Week 1B
Vector and Matrix

Christopher Ting

http://cting.x10host.com/

Hiroshima University

✉: cting@hiroshima-u.ac.jp
☎: +81 082-424-6451
: A1棟 131-1

Table of Contents

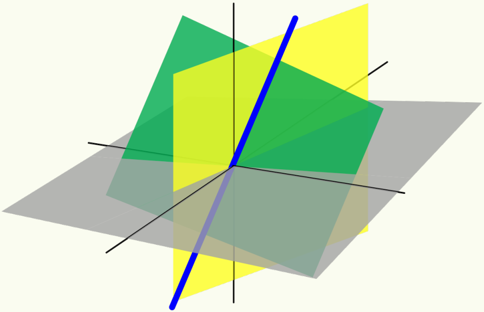
Inspiring Motivation

Application problems
without **tools** are lame,
and **tools**
without **application problems** are vain.

Learning Outcomes

- Make a strong case for the justification of learning linear algebra.
- Recall the constructs of vector and different types of matrices, including square matrix, symmetric & skew matrices, triangular matrix, commutative matrix, null matrix, and identity matrix.
- Apply matrix operations (addition, multiplication, inverse, transpose, trace) to solve problems.
- Prove the validity of a mathematical statement by the method of induction, as well as proof by contradiction.
- Explain the motivation for partitioning the matrix into sub-matrices.

Why Learn Linear Algebra?



- For the field of machine learning, it is a foundation.
- Some algorithms need linear algebra to express the idea.
- Programming languages (Matlab, Python etc) are based on vector and matrix structures, operations, and expressions.

Some Concrete Examples

Source: **10 Examples of Linear Algebra in Machine Learning**

- Dataset and Data Files
- Images and Photographs
- One Hot Encoding
- Linear Regression
- Regularization
- Principal Component Analysis
- Singular-Value Decomposition
- Latent Semantic Analysis
- Recommender Systems
- Deep Learning

Definitions of Vector and Row Vector

Definition 2.1 (p-Dimensional Column Vector).

Given p numbers, $a_1, a_2, \ldots, a_p$, a column vector $\boldsymbol{a}$ is an array (ordered list) of p rows where each row i is occupied by a_i. That is,

$$\boldsymbol{a} := \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}.$$

Definition 2.2 (q-Dimensional Row Vector).

Given q numbers, $a_1, a_2, \ldots, a_q$, a row vector $\boldsymbol{a}'$ is an array (ordered list) of q columns where each column j is occupied by a_j. That is,

$$\boldsymbol{a}' := \begin{bmatrix} a_1 & a_2 & \cdots & a_q \end{bmatrix}.$$

Scalars and Vectors

- A **scalar** a is a single number—just an ordinary number. It may be considered as a vector of one row and one column.
  - $\mathfrak{N}$: Natural numbers
  - $\mathfrak{Z}$: Integers
  - $\mathfrak{Q}$: Rational numbers
  - $\mathfrak{R}$: Real numbers
  - $\mathfrak{C}$: Complex numbers
- By default, a **vector** $\boldsymbol{a}$ is a list of p scalars arranged as a column.

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \in \mathfrak{R}^p.$$

Notations

- Vectors are usually denoted by small letters in bold font, such as

$$\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots.$$

- Suppose we have a collection of vectors. Let $i = 1, 2, ...$, and $j = 1, 2, \ldots$. From the definitions, row vector $\boldsymbol{a}_i'$ and vector $\boldsymbol{a}_j$ are expressed as

$$\boldsymbol{a}_i' := \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{iq} \end{bmatrix} \qquad (q\text{-dimensional row vector}),$$

$$\boldsymbol{a}_j := \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{bmatrix} \qquad (p\text{-dimensional column vector}).$$

Definitions of Matrix by Vectors

- Consider a set of p row vectors $\boldsymbol{a}_i'$, each having q columns. A $p \times q$ matrix $\boldsymbol{A}$ is defined as a column vector of p rows of $\boldsymbol{a}_i'$.

$$\boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix}.$$

- Consider a set of q column vectors $\boldsymbol{a}_j$, each having p rows. A $p \times q$ matrix $\boldsymbol{A}$ is defined as a row vector of q columns of $\boldsymbol{a}_j$.

$$\boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_q \end{bmatrix}.$$

Matrix 行列

- Effectively, a **matrix** $\boldsymbol{A}$ consists of $p \times q$ numbers a_{ij}, where $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots q$. These numbers are arranged as a rectangle. That is,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix}.$$

- Shorthand notations: $\boldsymbol{A}$, $[a_{ij}]$, or $[a_{ij}]_{p\times q}$.

