Linear Algebra

Author: Wenxiao Yang

Institute: University of Illinois at Urbana-Champaign; University of California Berkeley

Date: 2023 (Last Updated)

All models are wrong, but some are useful.

Contents

Chapter 1 Field and Vector Space

1.1 Field $(\mathbb{F}, +, \cdot)$

1.1.1 Definition of Field (@ Lec 02 of ECON 204)

Definition 1.1 (Field)

A **field** $\mathcal{F} = (\mathbb{F}, +, \cdot)$ is a 3-tuple consisting of a set $\mathbb{F}$ and two binary operations $+, \cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ such that

1. $+$ is associative, commutative.
2. Exists (unique) additive identity and (unique) additive inverse.
3. $\cdot$ is associative, commutative.
4. Exists (unique) multiplicative identity and (unique) multiplicative inverse.
5. Distributivity of multiplication over addition.

♣

Example 1.1 of Field

1. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ are field;
2. $\mathbb{N}$, $\mathbb{Z}$ are not field;
3. $\mathbb{Q}(\sqrt{2}) = \{q + r\sqrt{2} : q, r \in \mathbb{Q}\}$ is the smallest filed containing $\mathbb{Q} \cup \sqrt{2}$.

1.2 Vector Space

1.2.1 Definition of Vector Space (@ Lec 02 of ECON 204)

Definition 1.2 (Vector Space)

A vector space over a field $\mathbb{F}$, $(V, \mathbb{F}, +, \times)$, is a set V w/ an operation addition $+ : V \times V \to V$ and an operation scalar multiplication $\mathbb{F} \times V \to V$

1. $+$ is associative, commutative.
2. Exists (unique) additive identity and (unique) additive inverse.
3. $\cdot$ is associative (there is no need to consider commutativity).
4. Exists (unique) multiplicative identity (there is no need to consider inverse).
5. Distributivity of scalar multiplication over vector addition: $\forall \alpha \in \mathbb{F},\ v, u \in V,\ \alpha(v+u) = \alpha v + \alpha u$
6. Distributivity of scalar multiplication over scalar addition: $\forall \alpha, \beta \in \mathbb{F},\ v \in V,\ (\alpha+\beta)v = \alpha v + \beta v$

♣

We often say "V is a vector space over $\mathbb{F}$".

Example 1.2

1. $\mathbb{R}^n$ is a vector space over $\mathbb{R}$, for any $n \in \mathbb{N}$.
2. $\mathbb{Q}(\sqrt{2})$ is a vector space over $\mathbb{Q}$. (It is $\mathbb{Q}^2$, using (q, r) versus $q + r\sqrt{2}$).
3. $\mathbb{Q}(\sqrt[3]{2})$ is a vector space over $\mathbb{Q}$. (It is $\mathbb{Q}^3$, using (q, r, v) versus $q + r2^{\frac{2}{3}} + v2^{\frac{1}{3}}$).
4. $C([0, 1])$, the space of all continuous real-valued functions on $[0, 1]$, is a vector space over $\mathbb{R}$.

1.2.2 Theorem: A field is a vector space over its subfield

Theorem 1.1

$\mathbb{K} \subset \mathbb{F}$ is a subfield of a field $\mathbb{F}$. Then $\mathbb{F}$ is a vector space over $\mathbb{K}$. ♡

Example 1.3 Since $\mathbb{F} \subset \mathbb{F}[x]$, then $\mathbb{F}[x]$ is a vector space over $\mathbb{F}$.

1.2.3 Vector Subspace

Definition 1.3 (Vector Subspace)

Suppose that V is a vector space over $\mathbb{F}$. A vector subspace or just subspace is a nonempty subset $W \subset V$ closed under addition and scalar multiplication. i.e.

$$v + w \in W,\ av \in W,\ \forall v, w \in W,\ a \in \mathbb{F}$$

♣

Example 1.4 $\mathbb{K} \subset \mathbb{L} \subset \mathbb{F}$, then $\mathbb{L}$ is a subspace of $\mathbb{F}$ over $\mathbb{K}$.

1.2.4 Linear Independent

Definition 1.4 (Linear Independence)

A set $V \subseteq X$ is **linearly dependent** if there exist $v_1, ..., v_n \in V$ and $\alpha_1, ..., \alpha_n \in F$ not all zero such that $\sum_{i=1}^n \alpha_i v_i = 0$. ♣

Prove a set is linearly independent by $\sum_{i=1}^n \alpha_i v_i = 0 \Rightarrow \alpha_i = 0, \forall i$.

1.2.5 span V, basis, dimension

Definition 1.5 (Span)

If $V \subseteq X$, the span of V, denoted span V, is the set of all linear combinations of elements of V. The set $V \subseteq X$ **spans** X if span $V = X$. ♣

Definition 1.6 ((Hamel) Basis)

A **Hamel basis** (often just called a **basis**) of a vector space X is a **linearly independent** set of vectors in X that **spans** X. ♣

Theorem 1.2

Every vector space has a Hamel basis. ♡

Theorem 1.3 (Bases have the same cardinality)

Any two Hamel bases of a vector space X have the same cardinality (are numerically equivalent). ♡

Theorem 1.4 (Unique Representation)

Let V be a Hamel basis for X. Then every vector $x \in X$ has a ***unique representation*** *as a linear combination of a finite number of elements of V*

(with all coefficients nonzero, unique representation of $0 = \sum_{i \in \emptyset} \alpha_i v_i$). ♡

Proposition 1.1 (Proposition 2.4.10.)

Suppose V is a vector space over a field $\mathbb{F}$ having a basis $\{v_1, ..., v_n\}$ with $n \geq 1$.

(i). For all $v \in V$, $v = a_1 v_1 + ... + a_n v_n$ for exactly one $(a_1, ..., a_n) \in \mathbb{F}^n$ (unique representation).

(ii). If $w_1, ..., w_n$ span V, then they are linearly independent.

(iii). If $w_1, ..., w_n$ are linearly independent, then they span V. ♠

1.2.6 Dimension

Definition 1.7 (Dimension)

The dimension of a vector space X, denoted $\dim X$, is the cardinality of any basis of X. ♣

If a vector space V over $\mathbb{F}$ has a basis with n vectors, then V is said to be n-dimensional (over $\mathbb{F}$) or is said to have **dimension** n.

Definition 1.8 (Finite-Dimension)

Let X be a vector space. If $\dim X = n$ for some $n \in N$, then X is finite-dimensional. Otherwise, X is infinite-dimensional. ♣

1.2.7 Theorem: $|V| > \dim X \Rightarrow$ linearly dependent

Theorem 1.5 ($|V| > \dim X \Rightarrow V$ is linearly dependent)

Suppose $\dim X = n \in \mathbb{N}$. If $V \subseteq X$ and $|V| > n$, then V is linearly dependent. ♡

1.2.8 Theorem: $|V| = n$: Linear Indep $\Leftrightarrow$ Spans $\Rightarrow$ Basis

Theorem 1.6 (**$|V| = n$: Linear Indep $\Leftrightarrow$ Spans $\Rightarrow$ Basis**)

Suppose $\dim X = n \in \mathbb{N}$, *and* $V \subseteq X$ *and* $|V| = n$.

1. *If V is linearly independent, then V spans X, so V is a Hamel basis.*
2. *If V spans X, then V is linearly independent, so V is a Hamel basis.*

♡

1.2.9 Standard basis vectors

$e_1 = (1, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, 0, ..., 0, 1) \in \mathbb{F}^n$ are a basis for $\mathbb{F}^n$ called the **standard basis vectors**.

1.3 Linear Transformation

1.3.1 Definition of Linear Transformation

Definition 1.9 (Linear Transformation)

Given two vector spaces V and W over $\mathbb{F}$ a **linear transformation** is a function $T : V \to W$ such that for all $a \in \mathbb{F}$ and $v, w \in V$, we have

$$T(av) = aT(v) \text{ and } T(v + w) = T(v) + T(w)$$

♣

1.3.2 Composition of Linear Transformations is also Linear

Let $L(X, Y)$ denote the **set of all linear transformations** from X to Y over field $\mathbb{F}$.

Theorem 1.7

$L(X, Y)$ is a vector space of $\mathbb{F}$.

♡

Proposition 1.2 (Composition of Linear Transformations is also Linear)

Suppose $R \in L(X, Y)$, $S \in L(Y, Z)$, then $S \circ R \in L(X, Z)$.

♠

1.3.3 Function from a Basis extends uniquely to a Linear Transformation

Proposition 1.3 (Proposition 2.4.15.)

If V and W are vector spaces and $v_1, ..., v_n$ is a basis for V then any function from $\{v_1, ..., v_n\} \to W$ extends uniquely to a linear transformation $V \to W$.

♠

Any $v \in V$, $\exists(a_1, ..., a_n)$ s.t. $v = a_1 v_1 + ... + a_n v_n$. Then $T(v) = T(a_1 v_1 + ... + a_n v_n) = a_1 T(v_1) + ... + a_n T(v_n)$

1.3.4 Image, Kernel, Rank

Definition 1.10 (Image, Kernel, Rank)

Let $T \in L(X, Y)$,

- The **image** of T is $\text{Im } T = T(X) \subseteq Y$.
- The **kernel** of T is $\ker T = \{x \in X : T(x) = 0\}$.
- The **rank** of T is $\text{Rank } T = \dim(\text{Im } T)$.

