mathbf{v}_1}{\sigma_1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$\mathbf{u}_2 = \frac{\mathbf{A}\mathbf{v}_2}{\|\mathbf{A}\mathbf{v}_2\|} = \frac{\mathbf{A}\mathbf{v}_2}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

therefore

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

is the Singular Value Decomposition of $\mathbf{A}$.

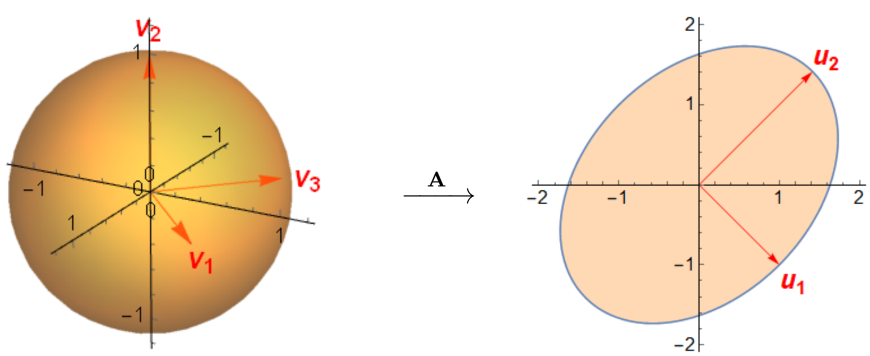


Figure 5.2: Multiplication by $\mathbf{A}$. It squashed the $\mathbf{v}_3$ direction to zero.

One useful application of SVD is to **find the bases of the fundamental subspaces**.

Theorem 5.27. Let $\mathbf{A}$ be $m \times n$ matrix with rank r, and $\mathbf{A} = \mathbf{U\Sigma V}^T$ be the SVD.

Assume $\mathbf{U} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix}$ and $\mathbf{V} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}$. Then

- $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal basis of $\mathrm{Col}\mathbf{A}$.
- $\{\mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$ is an orthonormal basis of $\mathrm{Nul}(\mathbf{A}^T)$.
- $\{\mathbf{v}_1, ..., \mathbf{v}_r\}$ is an orthonormal basis of $\mathrm{Row}\mathbf{A} = \mathrm{Col}\mathbf{A}^T$.
- $\{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$ is an orthonormal basis of $\mathrm{Nul}\mathbf{A}$.

Another application is the **least-square solution** which works like the example from QR decomposition in Chapter 3.

Definition 5.28. Let $\mathbf{U}_r = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r \\ | & & | \end{pmatrix}, \mathbf{V}_r = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_r \\ | & & | \end{pmatrix}$ be the submatrix consists of the first r columns. Then

$$\mathbf{A} = \begin{pmatrix} \mathbf{U}_r & * \end{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{V}_r^T \\ * \end{pmatrix} = \mathbf{U}_r \mathbf{D} \mathbf{V}_r^T.$$

The **pseudoinverse** of $\mathbf{A}$ is defined to be

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{D}^{-1} \mathbf{U}_r^T$$

The pseudoinverse satisfies for example

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$$

and

$$\mathbf{A}\mathbf{A}^+ = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}$$

because

$$\mathbf{A}\mathbf{A}^+ = (\mathbf{U}_r \mathbf{D} \mathbf{V}_r^T)(\mathbf{V}_r \mathbf{D}^{-1} \mathbf{U}_r^T) = \mathbf{U}_r \mathbf{D}\mathbf{D}^{-1} \mathbf{U}_r^T = \mathbf{U}_r \mathbf{U}_r^T = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}$$

Theorem 5.29. Given the equation $\mathbf{x} = \mathbf{A}\mathbf{b}$, the least-square solution is given by

$$\widehat{\mathbf{x}} = \mathbf{A}^{+}\mathbf{b} = \mathbf{V}_r\mathbf{D}^{-1}\mathbf{U}_r^T\mathbf{b}$$

Proof. Since $\mathbf{A}\widehat{\mathbf{x}} = \mathbf{A}\mathbf{A}^{+}\mathbf{b} = \mathrm{Proj}_{\mathrm{Col}\mathbf{A}}\mathbf{b}$, $\mathbf{A}\widehat{\mathbf{x}}$ is the closest point to $\mathbf{b}$ in Col$\mathbf{A}$. $\square$

APPENDIX A

Complex Matrix

If the field $K = \mathbb{C}$, we have many similar results corresponding to the real matrix case.

Definition A.1. If $c = x + iy \in \mathbb{C}$ is a complex number, we denote by $\bar{c} = x - iy \in \mathbb{C}$ the **complex conjugate** of c.

Instead of **transpose**, we have **adjoint**, which is conjugate transpose:

Definition A.2. If $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$ is an $m \times n$ complex matrix, then the **adjoint** is an $n \times m$ complex matrix given by

$$\mathbf{A}^* = \begin{pmatrix} \overline{a_{11}} & \cdots & \overline{a_{m1}} \\ \overline{a_{12}} & & \overline{a_{m2}} \\ \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{mn}} \end{pmatrix}$$

That is, we take the transpose, and then conjugate every element:

$$\left(a_{ij}\right)^* = \left(\overline{a_{ji}}\right)$$

The definition of an **inner product** is changed slightly, as already noted in Chapter 3.