- The **dimension** of matrix $\boldsymbol{A}$ is p by q.

Null Vector and Null Matrix

Definition 2.3 (Null Vector).

When all the numbers in a vector are zero, it is called the null vector and denoted by $\mathbf{0}$.

Definition 2.4 (Null Matrix).

When all the numbers in a matrix are zero, it is called the null matrix and denoted by $\mathbf{0}$, $0_{p\times q}$, $\left[0_{ij}\right]$, etc.

Matrix Operation: Transpose

Definition 2.5 (Transpose of a Matrix).

The transpose of a matrix is the matrix obtained by rotating each row into a column (clockwise) with the first element as pivot, i.e., for each row i:

$$\begin{bmatrix} \boxed{a_{i1}} & a_{i2} & \cdots & a_{iq} \end{bmatrix} \curvearrowright \begin{bmatrix} \boxed{a_{1i}} \\ a_{2i} \\ \vdots \\ a_{qi} \end{bmatrix}.$$

Let the symbol $^\top$ or $'$ indicate the transpose operation. Then

$$[a_{ij}]^\top \equiv [a_{ij}]' = [a_{ji}].$$

Illustration of Transpose

$$A = \begin{bmatrix} \boxed{\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{matrix}} \end{bmatrix}$$

$$A^\top = \begin{bmatrix} \boxed{\begin{matrix} 1 & 6 & 11 & 16 \\ 2 & 7 & 12 & 17 \\ 3 & 8 & 13 & 18 \\ 4 & 9 & 14 & 19 \\ 5 & 10 & 15 & 20 \end{matrix}} \end{bmatrix}$$

Properties of Transpose

- The transpose of a p by q matrix $\boldsymbol{A}$ is a q by p matrix.

Proof.

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_q \end{bmatrix}$. Then, given that each $\boldsymbol{a}_i$ is a vector of p rows,

$$\boldsymbol{A}^\top = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_q' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{1p} \\ a_{12} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{bmatrix} = \begin{bmatrix} a_{ji} \end{bmatrix}_{q \times p}.$$

$\square$

- The transpose of a transposed matrix $\boldsymbol{A}^\top$ is the matrix $\boldsymbol{A}$ itself, i.e.,

$$\left(\boldsymbol{A}^\top\right)^\top = \boldsymbol{A}.$$

Addition and Subtraction

Definition 3.1 (Matrix Addition and Subtraction).

Consider two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ that are both p by q. The addition and subtraction are defined as follows:

$$\boldsymbol{A} \pm \boldsymbol{B} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1q} \pm b_{1q} \\ a_{21} \pm b_{21} & a_{22} + b_{22} & \cdots & a_{2q} + b_{2q} \\ \vdots & \vdots & & \vdots \\ a_{p1} \pm b_{p1} & a_{p2} \pm b_{p2} & \cdots & a_{pq} \pm b_{pq} \end{bmatrix}.$$

Multiplication by a Scalar

Definition 3.2 (Scalar Multiplication).

Given a scalar c, a matrix $\boldsymbol{A}$ is said to be scaled by c times in that each number in $\boldsymbol{A}$ is multiplied by c. That is,

$$c\boldsymbol{A} = \left[ca_{ij}\right]_{p\times q}.$$

- Note that

$$c\boldsymbol{A} = \boldsymbol{A}c.$$

- Proof:

$$c\boldsymbol{A} = \left[ca_{ij}\right]_{p\times q} = \left[a_{ij}c\right]_{p\times q} = \boldsymbol{A}c.$$

$\square$

Theorem 3.3

Theorem 3.3 (Basic Laws of Matrix Addition and Scalar Multiplication).