♣

1.3.5 The Rank-Nullity Theorem: $\dim X = \dim \ker T + \text{Rank } T$

Theorem 1.8 (The Rank-Nullity Theorem)

Let X be a finite-dimensional vector space and $T \in L(X, Y)$. Then $\text{Im } T$ *and* $\ker T$ *are vector subspaces of Y and X respectively, and*

$$\dim X = \dim \ker T + \text{Rank } T$$

♡

1.3.6 Theorem: Linear Transformation T is 1-to-1 $\Leftrightarrow \ker T = \{0\}$

Theorem 1.9 (1-to-1 $\Leftrightarrow \ker T = \{0\}$)

$T \in L(X, Y)$ *is one-to-one if and only if* $\ker T = \{0\}$.

♡

1.3.7 Definition of Invertible Linear Transformation

Definition 1.11 (Invertible Linear Transformation)

$T \in L(X, Y)$ is **invertible** if there is a function $S : Y \to X$ such that

1. $S(T(x)) = x, \forall x \in X$.
2. $T(S(y)) = y, \forall y \in Y$.

♣

Proposition 1.4

$T \in L(X, Y)$ *is* ***invertible*** $\Leftrightarrow$ *bijective (1-to-1 and onto).*

♠

1.3.8 Theorem: Inverse of a Linear Transformation is also Linear

Theorem 1.10 (Inverse of a Linear Transformation is also Linear)

If $T \in L(X, Y)$ is invertible, then $T^{-1} \in L(Y, X)$, i.e. T^{-1} is linear.

♡

1.3.9 Theorem: Linear Transformation is completely determined by values on basis

Theorem 1.11 (Linear Transformation is completely determined by values on basis)

Let X, Y be two vector spaces over the same field F, and let $V = \{v_\lambda : \lambda \in \Lambda\}$ be a basis for X. Then a linear transformation $T \in L(X, Y)$ is completely determined by its values on V, that is:

1. *Given any set $\{y_\lambda : \lambda \in \Lambda\} \subseteq Y$, $\exists T \in L(X, Y)$ s.t.*

$$T(v_\lambda) = y_\lambda, \ \forall \lambda \in \Lambda$$

2. *If $S, T \in L(X, Y)$ and $S(v_\lambda) = T(v_\lambda)$ for all $\lambda \in \Lambda$, then $S = T$.*

♡

Proof 1.1

1. *By the definition of linear transformation, we can prove the result by proving $T(x) = T(\sum_i \alpha_i v_i) = \sum_i \alpha_i y_i$ is linear.*
2. *Suppose $S(v_\lambda) = T(v_\lambda)$. We prove the result by proving $S(x) = S(\sum_i \alpha_i v_i) = \sum_i \alpha_i S(v_i) = \sum_i \alpha_i T(v_i) = T(\sum_i \alpha_i v_i) = T(x)$.*

1.3.10 GL(V): set of invertible linear transformations $V \to V$

Given a vector space V over F, we let $GL(V) \subset L(V, V)$ denote the subset of **invertible linear transformations**.

$$GL(V) = \{T \in L(V, V) | T \textit{ is a bijection}\} = L(V, V) \cap Sym(V)$$

1.4 Isomorphism

1.4.1 Isomorphic: $\exists$ invertible $T \in L(X, Y)$

Definition 1.12 (Isomorphisms)

Two vector spaces X, Y over a field F are **isomorphic** if there is an invertible $T \in L(X, Y)$.

♣

We use notation $\cong$ to denote isomorphism.

1.4.2 Theorem: Isomorphic $\Leftrightarrow \dim X = \dim Y$

Theorem 1.12 (Isomorphic $\Leftrightarrow \dim X = \dim Y$)

*Two vector spaces X, Y over the same field are **isomorphic** ($\exists$ invertible $T \in L(X, Y)$) <u>if and only if</u>*

$$\dim X = \dim Y$$

♡

1.5 Quotient Vector Spaces

1.5.1 Quotient Vector Space: X/W

Given a vector space $(X, \mathbb{F}, +, \times)$ over $\mathbb{F}$ and a vector subspace $W \subseteq X$. Define an equivalence relation:

$$x \sim y \Leftrightarrow \exists w \in W \text{ s.t. } x = y + w$$

(or $x \sim y \Leftrightarrow x - y \in W$). The equivalence class of $x \in X$ with respect to $\sim$ is

$$[x] = \{x + w : w \in W\}$$

We define a set, which is read ”X mod W”, by

$$X/W = \{[x] : x \in X\} = \{x + W : x \in X\}$$

Define operations $+, \cdot$ in X/W by $\alpha[x] + \beta[y] = [\alpha x + \beta y]$, $\forall x, y \in X$ and $\forall \alpha, \beta \in \mathbb{F}$.

Definition 1.13 (Quotient Vector Spaces)

X/W forms a vector space over $\mathbb{F}$ with $+, \cdot$. ♣

1.5.2 Theorem: $\dim(X/W) = \dim X - \dim W$

Theorem 1.13 ($\dim(X/W) = \dim X - \dim W$)

If X is a vector space with $\dim X = n$ for some $n \in \mathbb{N}$ and W is a vector subspace of X, then

$$\dim(X/W) = \dim X - \dim W$$

♡

Proof 1.2

Given a basis for W, $\{w_1, ..., w_c\}$, and a basis for X/W, $\{[x_1], ..., [x_k]\}$. Show that $\{w_1, ..., w_c\} \cup \{x_1, ..., x_k\}$ is a basis for X.

1.5.3 Theorem: T is isomorphic to $X/\ker T$

Theorem 1.14 ($\operatorname{Im} T$ **is isomorphic to** $X/\ker T$)

Let X and Y be vector spaces over the same field F and $T \in L(X, Y)$. Then $\operatorname{Im} T = T(X)$ is isomorphic to $X/\ker T$, where $\ker T = \{x \in X : T(x) = 0\}$. ♡

Chapter 2 Matrix

2.1 Coordinate Representation

Let X be a finite-dimensional vector space over $\mathbb{R}$ with $\dim X = n$. Fix any Hamel basis $V = \{v_1, ..., v_n\}$ of X. According to the theorem 1.4, any $x \in X$ has a unique representation $x = \sum_{j=1}^n \beta_j v_j$ (here, we allow $\beta_j = 0$). Given this representation, we can write

$$\mathrm{crd}_V(x) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathbb{R}^n$$

Then, $\mathrm{crd}_V(x)$ is the vector of coordinates of x with respect to the basis V.

$$\mathrm{crd}_V(v_1) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathrm{crd}_V(v_2) = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \ \mathrm{crd}_V(v_n) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

crd_V is an isomorphism from X to $\mathbb{R}^n$.

2.2 Matrix Representation of a Linear Transformation

2.2.1 Definition: matrix corresponds to a linear transformation

Suppose $T \in L(X, Y)$, $\dim X = n$ and $\dim Y = m$. Fix bases

$$V = \{v_1, \ldots, v_n\} \text{ of } X$$

$$W = \{w_1, \ldots, w_m\} \text{ of } Y$$

$T(v_j) \in Y$, so

$$T(v_j) = \sum_{i=1}^m \alpha_{ij} w_i$$

Define a matrix corresponds to T by

$$\mathrm{Mtx}_{W,V}(T) = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}$$

where

$$[w_1, \ldots, w_m] \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} = [T(v_1), \ldots, T(v_n)]$$

Notice that the columns are the coordinates (expressed with respect to W) of $T(v_1), \ldots, T(v_n)$. Observe

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \vdots \\ \alpha_{m1} \end{pmatrix}$$

so

$$\text{Mtx}_{W,V}(T) \cdot \text{crd}_V(v_j) = \text{crd}_W(T(v_j))$$

$$\Rightarrow \text{Mtx}_{W,V}(T) \cdot \text{crd}_V(x) = \text{crd}_W(T(x)) \quad \forall x \in X$$

Multiplying a vector by a matrix does two things:

1. Computes the action of T.
2. Accounts for the change in basis.

Example 2.1 $X = Y = \mathbb{R}^2$, $V = \{(1,0),(0,1)\}$, $W = \{(1,1),(-1,1)\}$, $T = id$, that is, $T(x) = x$ for all x.

$\text{Mtx}_{W,V}(T) = \text{Mtx}_{W,V}(id)$ is the matrix that changes basis from V to W. How do we compute it?