Definition A.3. The inner product of $\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{C}^n$ is given by

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^* \mathbf{v} = \begin{pmatrix} \overline{u_1} & \cdots & \overline{u_n} \end{pmatrix} \begin{pmatrix} v_1 \\ \ddots \\ v_n \end{pmatrix} = \sum_{i=1}^{n} \overline{u_i} v_i \in \mathbb{C}$$

The important properties of inner product is property (4) from Theorem 3.2:

$$\mathbf{u} \cdot \mathbf{u} \geq 0, \qquad \text{and} \qquad \mathbf{u} \cdot \mathbf{u} = 0 \Longleftrightarrow \mathbf{u} = \mathbf{0}$$

Using this definition of inner product, **most of the definitions and results** of Chapter 3.1-3.3 applies (length, distance, projection, orthogonal, orthonormal basis etc.) However we cannot talk about "angle" anymore in the complex case.

The properties of determinant is slightly changed:

Theorem A.4. The determinant of the adjoint is the complex conjugate:

$$\det \mathbf{A}^* = \overline{\det \mathbf{A}}$$

Next, for a linear isometry, that is, the linear transformation that preserves inner product:

$$(T\mathbf{u}) \cdot (T\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

instead of **orthogonal matrix**, we have the **unitary matrix**:

Definition A.5. A square matrix $\mathbf{U}$ corresponding to a linear isometry is called a **unitary matrix**. It is invertible with

$$\mathbf{U}^{-1} = \mathbf{U}^*$$

The set of $n \times n$ unitary matrices is denoted by $U(n)$. It satisfies all the properties of Theorem 3.25.

From the properties of determinant, we have

Theorem A.6. The determinant of a unitary matrix is a complex number with norm 1:

$$|\det \mathbf{U}| = 1$$

We just have to replace **orthogonal matrix** by **unitary matrix** everywhere in the setting of $K = \mathbb{C}$. For example:

If $\mathbf{A} = \mathbf{QR}$ is the QR decomposition and $\mathbf{A}$ is a square matrix, then $\mathbf{Q}$ is a unitary matrix.

By the Fundamental Theorem of Algebra, any degree n polynomial $p(\lambda)$ has n roots (counted with multiplicity). Therefore

Theorem A.7. Every complex $n \times n$ square matrix $\mathbf{A}$ has n eigenvalues (counted with multiplicity)

Instead of **symmetric matrix**, we have **Hermitian matrix**:

Definition A.8. An $n \times n$ square matrix $\mathbf{A}$ is called **Hermitian** if

$$\mathbf{A}^* = \mathbf{A}$$

Next we talk about diagonalization:

Definition A.9. A matrix $\mathbf{A}$ is **unitary diagonalizable** if

$$\mathbf{A} = \mathbf{UDU}^{-1} = \mathbf{UDU}^*$$

for some unitary matrix $\mathbf{U}$ and diagonal matrix $\mathbf{D}$.

Then we have the **Spectral Theorem of Hermitian Matrices** similar to the one of symmetric matrices (see Theorem 5.7, 5.9 and 5.10):

Theorem A.10 (Spectral Theorem for Hermitian Matrices)**.** We have

- If $\mathbf{A}$ is Hermitian, then two eigenvectors from different eigenspaces are orthogonal.
- $\mathbf{A}$ is Hermitian if and only if it is unitary diagonalizable.
- If $\mathbf{A}$ is Hermitian, then it has n real eigenvalues (counted with multiplicity). The characteristic polynomial $p(\lambda)$ has n real roots (counted with multiplicity).

Finally, as mentioned in Chapter 5.4, since $p(\lambda)$ always has n roots:

Theorem A.11. Every complex matrix $\mathbf{A}$ is similar to a Jordan Canonical Form.

Summarize: We have the corresponding definitions and results for $n \times n$ matrices:

$K = \mathbb{R}$		$K = \mathbb{C}$
Real inner product: $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n$	$\longrightarrow$	Complex inner product: $\mathbf{u} \cdot \mathbf{v} = \overline{u_1} v_1 + \cdots + \overline{u_n} v_n$
Transpose: $\mathbf{A}^T : (a_{ij})^T = (a_{ji})$	$\longrightarrow$	Adjoint (conjugate transpose): $\mathbf{A}^* : (a_{ij})^* = (\overline{a_{ji}})$
$\det \mathbf{A}^T = \det \mathbf{A}$	$\longrightarrow$	$\det \mathbf{A}^* = \overline{\det \mathbf{A}}$
Orthogonal matrix: $\mathbf{U}^T \mathbf{U} = \mathbf{Id}$	$\longrightarrow$	Unitary matrix: $\mathbf{U}^* \mathbf{U} = \mathbf{Id}$
$\det \mathbf{U} = \pm 1$	$\longrightarrow$	$\lvert \det \mathbf{U} \rvert = 1$
Symmetric matrix: $\mathbf{A}^T = \mathbf{A}$	$\longrightarrow$	Hermitian matrix: $\mathbf{A}^* = \mathbf{A}$
Orthogonal diagonalizable: $\mathbf{A} = \mathbf{PDP}^T$	$\longrightarrow$	Unitary diagonalizable: $\mathbf{A} = \mathbf{UDU}^*$
Some matrix may not have eigenvalues	$\longrightarrow$	All matrix has n eigenvalues (with multiplicity)
Symmetric matrix has n real eigenvalues	$\longrightarrow$	Hermitian matrix has n real eigenvalues
Jordan Canonical Form: when $p(\lambda)$ has n roots	$\longrightarrow$	Jordan Canonical Form: always