For any $p \times q$ *matrices* $\boldsymbol{A}, \boldsymbol{B}$, *and* $\boldsymbol{C}$, *as well as scalars* a *and* b, *the following are true:*

1. *Commutativity* $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$
2. *Associativity* $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$
3. *Identity* $\boldsymbol{A} + \boldsymbol{0} = \boldsymbol{0} + \boldsymbol{A} = \boldsymbol{A}$
4. *Inverse* $\boldsymbol{A} + (-\boldsymbol{A}) = (-\boldsymbol{A}) + \boldsymbol{A} = \boldsymbol{0}$
5. *Commutativity of Scalar Multiplication* $a(b\boldsymbol{A}) = (ab)\boldsymbol{A}$
6. *Distributivity of Scalar Multiplication 1* $(a + b)\boldsymbol{A} = a\boldsymbol{A} + b\boldsymbol{A}$
7. *Distributivity of Scalar Multiplication 2* $a(\boldsymbol{A} + \boldsymbol{B}) = a\boldsymbol{A} + a\boldsymbol{B}$
8. $0\boldsymbol{A} = \boldsymbol{0}$
9. $1\boldsymbol{A} = \boldsymbol{A}$

Matrix Multiplication

Definition 3.4 (Matrix Multiplication).

Consider two matrices $\boldsymbol{A} = \left[a_{ij}\right]_{p\times n}$ and $\boldsymbol{B} = \left[b_{jk}\right]_{n\times q}$. Note that $\boldsymbol{A}$'s number of columns is equal to $\boldsymbol{B}$'s number of rows. The product $\boldsymbol{C} = \boldsymbol{AB}$ is then a $p \times q$ matrix, where

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk},$$

for $i = 1, 2, \ldots, p$ and $k = 1, 2, \ldots q$.

Essentially matrix multiplication is a series of "row times column".

Row Times Column

$$\begin{array}{cc} & \boxed{j\text{-th column}} \\ \boxed{i\text{-th row}} & \begin{pmatrix} a_{11} & \cdots & a_{1h} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ih} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{p1} & \cdots & a_{ph} & \cdots & a_{pn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1q} \\ \vdots & & \vdots & & \vdots \\ b_{h1} & \cdots & b_{hj} & \cdots & b_{hq} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nq} \end{pmatrix} \end{array}$$

- Both the i-th row of matrix $\boldsymbol{A}$ and the j-th column of matrix $\boldsymbol{B}$ has the same number of entries. That number is n. Hence, $c_{ij} = \sum_{h=1}^{n} a_{ih} b_{hj}$.

Dimensions after Multiplication

$\boxed{\boldsymbol{B} : n \text{ rows } q \text{ columns}}$

$$
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1n} \\
a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1} & a_{p2} & \dots & a_{pn}
\end{pmatrix}
\qquad
\begin{pmatrix}
b_{11} & b_{12} & \dots & b_{1q} \\
b_{21} & b_{22} & \dots & b_{2q} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n2} & \dots & b_{nq}
\end{pmatrix}
$$

$\boxed{\boldsymbol{A} : p \text{ rows } n \text{ columns}}$

$$
= \begin{pmatrix}
c_{11} & c_{12} & \dots & c_{1q} \\
c_{21} & c_{22} & \dots & c_{2q} \\
\vdots & \vdots & \ddots & \vdots \\
c_{p1} & c_{p2} & \dots & c_{pq}
\end{pmatrix}
$$

$\boxed{\boldsymbol{C} = \boldsymbol{AB} : p \text{ rows } q \text{ columns}}$

Vector Representation of Matrix Multiplication

- Write matrix A as $\begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix}$ and B as $\begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_q \end{bmatrix}$.
- The (i, j) component of the product $\boldsymbol{AB}$ is

$$\boldsymbol{a}_i'\boldsymbol{b}_j = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}.$$

- Moreover, $\boldsymbol{AB} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_q \end{bmatrix} = \begin{bmatrix} \boldsymbol{Ab}_1 & \boldsymbol{Ab}_2 & \cdots & \boldsymbol{Ab}_q \end{bmatrix}$, and

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_p' \end{bmatrix} \boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{B} \\ \boldsymbol{a}_2'\boldsymbol{B} \\ \vdots \\ \boldsymbol{a}_p'\boldsymbol{B} \end{bmatrix}.$$

Non-Commutativity

- In general, $\boldsymbol{AB} \neq \boldsymbol{BA}$.
- Example: Consider the following matrices. Calculate $\boldsymbol{AB}$ and $\boldsymbol{BA}$.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ -1 & 0 \\ -2 & 4 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{B} = \begin{bmatrix} -1 & 5 & 2 \\ 0 & 4 & 7 \end{bmatrix}$$

- Applying the rule of matrix multiplication,

$$\boldsymbol{AB} = \begin{bmatrix} -1 & 17 & 23 \\ 1 & -5 & -2 \\ 2 & 6 & 24 \end{bmatrix} \qquad \text{but} \qquad \boldsymbol{BA} = \begin{bmatrix} -10 & 5 \\ -18 & 28 \end{bmatrix}$$

Associativity, Distributivity, and Transpose

Theorem 3.5 (Associativity and Distributivity).