$[w_1, w_2]\text{Mtx}_{W,V}(id) = [v_1, v_2] \Rightarrow a_{11}w_1 + a_{21}w_2 = T(v_1)$ and $a_{12}w_1 + a_{22}w_2 = T(v_2) \Rightarrow \text{Mtx}_{W,V}(id) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

2.2.2 Theorem: Matrix Space $\cong$ Linear Transformation Space

Theorem 2.1 (Matrix Space $\cong$ Linear Transformation Space)

Let X and Y be vector spaces over the same field F, with $\dim X = n$, $\dim Y = m$. Then $L(X,Y)$, the space of linear transformations from X to Y, is isomorphic to $F_{m\times n}$, the vector space of $m \times n$ matrices over F. If $V = \{v_1, ..., v_n\}$ is a basis for X and $W = \{w_1, ..., w_m\}$ is a basis for Y, then

$$\text{Mtx}_{W,V}(T) \in L(L(X,Y), F_{m\times n})$$

and $\text{Mtx}_{W,V}$ is an isomorphism from $L(X,Y)$ to $F_{m\times n}$. ♡

2.2.3 Theorem: $\mathrm{Mtx}_{W,V}(T) \cdot \mathrm{Mtx}_{V,U}(S) = \mathrm{Mtx}_{W,U}(T \circ S)$

Theorem 2.2 ($\mathrm{Mtx}_{W,V}(T) \cdot \mathrm{Mtx}_{V,U}(S) = \mathrm{Mtx}_{W,U}(T \circ S)$)

Let X, Y, Z be finite-dimensional vector spaces over the same field F with bases U, V, W respectively. Let $S \in L(X, Y)$ and $T \in L(Y, Z)$. Then

$$\mathrm{Mtx}_{W,V}(T) \cdot \mathrm{Mtx}_{V,U}(S) = \mathrm{Mtx}_{W,U}(T \circ S)$$

i.e. matrix multiplication corresponds via the matrix representation isomorphism to composition of linear transformations. ♡

The theorem can be summarized by the following "Commutative Diagram:"

$$\begin{array}{ccccccc} & & S & & T & & \\ & X & \rightarrow & Y & \rightarrow & Z & \\ crd_U & \updownarrow & & \updownarrow crd_V & & \updownarrow crd_W & \\ & \mathbf{R}^n & \rightarrow & \mathbf{R}^m & \rightarrow & \mathbf{R}^r & \\ & & Mtx_{V,U}(S) & & Mtx_{W,V}(T) & & \end{array}$$

Figure 2.1: Commutative Diagram

We say the diagram commutes because you get the same answer any way you go around the diagram (in directions allowed by the arrows). The crd arrows go in both directions because crd is an isomorphism.

2.3 Change of Basis and Similarity

Let X be a finite-dimensional vector space with basis V. If $T \in L(X, X)$ it is customary to use the same basis in the domain and range. In this case, we use $\mathrm{Mtx}_V(T)$ to denotes $\mathrm{Mtx}_{V,V}(T)$.

2.3.1 Change of Basis

Question: If W is another basis for X, how are $Mtx_V(T)$ and $Mtx_W(T)$ related?

$$\begin{aligned} \mathrm{Mtx}_{V,W}(id) \cdot Mtx_W(T) \cdot Mtx_{W,V}(id) &= Mtx_{V,W}(id) \cdot Mtx_{W,V}(T \circ id) \\ &= Mtx_{V,V}(id \circ T \circ id) \\ &= Mtx_V(T) \end{aligned}$$

and

$$\mathrm{Mtx}_{V,W}(id)\cdot Mtx_{W,V}(id)=Mtx_{V,V}(id)\quad=\begin{pmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1\end{pmatrix}$$

So this says that

$$Mtx_V(T)=P^{-1}Mtx_W(T)P$$

for the invertible matrix

$$P=Mtx_{W,V}(id)$$

that is the change of basis matrix.

On the other hand, if P is any invertible matrix, then P is also a change of basis matrix for appropriate corresponding bases.

2.3.2 Similarity

Definition 2.1 (Similar)

Square matrices A and B are **similar** if

$$A=P^{-1}BP$$

for some invertible matrix P. ♣

2.3.3 Theorem: Similar Matrices ⇔ Same Linear Transformation for Two Bases

Theorem 2.3 (Similar Matrices ⇔ Same Linear Transformation for Two Bases)

Suppose that X is a finite-dimensional vector space.

1. *If $T\in L(X,X)$ then any two matrix representations of T are* ***similar****. That is, if U,W are any two bases of X, then $\mathrm{Mtx}_W(T)$ and $\mathrm{Mtx}_U(T)$ are similar.*
2. *Conversely, two similar matrices represent the same linear transformation T, relative to suitable bases. That is, given similar matrices A,B with $A=P^{-1}BP$ and any basis U, there is a basis W and $T\in L(X,X)$ such that*

$$B=\mathrm{Mtx}_U(T),\ A=\mathrm{Mtx}_W(T),\ P=\mathrm{Mtx}_{U,W}(id),\ P^{-1}=\mathrm{Mtx}_{W,U}(id).$$ ♡

2.4 Square Matrix $A_{n\times n}$: $det(A)$, singular

2.4.1 Determinant $det(A)$

1. A is singular if $det(A) = 0$, else non-singular.
2. If $det(A) \neq 0$, A^{-1} exists and $A^{-1} = \frac{adj(A)}{det(A)}$
3. $det(AB) = det(A)det(B)$

2.5 Eigenvalues and Eigenvectors

Definition 2.2 (Matrix Form: Eigenvalues, Eigenvectors)

A vector x is a **eigenvector** of a matrix A if Ax is parallel to x, that is if $Ax = \lambda x$ for some number $\lambda \in \mathbb{R}$. The number λ is called a **eigenvalue** of A.

i.e. the root of $(A - \lambda I_n)x = 0 \Leftrightarrow det(A - \lambda I_n) = 0$

♣

Definition 2.3 (Linear Transformation: Eigenvalues, Eigenvectors)

Let X be a vector space and $T \in L(X, X)$. We say that λ is an **eigenvalue** of T and $v \neq 0$ is an **eigenvector** corresponding to λ if $T(v) = \lambda v$.

♣

Theorem 2.4 (Matrix Form and LT Form are equivalent)

Let X be a finite-dimensional vector space, and U a basis. Then λ is an eigenvalue of T <u>if and only if</u> λ is an eigenvalue of $\mathrm{Mtx}_U(T)$. v is an eigenvector of T corresponding to λ <u>if and only if</u> $\mathrm{crd}_U(v)$ is an eigenvector of $\mathrm{Mtx}_U(T)$ corresponding to λ.

♡

2.6 Diagonalizable Matrix

Definition 2.4 (Diagonalizable)

Suppose X is a finite-dimensional vector space with basis U. Given a linear transformation $T \in L(X, X)$, let

$$A = \mathrm{Mtx}_U(T)$$

We say that A can be diagonalized (or is **diagonalizable**) if there is another basis W for X such that

$\mathrm{Mtx}_W(T)$ is diagonal, i.e.

$$\mathrm{Mtx}_W(T) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

♣

A $n \times n$ matrix A with n linearly independent eigenvalues u is said to be *diagonalizable*.

$$\begin{aligned} AU = A \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix} &= \begin{bmatrix} | & | & \cdots & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & \cdots & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ &= UD \\ \Rightarrow A &= UDU^{-1} \end{aligned}$$

Theorem 2.5 (Linearly Independent Eigenvectors $\Rightarrow A = UDU^{-1}$)

If an $n \times n$ matrix A has n linearly independent eigenvectors $u_1, \ldots, u_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$, then $A = UDU^{-1}$ where D is diagonal with entries $\lambda_1, \ldots, \lambda_n$, and U has columns $u_1, \ldots, u_n$. ♡

A is **similar** to D ($\exists P$ s.t. $A = PDP^{-1}$).

Not *diagonalizable* is also called *defective*.

Theorem 2.6 (Diagonalized $\Leftrightarrow \exists$ basis consisiting of eigenvectors)

Let X be an n-dimensional vector space, $T \in L(X, X)$, U any basis of X, and $A = \mathrm{Mtx}_U(T)$. Then the following are equivalent:

1. *A can be diagonalized.*
2. *There is a basis W for X consisting of eigenvectors of T.*
3. *There is a basis V for $\mathbb{R}^n$ consisting of eigenvectors of A.*

♡

Theorem 2.7 (Distinct Eigencalues Correspond to Distinct Eigenvectors)

Let X be a vector space and $T \in L(X, X)$.

1. *If $\lambda_1, ..., \lambda_m$ are distinct eigenvalues of T with corresponding eigenvectors $v_1, ..., v_m$, then $\{v_1, ..., v_m\}$ is linearly independent.*
2. *If $\dim X = n$ and T has n distinct eigenvalues, then X has a basis consisting of eigenvectors of T; consequently, if U is any basis of X, then $\text{Mtx}_U(T)$ is diagonalizable.*

♡

Corollary 2.1

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

♡

(Because the n associated eigenvectors are always linearly independent.)

2.7 Orthogonal, Orthonormal, Unitary

Two vectors a and b are orthogonal, if their dot product is equal to zero (they are perpendicular).

$$a \cdot b = 0$$

Two vectors a and b are orthonormal, if they are orthogonal **unit vectors**.

Formally,

Definition 2.5 (Orthonormal)

Let $\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$. A basis $V = \{v_1, ..., v_n\}$ of $\mathbb{R}^n$ is **orthonormal** if $v_i \cdot v_j = \delta_{ij}$.

♣

Definition 2.6 (Unitary)

A real $n \times n$ matrix A is **unitary** if $A^T = A^{-1}$.

♣

Theorem 2.8

A real $n \times n$ matrix A is ***unitary*** *if and only if the columns of A are* ***orthonormal****.*

♡

If A is unitary, let V be the set of columns of A and W be the standard basis of $\mathbb{R}^n$. Since A is unitary, it is invertible, so V is a basis of $\mathbb{R}^n$.

$$A^\top = A^{-1} = \text{Mtx}_{V,M}(id)$$

where $A = [v_1, ..., v_n]$: $A\text{Mtx}_{V,M}(id) = [v_1, ..., v_n]\text{Mtx}_{V,M}(id) = [m_1, ..., m_n] = \mathbb{I}_{n \times n} \Rightarrow \text{Mtx}_{V,M}(id) = A^{-1}$.

Since V is orthonormal, the transformation between bases W and V preserves all geometry, including lengths and angles.

2.8 Eigen Decomposition of Symmetric Matrices Results

Let A be a symmetric $n \times n$ matrix, i.e. $A^T = A$

Proposition 2.1

All eigenvalues of A are real.

Proposition 2.2

Eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof 2.1

Let λ_1, λ_2 be eigenvalues s.t. $\lambda_1 \neq \lambda_2$.

$$\begin{aligned} Au_1 = \lambda_1 u_1;\ Au_2 = \lambda_2 u_2 \\ \lambda_1 u_1^T u_2 = (Au_1)^T u_2 = u_1^T A^T u_2 \\ = u_1^T A u_2 = u_1^T (Au_2) = \lambda_2 u_1^T u_2 \\ \Rightarrow u_1^T u_2 = 0 \text{ Since } \lambda_1 \neq \lambda_2 \end{aligned}$$

Proposition 2.3

If λ is an eigenvalue with multiplicity k, we can find k orthogonal eigenvectors for λ.

Multiplicity: the number of times an element is repeated in a multiset.

2.9 Diagonalization of Real Symmetric Matrices

Theorem 2.9 (Diagonalization of Real Symmetric Matrices)

Let $T \in L(\mathbb{R}^n, \mathbb{R}^n)$ and W be the standard basis of $\mathbb{R}^n$. Suppose that $Mtx_W(T)$ is symmetric. Then the eigenvectors of T are all real, and there is an orthonormal basis $V = \{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of T, so that $Mtx_W(T)$ is diagonalizable:

$$Mtx_W(T) = Mtx_{W,V}(id) \cdot Mtx_V(T) \cdot Mtx_{V,W}(id)$$

where $Mtx_V(T)$ is diagonal and the change of basis matrices $Mtx_{V,W}(id)$ and $Mtx_{W,V}(id)$ are unitary.

A real symmetric matrix $A_{n \times n}$ can be written as

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U \Omega U^T$$

u_i are orthonormal eigenvectors. λ_i are eigenvalues.

Where $U = [u_1, u_2, ..., u_n]$, $\Omega = diag(\lambda_1, ..., \lambda_n)$

Since u_i are orthonormal eigenvectors, $U^TU = I \Rightarrow U^T = U^{-1}$. U is an orthogonal matrix.

2.9.1 Proposition: $\lambda_{\min}\|x\|^2 \leq x^TAx \leq \lambda_{\max}\|x\|^2$

Proposition 2.4

For any $x \in \mathbb{R}^n$,

$$\lambda_{\min}\|x\|^2 \leq x^TAx \leq \lambda_{\max}\|x\|^2$$

Proof 2.2

Since u_i are orthonormal and linearly independent. $x = \sum_{i=1}^n \alpha_i u_i$ for some $\alpha_i \in \mathbb{R}, i = 1, ..., n$

$$\begin{aligned}
x^TAx &= (\sum_{i=1}^n \alpha_i u_i)^T A (\sum_{j=1}^n \alpha_j u_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j u_i^T A u_j \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j u_i^T (A u_j) \\
&= \sum_{i=1}^n \alpha_i^2 \lambda_i
\end{aligned}$$

$$\Rightarrow \lambda_{\min}\|x\|^2 \leq x^TAx \leq \lambda_{\max}\|x\|^2$$

The first equation holds if x is the eigenvector for $\lambda_{\min}$. The second equation holds if x is the eigenvector for $\lambda_{\max}$.

2.9.2 Proposition: λ^2 is the eigenvalue of A^2 and A^TA

Proposition 2.5

If λ is an eigenvalue of A, then λ^2 is the eigenvalue of A^2 and A^TA, the corresponding eigenvector doesn't change.

Proof 2.3

$$\begin{aligned}
Ax_1 &= \lambda x_1 \\
A^2x_1 = A(Ax_1) = \lambda A x_1 &= \lambda^2 x_1 \\
A^TAx_1 = A^T(Ax_1) = \lambda A^T x_1 &= \lambda^2 x_1
\end{aligned}$$

2.10 Trace

$A_{n\times n}, Tr(A) = \sum_{i=1}^{n} A_{kk}$

$$det(A) = \prod_{i=1}^{n} \lambda_i,\ Tr(A) = \sum_{i=1}^{n} \lambda_i$$

Proposition 2.6 (Invariance Property)

$A_{m\times n}, B_{n\times k}, C_{k\times m}, Tr(ABC) = Tr(CAB) = Tr(BCA).$ ♠

2.11 Jacobian matrix

Suppose $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a function such that each of its first-order partial derivatives exist on $\mathbb{R}^n$. This function takes a point $\mathbf{x} \in \mathbb{R}^n$ as input and produces the vector $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$ as output. Then the Jacobian matrix of $\mathbf{f}$ is defined to be an $m \times n$ matrix, denoted by $\mathbf{J}$, whose (i, j) th entry is $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$, or explicitly

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

where $\nabla^{\mathrm{T}} f_i$ is the transpose (row vector) of the gradient of the i component.

2.12 Hessian matrix

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function taking as input a vector $\mathbf{x} \in \mathbb{R}^n$ and outputting a scalar $f(\mathbf{x}) \in \mathbb{R}$. If all second partial derivatives of f exist and are continuous over the domain of the function, then the Hessian matrix $\mathbf{H}$ of f is a square $n \times n$ matrix, usually defined and arranged as follows:

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

or, by stating an equation for the coefficients using indices i and j,

$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

The Hessian matrix is a symmetric matrix, since the hypothesis of continuity of the second derivatives implies that the order of differentiation does not matter (Schwarz's theorem).

The determinant of the Hessian matrix is called the Hessian determinant.

2.13 Positive Definite Matrices

2.13.1 Definition

We say that a symmetric $n \times n$ matrix A is:

(1). **positive semidefinite (PSD)** (written $A \succeq 0$) if $x^T Ax \geq 0$ for all x.

(2). **positive definite (PD)** (written $A \succ 0$) if $x^T Ax > 0$ for all $x \neq 0$.

(3). **negative semidefinite (NSD)** (written $A \preceq 0$) if $x^T Ax \leq 0$ for all x.

(4). **negative definite (ND)** (written $A \prec 0$) if $x^T Ax < 0$ for all $x \neq 0$.

(5). **indefinite** (not written in any particular way) if none of the above apply.

$x^T Ax$ is a function of x called the quadratic form associated to A.

A is ND(NSD) $\Leftrightarrow -A$ is PD(PSD)

Note: $A^T A$ is **positive semidefinite**, since $x^T A^T Ax = \|Ax\|^2 \geq 0$.

Note: We can extend definition to non-symmetric $n \times n$

$$x^T Ax = x^T A^T x \Rightarrow x^T Ax = x^T (\frac{A + A^T}{2})x$$

2.13.2 Condition number (for PD matrix)

Condition number (for PD matrix):

$$\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}} > 0$$

2.13.3 Diagonal matrix situation

$$D = \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{bmatrix}$$

Lemma 2.1

If $d_1, ...d_n$ are all nonnegative, then $D \succeq 0$;

If $d_1, ...d_n$ are all positive, then $D \succ 0$;

If $d_1, ...d_n$ are all nonpositive, then $D \preceq 0$;

If $d_1, ...d_n$ are all negative, then $D \prec 0$;

♡

2.13.4 Using eigenvalues

If A is an $n \times n$ symmetric matrix, then it can be factored as

$$A = Q^T \Lambda Q = Q^T \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{bmatrix} Q$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A and the columns of Q are the corresponding eigenvectors.

We can get $x^T Ax = x^T Q^T \Lambda Qx = (Qx)^T \Lambda (Qx)$

If we substitute $y = Qx$:

$x^T Ax = y^T \Lambda y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_n y_n^2$

Theorem 2.10

If $\lambda_1, ...\lambda_n$ are all non-negative, then symmetric matrix $A \succeq 0$;

If $\lambda_1, ...\lambda_n$ are all positive, then $A \succ 0$;

If $\lambda_1, ...\lambda_n$ are all non-positive, then $A \preceq 0$;

If $\lambda_1, ...\lambda_n$ are all negative, then $A \prec 0$;

if it has both positive and negative eigenvalues, then A is indefinite

2.13.5 Sylvester's Criterion

Consider a $n \times n$ matrix A:

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

Denote its $k \times k$ submatrix $A^{(k)}$:

$$A^{(k)} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1k} \\ a_{21} & a_{22} & ... & a_{2k} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kk} \end{bmatrix}$$

Let $\Delta_k = det(A^{(k)})$

$$det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)...(\lambda_n - x)$$

Note *The left side gives the coefficient of x is 1.*

By setting $x = 0$, we get

$$det(A) = \lambda_1 \lambda_2 ... \lambda_n$$

When $A \succ 0$, all the eigenvalues are positive, so $det(A) > 0$ as well.