For any matrices A, B, and C such that addition and multiplication are feasible.

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

Theorem 3.6 (Properties of Transpose).

For any matrices A, B and scalar c such that addition and multiplication are possible, the following equations hold:

1. $(A + B)^\top = A^\top + B^\top$
2. $(cA)^\top = cA^\top$
3. $(AB)^\top = B^\top A^\top$

Proof of $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$

- Let $\boldsymbol{A} = \left[a_{ij}\right]_{p\times n}$, $\boldsymbol{B} = \left[b_{ij}\right]_{n\times q}$, and $\boldsymbol{C} := \boldsymbol{AB} = \left[c_{ij}\right]_{p\times q}$.
- By definition, $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Hence, the (i,j) element of $(\boldsymbol{AB})^\top$ is

$$c_{ji} = \sum_{k=1}^{n} a_{jk}b_{ki}.$$

- On the other hand, let $\boldsymbol{A}^\top := \left[\widetilde{a}_{ij}\right]_{n\times p}$ and $\boldsymbol{B}^\top := \left[\widetilde{b}_{ij}\right]_{q\times n}$. By the definition of transpose, we have $\widetilde{a}_{ij} = a_{ji}$ and $\widetilde{b}_{ij} = b_{ji}$.
- Let $\boldsymbol{D} = \boldsymbol{B}^\top \boldsymbol{A}^\top$. It follows that the (i,j) component of $\boldsymbol{D}$ is

$$d_{ij} = \sum_{k=1}^{n} \widetilde{b}_{ik}\widetilde{a}_{kj} = \sum_{k=1}^{n} b_{ki}a_{jk} = c_{ji}.$$

- Hence $c_{ji} = d_{ij}$, which implies that $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$. $\square$

Definitions of Square Matrix, Diagonal Matrix etc

Definition 4.1 (Square Matrix).

An n by n matrix is called the n-dimensional square matrix. The diagonal elements of the square matrix are a_{ii} where $i = 1, 2, \ldots, n$.

Definition 4.2 (Diagonal Matrix and Identity Matrix).

(1) A diagonal matrix is a square matrix for which all the non-diagonal elements are zero.

(2) An identity matrix is a diagonal matrix where all the diagonal elements are 1.

Definition 4.3 (Kronecker's Delta).

For any natural numbers i and j, the **Kronecker**'s delta is defined as

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

Standard Basis

Definition 4.4 (Standard Basis Vectors).

Consider the vector representation of n-dimensional identity matrix:

$$\boldsymbol{I}_n = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix},$$

where

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

These n vectors are called standard basis vectors.

- $\{\boldsymbol{e}_i\}_{i=1}^n$ are also called standard vectors or unit vectors.

Scalar Matrix

- With respect to any scalar c, the diagonal matrix $c\boldsymbol{I}_n$ is called the **scalar matrix**. In particular, $\mathbf{0} = 0\boldsymbol{I}_n$ and $\boldsymbol{I}_n = 1\boldsymbol{I}_n$ are scalar matrices.

Theorem 4.5 (Identity Matrix and Scalar Multiplication).

With respect to any p by q matrix $\boldsymbol{A}$, the following commutative equations hold:

$$\mathbf{0}_p\boldsymbol{A} = \boldsymbol{A}\mathbf{0}_q = \mathbf{0}_{p\times q}, \tag{1}$$

$$\boldsymbol{I}_p\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_q = \boldsymbol{A}. \tag{2}$$

In general, for any scalar c,

$$\left(c\boldsymbol{I}_p\right)\boldsymbol{A} = \boldsymbol{A}\left(c\boldsymbol{I}_q\right) = c\boldsymbol{A}. \tag{3}$$