$A \succ 0 \Rightarrow \vec{x}^T A \vec{x} > 0$ for all $\vec{x} \neq \vec{0}$. Consider $\vec{x} \in \mathbb{R}^n$ with $x_{k+1} = \cdots = x_n = 0$. $\vec{x} = [x_1, x_2, ..., x_k, 0, ...0]^T$.

Then,

$$\vec{x}^T A \vec{x} = [x_1, x_2, ..., x_k, 0, ...0] \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{x}^T A^{(k)} \vec{x}$$

Then we know $A \succ 0 \Rightarrow A^{(k)} \succ 0$

We expect $A^{(k)} \succ 0 \Rightarrow \Delta_k > 0$ for all k.

Theorem 2.11 (Sylvester's criterion)

Let A be $n \times n$ symmetric matrix

1. $A \succ 0$ *iff* $\Delta_i > 0 \; \forall i = 1, ..., n$
2. $A \prec 0$ *iff* $(-1)^i \Delta_i > 0 \; \forall i = 1, ..., n$
3. *A is indefinite if the first $\Delta_k$that breaks each pattern respectively is the wrong sign (rather than 0).*

Proposition 2.7

1. *Symmetric matrix A is PD*

   $\Leftrightarrow$ *All eigenvalues of A are > 0*

   $\Leftrightarrow \Delta_i > 0 \; \forall i = 1, ..., n$
2. *Symmetric matrix A is PSD*

   $\Leftrightarrow$ *All eigenvalues of A are ≥ 0*

   $\Leftrightarrow \Delta_i \geq 0 \; \forall i = 1, ..., n$
3. *For ND and NSD, test $-A$ instead of A*

2.13.6 Cholesky Decomposition: $A = LL^T$

Proposition 2.8 (Cholesky Decomposition)

Suppose $\boldsymbol{A}$ is a real matrix with symmetric positive-definite. $\boldsymbol{A}$ can be decomposed into the product of two matrices,

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$$

where $\boldsymbol{L}$ is a real lower triangular matrix with positive diagonal entries. ♠

Chapter 3 Linear Maps between Normed Spaces

Definition 3.1 (Bounded Linear Transformation)

Suppose X, Y are normed vector spaces and $T \in L(X, Y)$. We say T is **bounded** if

$$\exists \beta \in \mathbb{R} \text{ s.t. } \|T(x)\|_Y \leq \beta \|x\|_X, \forall x \in X$$

Note this implies that T is Lipschitz with constant β. ♣

Theorem 3.1

Let X, Y be normed vector spaces and $T \in L(X, Y)$. Then

$$\begin{aligned} &T \text{ is continuous at some point } x_0 \in X \\ \Leftrightarrow &T \text{ is continuous at every } x \in X \\ \Leftrightarrow &T \text{ is uniformly continuous on } X \\ \Leftrightarrow &T \text{ is Lipschitz} \\ \Leftrightarrow &T \text{ is bounded} \end{aligned}$$

♡

Note *Every linear map on a finite-dimensional normed vector space is bounded (and thus continuous, uniformly continuous, and Lipschitz continuous).*

Theorem 3.2 (LP in finite-dimensional normed vector space is bounded)

Let X, Y be normed vector spaces with dim $X = n$. Every $T \in L(X, Y)$ is bounded (and thus continuous, uniformly continuous, and Lipschitz continuous). ♡

Definition 3.2 (Topological Isomorphism)

A **topological isomorphism** between normed vector spaces X and Y is a linear transformation $T \in L(X, Y)$ that is invertible (one-to-one, onto), continuous, and has a continuous inverse.

Two normed vector spaces X and Y are **topologically isomorphic** if there is a topological isomorphism $T : X \to Y$. ♣

Suppose X and Y are normed vector spaces. We define

$$B(X, Y) = \{T \in L(X, Y) : T \text{ is bounded}\}$$

and

$$\begin{aligned} \|T\|_{B(X,Y)} &= \sup \left\{ \frac{\|T(x)\|_Y}{\|x\|_X}, x \in X, x \neq 0 \right\} \\ &= \sup\{\|T(x)\|_Y : \|x\|_X = 1\} \end{aligned}$$

Theorem 3.3 (($B(X,Y), \|\cdot\|_{B(X,Y)}$) **is normed vector space)**

Let X, Y be normed vector spaces. Then

$$\left(B(X,Y), \|\cdot\|_{B(X,Y)}\right)$$

is a ***normed vector space****.*

Theorem 3.4

Let $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ $(= B(\mathbb{R}^n, \mathbb{R}^m))$ with matrix $A = \{a_{ij}\}_{i,j}$ with respect to the standard bases. Let

$$M = \max\{|a_{ij}| : 1 \leq i \leq m, 1 \leq j \leq n\}$$

Then

$$M \leq \|T\| \leq M\sqrt{mn}$$

Theorem 3.5

Let $R \in L(\mathbb{R}^m, \mathbb{R}^n)$ and $S \in L(\mathbb{R}^n, \mathbb{R}^p)$. Then

$$\|S \circ R\| \leq \|S\|\|R\|$$

Define

$$\Omega\left(\mathbb{R}^n\right) = \left\{T \in L\left(\mathbb{R}^n, \mathbb{R}^n\right) : T \text{ is invertible }\right\}$$

Theorem 3.6

Suppose $T \in L\left(\mathbb{R}^n, \mathbb{R}^n\right)$ and E is the standard basis of $\mathbb{R}^n$. Then

$$\begin{aligned}
&T \textit{ is invertible} \\
&\Leftrightarrow \ker T = \{0\} \\
&\Leftrightarrow \det\left(Mtx_E(T)\right) \neq 0 \\
&\Leftrightarrow \det\left(Mtx_{V,V}(T)\right) \neq 0 \textit{ for every basis } V \\
&\Leftrightarrow \det\left(Mtx_{V,W}(T)\right) \neq 0 \textit{ for every pair of bases } V, W
\end{aligned}$$

Theorem 3.7

If $S, T \in \Omega\left(\mathbb{R}^n\right)$, then $S \circ T \in \Omega\left(\mathbb{R}^n\right)$ and

$$(S \circ T)^{-1} = T^{-1} \circ S^{-1}$$

Theorem 15 (Thm. 4.14) Let $S, T \in L\left(\mathbb{R}^n, \mathbb{R}^n\right)$. If T is invertible and

$$\|T - S\| < \frac{1}{\|T^{-1}\|}$$

then S is invertible. In particular, $\Omega\left(\mathbb{R}^n\right)$ is open in $L\left(\mathbb{R}^n, \mathbb{R}^n\right) = B\left(\mathbb{R}^n, \mathbb{R}^n\right)$. ♡

Theorem 3.8

The function $(\cdot)^{-1} : \Omega\left(\mathbb{R}^n\right) \to \Omega\left(\mathbb{R}^n\right)$ that assigns T^{-1} to each $T \in \Omega\left(\mathbb{R}^n\right)$ is continuous. ♡

Chapter 4 Euclidean geometry basics

4.1 Metrics

4.1.1 Vector's Norm

Vector $x \in \mathbb{R}^n$-n-dim Euclidean space

$$x = (x_1, \ldots, x_n) \equiv \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Norm of x, $\|x\|$ satisfies properties:

(a) $\|x\| \geqslant 0$

(b) $\|x\| = 0 \Leftrightarrow x = 0$

(c) $\|cx\| = |c|\|x\|$, for $c \in \mathbb{R}$

(d) $\|x + y\| \leqslant \|x\| + \|y\| \longleftarrow$ Triangle Ineq.

Enclidean Norm (default $\rho = 2$): $\|x\| = \sqrt{x^\top x} = \sqrt{\sum_{i=1}^n x_i{}^2}$

Other norms:

1. l_1-norm : $\|x\|_1 = \sum_{i=1}^n |x_i|$
2. l_ρ-norm : $\|x\|_\rho = \sqrt[\rho]{\sum_{i=1}^n |x_i|^\rho}$
3. Supremum norm or l_∞-norm : $\|x\|_\infty = \max_i |x_i|$

4.1.2 Matrix's Norm

Definition 4.1 (Matrix Norm)

The matrix norm defined in real vector space is a function $\|\cdot\| : \mathbb{R}^{n\times m} \to \mathbb{R}$ that satisfies following properties: for all scalars $\alpha \in \mathbb{R}$ and matrices $A, B \in \mathbb{R}^{n\times m}$

(1). $\|A\| \geq 0$

(2). $\|A\| = 0 \Leftrightarrow A = 0$

(3). $\|\alpha A\| = |\alpha|\|A\|$

(4). $\|A + B\| \leq \|A\| + \|B\|$ (sub-additive)

(5). $\|AB\| \leq \|A\|\|B\|$ (sub-multiplicative)

In some definitions, it may not hold, but every norm can be rescaled to be sub-multiplicative.

Here, we let it hold.

Definition 4.2 (Common Norms)

$A \in \mathbb{R}^{n \times m}$ is a matrix.

- ($\rho = 1$, *Default*): $\|A\| = \|A\|_1 = \max_{\|x\|_1=1} \|Ax\| = \max_j \sum_{i=1}^n |A_{ij}|$

  i.e., find the column with the highest sum of absolute values.

  $\|A\| \geqslant \frac{\|Ax\|}{\|x\|}$, $\|Ax\| \leqslant \|A\|\|x\|$
- (Frobenius Norm): $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$

  Frobenius norm property:

$$\|A\|_F^2 = \langle A, A \rangle = \| \underbrace{\text{vec}(A)}_{mn\times 1} \|^2 = trace(A^T A) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2$$

- ($\rho = \infty$): $\|A\|_\infty = \max_j \sum_{j=1}^n |A_{ij}|$

  i.e., find the row with the highest sum of absolute values.
- ($\rho = 2$, Spectral Norm, Euclidean Norm): $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A)$.

  σ_k is the singular value

Proposition 4.1

Consider Frobenius norm. Suppose a, b are 2 column vectors, we have

$$\|ab^T\|_F = tr(ba^T ab^T)^{\frac{1}{2}} = tr(b^T ba^T a)^{\frac{1}{2}} = \|a\|\|b\|$$

4.1.3 Spectral Radius

Definition 4.3 (Spectral Radius)

For $n \times n$ matrix A, **spectral radius** is the maximum eigenvalue.

$$S(A) = \max_{i=1,\dots,n} |\lambda_i|$$

Proposition 4.2

$S(A) \leq \|A\|$

Proof 4.1

$\|A\| = \max_{\|x\|=1} \|Ax\| \geq \|Au\| = |\lambda|\|u\| = |\lambda|$, *where u is the corresponding eigenvector of λ for A.*

Proposition 4.3

For symmetric $A_{n\times n}$, $S(A) = \|A\|$

Proof 4.2

$\forall x \in \mathbb{R}^n$, decompose it by u_i. Since u_i are orthonormal and linearly independent. $x = \sum_{i=1}^n \alpha_i u_i$ for some $\alpha_i \in \mathbb{R}, i = 1, ..., n$. $\|x\|^2 = \sum_{i=1}^n |\alpha_i|^2$

$$\|Ax\|^2 = \|\sum_{i=1}^n \alpha_i A u_i\|^2 = \|\sum_{i=1}^n \alpha_i \lambda_i u_i\|^2 = \sum_{i=1}^n |\alpha_i|^2 |\lambda_i|^2$$
$$\leq \sum_{i=1}^n |\alpha_i|^2 S(A)^2 = S(A)^2 \Rightarrow \|A\| \leq S(A)$$

Since we proved $S(A) \leq \|A\|$ before, $S(A) = \|A\|$.

4.1.4 Euclidean distance, inner product

Euclidean distance on $\mathbb{R}^n$:

$$|x - y| = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2}$$

Euclidean inner product:

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n = x^T y$$

Also written as $< x, y >$

Useful fact:

$$< x, y >= \cos(\theta) \|x\|_2 \|y\|_2$$

θ is the angle between x and y.

Two important results for Euchidean norm:

1) Pythagorean Theorem: If $x^\top y = 0$,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2$$

2) Cauchy - Schwarz Inequality:

$$< x, y >= \left| x^\top y \right| \leqslant \|x\|_2 \|y\|_2$$

$$" = " \text{ iff } x = \alpha y \text{ for some } \alpha \in \mathbb{R}$$

4.2 Inequalities

Proposition 4.4 (Expectation Inequality)

For any random variable Y with $\|Y\| < \infty$, we have

$$\|\mathbb{E}[Y]\| \leq \mathbb{E}\|Y\|$$

Proposition 4.5 (Cauchy - Schwarz Inequality)

For any random variables X and Y,

$$\mathbb{E}[\|X^T Y\|] \leq \mathbb{E}[\|X\|^2]^{\frac{1}{2}}\mathbb{E}[\|Y\|^2]^{\frac{1}{2}}$$

4.3 General Inner Products

4.3.1 Inner Product

Definition 4.4

An **inner product** $*$ is a function that maps two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ to a single value $\vec{x} * \vec{y} \in \mathbb{R}$, satisfying the following axioms:

1. **Bilinear** (linearity in both arguments): for all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$, we have

$$(a\vec{x} + b\vec{y}) * \vec{z} = a(\vec{x} * \vec{z}) + b(\vec{x}\vec{y} * \vec{z})$$

$$\vec{x} * (a\vec{y} + b\vec{z}) = a(\vec{x} * \vec{y}) + b(\vec{x} * \vec{z})$$

2. **Symmetric** i.e. for all $\vec{x}, \vec{y} \in \mathbb{R}^n$,

$$\vec{x} * \vec{y} = \vec{y} * \vec{x}$$

3. **Positivity** i.e. for all $\vec{x} \in \mathbb{R}^n$,

$$\vec{x} * \vec{x} \geq 0$$

   with equality if and only if $\vec{x} = \vec{0}$

Definition 4.5

Every inner product $*$ defines a corresponding norm $\|\cdot\|_*$ as $\|\vec{x}\|_* = \sqrt{\vec{x} * \vec{x}}$

4.3.2 Theorem: $*$ is inner product iff $\vec{x} * \vec{y} = \vec{x}^T H \vec{y}$ for some symmetric H

Theorem 4.1

An operation $* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an inner product if and only if it can be written as $\vec{x} * \vec{y} = \vec{x}^T H \vec{y}$ for some symmetric positive definite $n \times n$ matrix H.

Proof 4.3

Define $H := [H_{ij}] = [\vec{e}_i * \vec{e}_j]$

1. *Bilinear:*

$$
\begin{aligned}
\vec{x} * \vec{y} &= \left(\sum_{i=1}^{n} x_i \vec{e}_i\right) * \left(\sum_{j=1}^{n} y_j \vec{e}_j\right) \\
&= \sum_{i=1}^{n} x_i \left(\vec{e}_i * \sum_{j=1}^{n} y_j \vec{e}_j\right) \quad \text{(by linearity)} \\
&= \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} y_j (\vec{e}_i * \vec{e}_j)\right) \quad \text{(by linearity)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \left(\vec{e}_i * \vec{e}_j\right) y_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i H_{ij} y_j = \vec{x}^T H \vec{y}
\end{aligned}
$$

2. *Symmetric $\Leftrightarrow H = H^T$:*

$$\vec{x} * \vec{y} = \vec{x}^T H \vec{y} = \left(\vec{x}^T H \vec{y}\right)^T = \vec{y}^T H^T \vec{x} = \vec{y}^T H \vec{x} = \vec{y} * \vec{x}$$

3. *Positivity $\Leftrightarrow H \succ 0$: $\vec{x}^T H \vec{x} \geq 0$ with equality only if $\vec{x} = 0$*

As we know that a symmetric matrix H is positive definite if and only if we can write $H = B^T B$ for some invertible matrix B.

$$\vec{x} * \vec{y} = \vec{x}^T H \vec{y} = \vec{x}^T B^T B \vec{y} = (B\vec{x})^T B \vec{y} = (B\vec{x}) \cdot (B\vec{y})$$

Definition 4.6

Given a positive definite matrix H, let the associated inner product be $\vec{x} \cdot_H \vec{y} = \vec{x}^T H \vec{y}$ and the associated norm be $\|\vec{x}\|_H = \sqrt{\vec{x}^T H \vec{x}}$

4.4 Isometry

An **isometry** of $\mathbb{R}^n$ is a bijection $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ that preserves distance, which means,

$$|\Phi(x) - \Phi(y)| = |x - y|, \; \forall x, y \in \mathbb{R}^n$$

We use $\text{Isom}(\mathbb{R}^n)$ denotes the set of all isometries of $\mathbb{R}^n$,

$$\text{Isom}(\mathbb{R}^n) = \{\Phi : \mathbb{R}^n \to \mathbb{R}^n || \Phi(x) - \Phi(y)| = |x - y|,\ \forall x, y \in \mathbb{R}^n\}$$

Proposition 4.6

$\Phi, \Psi \in \text{Isom}(\mathbb{R}^n)$, *then* $\Phi \circ \Psi, \Phi^{-1} \in \text{Isom}(\mathbb{R}^n)$

Proof 4.4

Since Φ, Ψ *are bijections, so is* $\Phi \circ \Psi$*. Moreover,*

$$|\Phi \circ \Psi(x) - \Phi \circ \Psi(y)| = |\Phi(\Psi(x)) - \Phi(\Psi(y))| = |\Psi(x) - \Psi(y)| = |x - y|$$

Since $id \in \text{Isom}(\mathbb{R}^n)$*,*

$$|x - y| = |id(x) - id(y)| = |\Phi \circ \Phi^{-1}(x) - \Phi \circ \Phi^{-1}(y)| = |\Phi^{-1}(x) - \Phi^{-1}(y)|$$

4.5 Linear isometries i.e. orthogonal group

There is a matrix $A \in GL(n, \mathbb{R})$ i.e. a *invertible linear transofrmations* $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T_A(v) = Av$.