Proof of Theorem 4.5

- We first prove (3).
- Let $\boldsymbol{I}_p = \left[\delta_{ij}\right]_{p\times p}$ and $\boldsymbol{A} = \left[a_{ij}\right]_{p\times q}$.
- From the definition of matrix multiplication, the (i,j) component of $(c\boldsymbol{I}_p)\boldsymbol{A}$ is

$$\sum_{k=1}^{p} c\delta_{ik}a_{kj} = (c\delta_{ii})a_{ij} = ca_{ij}.$$

  Hence, $(c\boldsymbol{I}_p)\boldsymbol{A} = c\boldsymbol{A}$.
- Using the same technique, we obtain $\boldsymbol{A}(c\boldsymbol{I}_q) = c\boldsymbol{A}$.
- The special case of $c = 0$ give rise to (1).
- The special case of $c = 1$ leads to (2). $\square$

Symmetric and Skew-Symmetric Matrices

Definition 4.6 (Symmetric Matrix).

A square matrix A is said to be a symmetric matrix when

$$\boldsymbol{A}^\top = \boldsymbol{A}.$$

That is, $a_{ij} = a_{ji}$.

Definition 4.7 (Skew Symmetric Matrix).

A square matrix A is said to be a skew symmetric matrix when

$$\boldsymbol{A}^\top = -\boldsymbol{A}.$$

That is, $a_{ij} = -a_{ji}$.

Triangular Matrices

Definition 4.8 (Triangular Matrices).

Given a square matrix $\boldsymbol{A} = [a_{ij}]$,

1. When $a_{ij} = 0$ for $i > j$, $\boldsymbol{A}$ is called the upper triangular matrix.
2. When $a_{ij} = 0$ for $i < j$, $\boldsymbol{A}$ is called the lower triangular matrix.

All in all, they are called triangular matrices.

Upper: $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$ Lower: $\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$

Non-Commutativity of Square Matrices

- In general, for any two square matrices, $AB \neq BA$, that is, **non-commutative**.

Example 4.9.

- Let $A = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 5 \\ 0 & 4 \end{bmatrix}$.
- We find that $AB = \begin{bmatrix} 2 & 14 \\ 3 & 11 \end{bmatrix}$, whereas $BA = \begin{bmatrix} 17 & -4 \\ 12 & -4 \end{bmatrix}$.

- But when $AB = BA$, they are said to be **commutative**.
- For any natural number m, A^m is the **power matrix** of A, which is constructed by multiplying A with itself m times.

$$A^m = \underbrace{AA \cdots A}_{m}.$$

- Obviously, A^m and A are commutative.

The Cayley-Hamilton Theorem for 2×2 Matrices

Theorem 4.10 (The Cayley-Hamilton Theorem for 2-D Matrix).

For any arbitrary 2-dimensional matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

$$\boldsymbol{A}^2 - (a+d)\boldsymbol{A} + (ad - bc)\boldsymbol{I} = \boldsymbol{0}.$$

Proof.

- $\boldsymbol{A}$ is commutative with any arbitrary scalar matrix, and thus we have
$$\big(\boldsymbol{A} - a\boldsymbol{I}\big)\big(\boldsymbol{A} - d\boldsymbol{I}\big) = \boldsymbol{A}^2 - (a+d)\boldsymbol{A} + ad\boldsymbol{I}.$$
- Now, $\boldsymbol{A} - a\boldsymbol{I} = \begin{bmatrix} 0 & b \\ c & d-a \end{bmatrix}$ and $\boldsymbol{A} - d\boldsymbol{I} = \begin{bmatrix} a-d & b \\ c & 0. \end{bmatrix}$
- It follows that $\big(\boldsymbol{A} - a\boldsymbol{I}\big)\big(\boldsymbol{A} - d\boldsymbol{I}\big) = \begin{bmatrix} 0 & b \\ c & d-a \end{bmatrix} \begin{bmatrix} a-d & b \\ c & 0. \end{bmatrix} = \begin{bmatrix} bc & 0 \\ 0 & bc \end{bmatrix} = bc\boldsymbol{I}$.
- Hence, $\boldsymbol{A}^2 - (a+d)\boldsymbol{A} + ad\boldsymbol{I} = bc\boldsymbol{I}$. □

Regular/Non-Singular/Invertible Matrices

Definition 4.11 (Regular Matrix).

Let A be a square matrix. If X exists such that

$$AX = XA = I,$$

then we say that A is **regular**, **non-singular** or **invertible**, and we write

$$X := A^{-1},$$

and define it as the **inverse matrix**.

- When A is regular, its inverse is unique.