$$T_A(v) \cdot T_A(w) = (Av) \cdot (Aw) = (Av)^t (Aw) = v^t A^t A w$$

$$A^t A = I \Leftrightarrow T_A(v) \cdot T_A(w) = v \cdot \Leftrightarrow T_A \in \text{Isom}(\mathbb{R}^n)$$

We define the all isometries in *invertible linear transofrmations* $\mathbb{R}^n \to \mathbb{R}^n$ as **orthogonal group**

$$O(n) = \{A \in GL(n, \mathbb{R}) | A^t A = I\} \subset GL(n, \mathbb{R})$$

4.6 Special orthogonal group

$O(n)$ are the matrices representing linear isometries of $\mathbb{R}^n$. $1 = det(I) = det(A^t A) = det(A^t) det(A) = det(A)^2 \Rightarrow det(A) = 1$ or $det(A) = -1$. We use **special orthogonal group** represents A with $det(A) = 1$,

$$SO(n) = \{A \in O(n) | det(A) = 1\}$$

4.7 Translation

Define a *translation* by $v \in \mathbb{R}^n$,

$$\tau_v : \mathbb{R}^n \to \mathbb{R}^n,\ \tau_v(x) = x + v$$

Note *[Exercise 2.5.3]* $\forall v \in \mathbb{R}^n, \tau_v$ *is an isometry.*

Proof 4.5

$|\tau_v(x) - \tau_v(y)| = |(x+v) - (y+v)| = |x-y|$

4.8 All isometries can be represented by a composition of *a translation* and *an orthogonal transformation*

Since *the composition of isometries is an isometry,* $\forall A \in O(n)$ and $v \in \mathbb{R}^n$, the composition

$$\Phi_{A,v}(x) = \tau_v(T_A(x)) = Ax + v$$

is an isometry. **which could account for all isometries**.

Theorem 4.2

$\mathrm{Isom}(\mathbb{R}^n) = \{\Phi_{A,v} | A \in O(n), v \in \mathbb{R}^n\}$

Chapter 5 Algebra Computation

5.1 Hessian Matrix

Definition 5.1

The Hessian of f at point x is an $n \times n$ symmetric matrix denoted by $\nabla^2 f(x)$ with $[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ ♣

5.2 Taylor's Expansion

Definition 5.2 (Taylor's Expansion of Vector)

$$f(y) - f(x) = \nabla f(x)^T (y-x) + \frac{1}{2}(x-y)^T \nabla^2 f(x)(x-y) + o(\|x-y\|^2)$$

♣

5.3 Random Vectors and Random Matrices

5.3.1 Mean

Definition 5.3 (Mean of a random vector)

The mean of a $d-$dimensional random vector $\vec{x}$ is

$$\mathbb{E}(\vec{x}) = \begin{pmatrix} \mathbb{E}(x_1) \\ \mathbb{E}(x_2) \\ \cdots \\ \mathbb{E}(x_d) \end{pmatrix}$$

♣

Definition 5.4 (Mean of a random matrix)

The mean of a $d_1 \times d_2$ matrix with random entries $\vec{X}$ is

$$\mathbb{E}(\vec{X}) = \begin{pmatrix} \mathbb{E}(\vec{X}[1,1]) & \mathbb{E}(\vec{X}[1,2]) & \cdots & \mathbb{E}(\vec{X}[1,d_2]) \\ \mathbb{E}(\vec{X}[2,1]) & \mathbb{E}(\vec{X}[2,2]) & \cdots & \mathbb{E}(\vec{X}[2,d_2]) \\ \cdots & \cdots & \cdots & \cdots \\ \mathbb{E}(\vec{X}[d_1,1]) & \mathbb{E}(\vec{X}[d_1,2]) & \cdots & \mathbb{E}(\vec{X}[d_1,d_2]) \end{pmatrix}$$

♣

Lemma 5.1 (Linearity of expectation for random vectors and matrices)

Let $\vec{x}$ be a $d-$dimensional random vector and $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times d}$, then

$$\mathbb{E}(A\vec{x} + b) = A\mathbb{E}(\vec{x}) + b$$

Similarly let, $\vec{X}$ be a $d_1 \times d_2$ random matrix and $B \in \mathbb{R}^{m \times d_2}$ and $A \in \mathbb{R}^{m \times d_1}$, then

$$\mathbb{E}(A\vec{X} + B) = A\mathbb{E}(\vec{X}) + B$$

♡

Definition 5.5 (Sample mean of multivariate data)

Let $X := \{x_1, x_2, ..., x_n\}$ denote a set of $d-$dimensional vectors of real-valued data. The sample mean is the entry-wise average

$$\mu_X := \frac{\sum_{i=1}^n x_i}{n}$$

♣

5.3.2 Variance, Covariance

Definition 5.6 (Covariance matrix of a vector)

The covariance matrix of a $d-$dimensional random vector $\vec{x}$ is the $d \times d$ matrix

$$\Sigma_{\vec{x}} = \mathbb{E}[(\vec{x} - \mathbb{E}(\vec{x}))^T(\vec{x} - \mathbb{E}(\vec{x}))] = \begin{bmatrix} \text{Var}(\vec{x}[1]) & \cdots & \text{Cov}(\vec{x}[1], \vec{x}[d]) \\ \cdots & \cdots & \cdots \\ \text{Cov}(\vec{x}[d], \vec{x}[1]) & \cdots & \text{Var}(\vec{x}[d]) \end{bmatrix}$$

♣

Lemma 5.2

For any random vector $\vec{x}$ with covariance matrix $\Sigma_{\vec{x}}$, and any vector v:

$$\text{Var}(v^T\vec{x}) = v^T\Sigma_{\vec{x}}v$$

♡

Definition 5.7 (Sample covariance matrix)

Let $X := \{x_1, x_2, ..., x_n\}$ denote a set of $d-$dimensional vectors of real-valued data. The sample covariance matrix equals

♣

Variance-Covariance matrix Σ:

$$\Sigma_{m \times m} = Cov(\mathbf{Z}) = \mathbb{E}((\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^T) = \begin{bmatrix} Var(Z_1) & \cdots & Cov(Z_1, Z_m) \\ \cdots & \cdots & \cdots \\ Cov(Z_m, Z_1) & \cdots & Var(Z_m) \end{bmatrix}$$

Affine Transformation

(1)

$$\mathbf{W} = \mathbf{a}_{n \times 1} + \mathbf{B}_{n \times m}\mathbf{Z}_{m \times 1}$$

$$\mathbb{E}(\mathbf{W}) = \mathbf{a} + \mathbf{B}\mu,\ Cov(\mathbf{W}) = \mathbf{B}\Sigma\mathbf{B}^T$$

(2)

$$\mathbf{W} = \mathbf{v}^T\mathbf{Z} = v_1 Z_1 + ... + v_m Z_m$$

$$\mathbb{E}(\mathbf{W}) = \mathbf{v}^T\mu = \sum_{i=1}^{m} v_i\mu_i$$

$$Var(\mathbf{W}) = \mathbf{v}^T\Sigma\mathbf{v} = \sum_{i=1}^{m} v_i^2 Var(Z_i) + 2\sum_{i<j} v_i v_j Cov(Z_i, Z_j)$$

$$\text{i.e. } \mathbb{E}(\mathbf{AZ}) = \mathbf{A}\mathbb{E}(Z);\ Var(\mathbf{AZ}) = \mathbf{A}Var(\mathbf{Z})\mathbf{A}^T$$

(3)

$$Cov(\mathbf{AX}, \mathbf{BY}) = \mathbb{E}[(\mathbf{AX} - \mathbf{A}\mathbb{E}(X))(\mathbf{BY} - \mathbf{B}\mathbb{E}(Y))^T] = \mathbf{A}\mathbb{E}[(\mathbf{X} - \mathbb{E}(X))(\mathbf{Y} - \mathbb{E}(Y))^T]\mathbf{B}^T = \mathbf{A}Cov(\mathbf{X}, \mathbf{Y})\mathbf{B}^T$$

5.4 Matrix Multiplication

(1). $A(BC) = (AB)C$.

(2). $A(B + C) = AB + AC$.

(2). $(B + C)A = BA + CA$.

(3). No commutative: $AB \neq BA$.