Theorem 4.12: Properties of Inverse

Theorem 4.12 (Properties of Inverse).

Suppose the square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are regular and have the same dimension. Then the following statements hold:

(1) *If $\boldsymbol{A}^{-1}$ is regular, then $\left(\boldsymbol{A}^{-1}\right)^{-1} = \boldsymbol{A}$.*

(2) *If $\boldsymbol{A}^\top$ is regular, then $\left(\boldsymbol{A}^\top\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^\top$*

(3) *If $\boldsymbol{A}\boldsymbol{B}$ is regular, then $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$*

(4) *If $\boldsymbol{A}^h$ is regular, then $\left(\boldsymbol{A}^h\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^h$, where h is any natural number.*

Proof of Theorem 4.12

Proof of (2) $(A')^{-1} = (A^{-1})'$.

- Statement (1) is clear from the definition.
- To prove (2), with respect to $AA^{-1} = I$, perform the transpose. In view of Theorem 3.6's $(AB)' = B'A'$, we have

$$(AA^{-1})' = (A^{-1})'A' = I' = I.$$

Then multiply both sides from the right by $(A')^{-1}$ to obtain

$$(A^{-1})'A'(A')^{-1} = I(A')^{-1}.$$

Since $A'(A')^{-1} = I$, statement (2) ensues.

Proof of Theorem 4.12 (Cont'd)

Proof of (3) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ and (4) $(\boldsymbol{A}^h)^{-1} = (\boldsymbol{A}^{-1})^h$.

- To prove assertion (3), we compute and obtain $\boldsymbol{AB}(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{A}(\boldsymbol{BB}^{-1})\boldsymbol{A}^{-1} = \boldsymbol{AA}^{-1} = \boldsymbol{I}$. Likewise, we find that $(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1})\boldsymbol{AB} = \boldsymbol{I}$. Hence, we obtain $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
- We use the method of induction to prove (4), starting with $h = 1$, which obviously holds. Suppose it also holds for $h = k > 1$. We need to show that it also holds for $h = k + 1$.
- We start from $(\boldsymbol{A}^{k+1})^{-1}$. Now, let $\boldsymbol{B} = \boldsymbol{A}^k$ and substitute it into the proven (3) to obtain $(\boldsymbol{A}^{k+1})^{-1} = (\boldsymbol{AA}^k)^{-1} = (\boldsymbol{A}^k)^{-1}\boldsymbol{A}^{-1}$. Apply (3) another k times, we get $(\boldsymbol{A}^{k+1})^{-1} = \underbrace{\boldsymbol{A}^{-1}\boldsymbol{A}^{-1}\cdots\boldsymbol{A}^{-1}}_{k+1} = (\boldsymbol{A}^{-1})^{k+1}$.
- Hence, it also holds for $h = k + 1$ and by induction, statement (4) must be true. □

2-Dimensional Regular Square Matrix

Theorem 4.13.

The square matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is regular.* $\iff ad - bc \neq 0.$

Proof.

- Suppose $\boldsymbol{A}$ is regular. We want to show that it implies $ad - bc \neq 0$.
- We use the method of proof by contradiction: Given that $\boldsymbol{A}$ is regular, suppose $ad - bc = 0$.
- By Theorem 4.10, we have $\boldsymbol{A}^2 = (a+d)\boldsymbol{A}$.
- Multiplying both sides by $\boldsymbol{A}^{-1}$, we obtain $\boldsymbol{A} = (a+d)\boldsymbol{I}$, which implies that

$$\boldsymbol{A} - (a+d)\boldsymbol{I} = \begin{bmatrix} -d & b \\ c & -a \end{bmatrix} = \boldsymbol{0}, \quad \Longrightarrow \quad a = b = c = d = 0.$$

- Thus, $\boldsymbol{A} = \boldsymbol{0}$, which contradicts the supposition that $\boldsymbol{A}$ is regular.

Proof of Theorem 4.13 (Cont'd)

Proof (Cont'd).

- Next, we start with $ad - bc \neq 0$ and show that it leads to $\boldsymbol{A}$ being regular.
- Now, $\boldsymbol{A}^2 = \begin{bmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix}$ and $(a+d)\boldsymbol{A} = \begin{bmatrix} a^2 + ad & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix}$, resulting in $\boldsymbol{A}^2 - (a+d)\boldsymbol{A} = (bc - ad)\boldsymbol{I}$.
- Divide both sides by $-\dfrac{1}{ad - bc}$, we obtain

$$-\frac{1}{ad-bc}\left(\boldsymbol{A}^2 - (a+d)\boldsymbol{A}\right) = \boldsymbol{A}\left(-\frac{1}{ad-bc}\left(\boldsymbol{A} - (a+d)\boldsymbol{I}\right)\right) = \boldsymbol{I}. \tag{4}$$

- In other words, $\left(-\dfrac{1}{ad-bc}\left(\boldsymbol{A} - (a+d)\boldsymbol{I}\right)\right)$ is the inverse of $\boldsymbol{A}$, which means that $\boldsymbol{A}$ is regular. □

Inverse of a 2×2 Matrix

Corollary 4.14.

When $ad - bc \neq 0$, the inverse matrix of $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

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

Proof.

From (4),

$$\boldsymbol{A}^{-1} = \left(-\frac{1}{ad - bc} (\boldsymbol{A} - (a+d)\boldsymbol{I}) \right) = -\frac{1}{ad - bc} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} a+d & 0 \\ 0 & a+d \end{bmatrix} \right)$$
$$= -\frac{1}{ad - bc} \begin{bmatrix} -d & b \\ c & -a \end{bmatrix}.$$

Trace

Definition 4.15 (Trace of a Square Matrix).

With respect to any square matrix $\boldsymbol{A} = \left[a_{ij}\right]_{n\times n}$, the trace of $\boldsymbol{A}$ is defined as

$$\operatorname{Tr}(\boldsymbol{A}) := \sum_{i=1}^{n} a_{ii}.$$

Properties of Trace

Theorem 4.16 (Properties of Trace).

With respect to the square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and the scalar c, the following statements hold:

1. $\mathrm{Tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{Tr}(\boldsymbol{A}) + \mathrm{Tr}(\boldsymbol{B})$, $\qquad \mathrm{Tr}(c\boldsymbol{A}) = c\mathrm{Tr}(\boldsymbol{A})$
2. $\mathrm{Tr}(\boldsymbol{A}') = \mathrm{Tr}(\boldsymbol{A})$
3. $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{Tr}(\boldsymbol{B}\boldsymbol{A})$
4. $\mathrm{Tr}(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}) = \mathrm{Tr}(\boldsymbol{A})$, *where $\boldsymbol{P}$ is regular.*
5. *When $\boldsymbol{A}$ is a real-valued matrix,* $\mathrm{Tr}(\boldsymbol{A}^\top\boldsymbol{A}) \geqq 0$.

Example of Trace

Example 4.17.

Consider a matrix $\boldsymbol{A} = \begin{bmatrix} 3 & x & 2x \\ 0 & x^2 & 1 \\ 2 & 3x & 4 \end{bmatrix}$. Find all values of x such that $\mathrm{Tr}\left(\boldsymbol{A}\right) = 23$.

- The trace of a matrix is defined only on the entries of a matrix in the main diagonal. It follows that $\mathrm{Tr}\left(\boldsymbol{A}\right) = 3 + x^2 + 4$.
- Since $\mathrm{Tr}\left(\boldsymbol{A}\right) = 23$, a quadratic equation is obtained as follows:

$$23 = 3 + x^2 + 4.$$

- Upon solving the equation, we find that the solutions are -4 and 4.

Definition of Matrix Partition

Definition 5.1 (Partition of a Matrix).

Suppose $(k-1)$ horizontal lines and $(\ell-1)$ vertical lines are inserted into a $p \times q$ matrix $\boldsymbol{A}$ such that

$$\boldsymbol{A} = \left[\begin{array}{c|c|c} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1\ell} \\ \hline \vdots & \vdots & \vdots \\ \hline \boldsymbol{A}_{1k} & \cdots & \boldsymbol{A}_{k\ell} \end{array}\right], \tag{5}$$

where $\boldsymbol{A}_{ij}$ is a $p_i \times q_j$ matrix, $\sum_{i=1}^{k} p_i = p$, and $\sum_{j=1}^{\ell} q_j = q$.

The matrix $\boldsymbol{A}$ in (5) is called block partition or small matrix partition. The resulting matrix is called the partitioned matrix.

Basic Algebra of Partitioned Matrices

Theorem 5.2 (Summation and Scalar Multiplication).