5.5 Matrix Derivation

$$\frac{\partial x^T Q x}{\partial x} = 2Qx$$

https://zhuanlan.zhihu.com/p/24709748

https://blog.csdn.net/daaikuaichuan/article/details/80620518

Vector by vector:

Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{y}$ and $\mathbf{x}^T$	Denominator layout, i.e. by $\mathbf{y}^T$ and $\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$v = v(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial v\mathbf{u}}{\partial \mathbf{x}} =$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial v}{\partial \mathbf{x}}$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}\mathbf{u}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

Figure 5.1: Denominator layout means $x \in \mathbb{R}^{n\times 1}$

$$
\begin{aligned}
\frac{\partial u}{\partial x^T} &= (\frac{\partial u^T}{\partial x})^T \\
\frac{\partial u^T v}{\partial x} &= \frac{\partial u^T}{\partial x} v + \frac{\partial v^T}{\partial x} u^T \\
\frac{\partial u v^T}{\partial x} &= \frac{\partial u}{\partial x} v^T + u \frac{\partial v^T}{\partial x} \\
\frac{\partial x^T x}{\partial x} &= 2x \\
\frac{\partial x^T A x}{\partial x} &= (A + A^T)x
\end{aligned}
$$

where $x, u, v \in \mathbb{R}^{n\times 1}$

Note:

$$\frac{d\|Aw-b\|^2}{dw} = \frac{d(Aw-b)^T(Aw-b)}{dw} = \frac{d(Aw-b)^T}{dw}(Aw-b) + \frac{d(Aw-b)^T}{dw}(Aw-b) = 2A^T(Aw-b)$$

Matrix by vector:

$$\frac{\partial AB}{\partial x} = \frac{\partial A}{\partial x} B + A \frac{\partial B}{\partial x}$$

Matrix by matrix:

$$
\begin{aligned}
\frac{\partial u^T X v}{\partial X} &= u v^T \\
\frac{\partial u^T X^T X u}{\partial X} &= 2 X u u^T \\
\frac{\partial [(Xu - v)^T (Xu - v)]}{\partial X} &= 2(Xu - v) u^T
\end{aligned}
$$

Trace:

$$tr(a) = a$$

$$tr(AB) = tr(BA)$$

$$tr(ABC) = tr(CAB) = tr(BCA)$$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$$tr(A) = tr(A^T)$$

$$\frac{\partial tr(ABA^TC)}{\partial A} = CAB + C^TAB^T$$

5.6 Inverse Matrix

Proposition 5.1 (Inverse of 2×2 Matrix)

The inverse of 2×2 matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Proposition 5.2 (Woodbury matrix identity)

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{VA}^{-1},$$

where $\boldsymbol{A}, \boldsymbol{U}, \boldsymbol{C}$ and $\boldsymbol{V}$ are conformable matrices: $\boldsymbol{A}$ is $n \times n$, $\boldsymbol{C}$ is $k \times k$, $\boldsymbol{U}$ is $n \times k$, and $\boldsymbol{V}$ is $k \times n$.

This can be derived using blockwise matrix inversion.

Proposition 5.3 (Blockwise inversion)

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B}\right)^{-1} \\ -\left(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CA}^{-1} & \left(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}$$

Henderson, H. V., & Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review*, 23(1), 53-60. In this paper, they also provide a vector case.

Corollary 5.1 (Vector Case)

$$(\boldsymbol{A}+b\boldsymbol{uv}')^{-1} = \boldsymbol{A}^{-1} - \frac{b}{1+b\boldsymbol{v}'\boldsymbol{A}^{-1}\boldsymbol{u}}\boldsymbol{A}^{-1}\boldsymbol{uv}'\boldsymbol{A}^{-1}$$

where b is constant and $\boldsymbol{u}, \boldsymbol{v}$ are column vectors.

5.6.1 Block Matrix Determinant

If A is invertible,

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

5.7 Linear Regression: Least Square

Minimize$_w \mathcal{R}(w) = \|Xw - y\|^2$

5.7.1 Normal Equations

$$\nabla_w \|Xw - y\|^2 = 2X^T(Xw - y) = 0$$

$$\Rightarrow X^T X w = X^T y$$

These are called the **normal equations**.

Proposition 5.4

$\hat{w}$ satisfies $\mathcal{R}(\hat{w}) = \min_w \mathcal{R}(w)$ if and only if $\hat{w}$ satisfies the normal equations. (i.e. prove its is the global minimum)

Proof 5.1

Consider $\boldsymbol{w}$ with $\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{w} = \boldsymbol{X}^\top \boldsymbol{y}$, and any $\boldsymbol{w}'$; then

$$\begin{aligned} \left\|\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{y}\right\|^2 &= \left\|\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{X}\boldsymbol{w} + \boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\right\|^2 \\ &= \left\|\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{X}\boldsymbol{w}\right\|^2 + 2\left(\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{X}\boldsymbol{w}\right)^\top \left(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\right) + \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|^2 \end{aligned}$$

Since

$$\left(\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{X}\boldsymbol{w}\right)^\top \left(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\right) = \left(\boldsymbol{w}' - \boldsymbol{w}\right)^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^\top \boldsymbol{y}\right) = 0$$

then

$$\left\|\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{y}\right\|^2 = \left\|\boldsymbol{X}\boldsymbol{w}' - \boldsymbol{X}\boldsymbol{w}\right\|^2 + \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|^2 \geq \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|^2$$

Chapter 6 Matrix Decomposition

6.1 Cholesky Decomposition

Suppose $\boldsymbol{A}$ is a real matrix with symmetric positive-definite. $\boldsymbol{A}$ can be decomposed into the product of two matrices, $\boldsymbol{L}$ and $\boldsymbol{L}^T$,

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$$

where $\boldsymbol{L}$ is a real lower triangular matrix with positive diagonal entries.

6.2 LU Decomposition (Restricted to Square)

Triangular matrix saves time when computing $Ax = b$.

Let A be a square matrix. An LU factorization refers to the factorization of A, with proper row and/or column orderings or permutations, into two factors - a lower triangular matrix L and an upper triangular matrix U: $A = LU$. In the lower triangular matrix all elements above the diagonal are zero, in the upper triangular matrix, all the elements below the diagonal are zero. For example, for a 3×3 matrix A, its LU decomposition looks like this:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}.$$

$$A = PLU$$

P is a permutation matrix (used to swap row, only one 1 in every row). P is orthogonal, so $P^{-1} = P^T$.

Solve $Ax = b$:

$$Ax = b$$

$$PLUx = b$$

Let $y = Ux$, then solve PLy=b

$$Ly = P^T b$$

Complexity: $O(n^3)$

6.3 SVD: Singular Value Decomposition

For a $n \times m$ matrix A with rank r,

$$\begin{aligned} A_{n\times m} =& U_{n\times n}\Sigma_{n\times m}V^T_{m\times m} \\ =& \sum_{i=1}^{r} s_i u_i v_i^T \\ =& \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} s_1 & & & & & & \\ & s_2 & & & & & \\ & & \ddots & & & & \\ & & & s_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & \cdots & | \end{bmatrix}^T \end{aligned}$$

U, V are orthogonal matrices. $u_i \in \mathbb{R}^{n\times 1}$ are left singular vectors, $v_i \in \mathbb{R}^{m\times 1}$ are right singular vectors. $s_i, i = 1, ..., r$ are singular values (absolute values of eigenvalues of a normal matrix).

Complexity:$O(mn^2 + n^3)$

6.3.1 Pseudo-inverse

We can't compute the inverse matrix of a singular matrix. We can use pseudo-inverse matrix.

$$A^+_{m\times n} = \sum_{i=1}^{r} \frac{1}{s_i} v_i u_i^T = V\Sigma^+U^T$$

Where

$$\Sigma^+ = \begin{bmatrix} \frac{1}{s_1} & & & & & & \\ & \frac{1}{s_2} & & & & & \\ & & \ddots & & & & \\ & & & \frac{1}{s_r} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

The SVD may not be unique, but the pseudo-inverse of A, A^+ is unique.

$$AA^+ = \sum_{i=1}^{r} u_i u_i^T = \begin{bmatrix} I_{r\times r} & O_{r\times n-r} \\ O_{n-r\times r} & O_{n-r\times n-r} \end{bmatrix}_{n\times n}$$

$$A^+A = \sum_{i=1}^{r} v_i v_i^T = \begin{bmatrix} I_{r\times r} & O_{r\times m-r} \\ O_{m-r\times r} & O_{m-r\times m-r} \end{bmatrix}_{m\times m}$$

If A^{-1} exists, $A^{-1} = A^+$.

6.3.2 Analysis of A^TA and AA^T

$$\begin{aligned} A^TA &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T\Sigma V^T \\ &= V\Sigma^2 V^T \\ \Rightarrow V &= A^T U\Sigma^+ \end{aligned}$$

Columns of V are the eigenvectors of A^TA.

The diagonal entries of Σ^2, $s_1^2, s_2^2, ..., s_r^2$ are the eigenvalues of A^TA.

Similarly:

$$AA^T = U\Sigma^2U^T$$

$$\Rightarrow U = AV\Sigma^+$$

Columns of U are the eigenvectors of AA^T.

Fact: A^TA is positive semi-definite.

6.3.3 Solve Normal Equations

Solve $X^TXw = X^Ty$,

$$\hat{w}_{ols} = X^+y$$

$$X^TX\hat{w}_{ols} = X^TXX^+y = (X^T(XX^+))y = X^Ty$$

6.3.4 Low-Rank Approximation

For a $n\times m$ matrix A with rank r, $A = \sum_{i=1}^{r} s_i u_i v_u^T$.

Rank-k approximation for A is

$$A_k = \sum_{i=1}^{k} s_i u_i v_u^T$$

Where $s_1 \geq s_2 \geq \cdots \geq 0$

Bibliography

[1] MATH 417: Christopher J Leininger Introduction to Abstract Algebra (Draft) 2017.

[2] MATH 484

[3] ECE 490

[4] STAT 425

[5] CS/MATH 357