Let the partitioned matrices be $\boldsymbol{A} = \left[\boldsymbol{A}_{ij}\right]_{k\times\ell}$ and $\boldsymbol{B} = \left[\boldsymbol{B}_{ij}\right]_{k\times\ell}$, such that for each $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$, the matrices $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_{ij}$ are of the same dimensions. Then the sum $\boldsymbol{A} + \boldsymbol{B}$ and scalar multiplication $c\boldsymbol{A}$ are block partitioned as, respectively,

$$\boldsymbol{A} + \boldsymbol{B} = \left[\boldsymbol{A}_{ij} + \boldsymbol{B}_{ij}\right]_{k\times\ell} \quad \text{and} \quad c\boldsymbol{A} = \left[c\boldsymbol{A}_{ij}\right]_{k\times\ell}.$$

Theorem 5.3 (Multiplication of Partitioned Matrix).

Let the partitioned matrices be $\boldsymbol{A} = \left[\boldsymbol{A}_{ij}\right]_{k\times\ell}$ and $\boldsymbol{B} = \left[\boldsymbol{B}_{ij}\right]_{k\times\ell}$, such that for each $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$, the matrices $\boldsymbol{A}_{ik}\boldsymbol{B}_{kj}$ are defined. Then $\boldsymbol{C} = \left[\boldsymbol{C}_{ij}\right]_{p\times q} := \boldsymbol{A}\boldsymbol{B}$ can be block partitioned as

$$\boldsymbol{C}_{ij} = \sum_{k=1}^{q} \boldsymbol{A}_{ik}\boldsymbol{B}_{kj}.$$

Inverse of a Partitioned Matrix: Question

- Consider two n-dimensional square matrices $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}, \end{bmatrix}$ where $\boldsymbol{A}_{11}$ and $\boldsymbol{B}_{11}$ are k-dimensional square matrices.
- Moreover, suppose $\boldsymbol{A}$ is regular and $\boldsymbol{A}^{-1} = \boldsymbol{B}$
- **What is B in terms of the sub-matrices A_{ij}?**
- Inverse calculation is computationally intensive.
- Matrix partition makes it possible for inverse calculation to be processed in parallel.

Inverse of a Partitioned Matrix: Solution

- Since $\boldsymbol{B} = \boldsymbol{A}^{-1}$,

$$\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} & \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{0}_{(n-k)\times k} & \boldsymbol{I}_{n-k}. \end{bmatrix}.$$

- We need to solve

$$\boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} = \boldsymbol{I}_k, \tag{6}$$

$$\boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} = \boldsymbol{0}_{k\times(n-k)}, \tag{7}$$

$$\boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} = \boldsymbol{0}_{(n-k)\times k}, \tag{8}$$

$$\boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} = \boldsymbol{I}_{n-k}. \tag{9}$$

Inverse of a Partitioned Matrix: Solution (つづき)

- Assume that A_{11} and A_{22} are regular. From (7) and (8), we obtain

$$B_{12} = -A_{11}^{-1}A_{12}B_{22}, \tag{10}$$

$$B_{21} = -A_{22}^{-1}A_{21}B_{11}. \tag{11}$$

- Plugging B_{21} into (6), we obtain

$$\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)B_{11} = I_k.$$

- Plugging B_{12} into (9), we get

$$\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)B_{22} = I_{n-k}.$$

Inverse of a Partitioned Matrix: Solution (つづき)

- Hence, we find

$$B_{11} = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}, \tag{12}$$

$$B_{22} = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}. \tag{13}$$

- Plugging (12) and (13) back to (10) and (11), we obtain

$$B_{12} = -A_{11}^{-1}A_{12}\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1},$$

$$B_{21} = -A_{22}^{-1}A_{21}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}.$$

- Thus we have obtained B in terms of A's partitioned matrices.

Takeaways

- Linear algebra is a mathematical foundation of AI and data science.
- Vector $\boldsymbol{a}$ provides an n-dimensional structure to a set of n numbers.
- By default, vector $\boldsymbol{a}$ is a column vector. Its transpose $\boldsymbol{a}'$ (or $\boldsymbol{a}^\top$) is a row vector.
- Matrix provides an ordered structure to a set of vectors.
- Matrix multiplication is not element-wise; it is the sum of row $\times$ column.
- Inverse is an analog of the usual division.
- Transpose and trace are unique to matrices.
- Partitioning of a matrix enables parallel processing of the computation of the inverse matrix.

Keywords