Session 4

Introduction to Matrix Algebra

Anthony Tay

This Version: 17 Sep 2025

Session 4

- Introduction to Matrix Algebra
  - Different types of matrices
  - Matrix operations
  - Partitioned matrices
  - Vectors and matrices of random variables
  - Differentiation of matrix forms

Matrix Definitions

$$\text{rows}\quad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

columns

Matrix with m rows, n columns

- **Order** or **Dimension** $m \times n$
- If $m = n$: **square matrix**
- If $m > 1$ and $n = 1$: **column vector**
- If $m = 1$ and $n > 1$: **row vector**
- If $m = 1$ and $n = 1$: **scalar**

a_{ij} is the (i, j)th **element** or **term** of the matrix

Other notational conventions:

- $(a_{ij})_{m\times n}$ refers to a $m \times n$ matrix with (i, j)th element a_{ij}
- $(A)_{ij}$ refers to the (i, j)th element of the matrix A

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

Given two matrices A and B of the same dimensions:

- Equality: $A = B \iff (A)_{ij} = (B)_{ij}$ for all $i = 1, \ldots, m$; $j = 1, \ldots, n$
- Addition: $(A + B)_{ij} = (A)_{ij} + (B)_{ij}$ for all $i = 1, \ldots, m$; $j = 1, \ldots, n$
  - Matrix addition is element-by-element addition
- Hadamard Product: $(A \odot B)_{ij} = (A)_{ij}(B)_{ij}$
  - Hadamard Product is element-by-element multiplication
  - Alternative notations for Hadamard Product: $A \circ B$, $A * B$

For any matrix A and any scalar $\alpha \in \mathbb{R}$

- Scalar Multiplication: $(\alpha A)_{ij} = (A\alpha)_{ij} = \alpha(A)_{ij}$ for all $i = 1, \ldots, m$; $j = 1, \ldots, n$

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

Examples: If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$, then

- $A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix}$
- $A \odot B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \end{bmatrix}$
- $\alpha A = A\alpha = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \end{bmatrix}$
- $A - B = A + (-1)B = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & a_{13} - b_{13} \\ a_{21} - b_{21} & a_{22} - b_{22} & a_{23} - b_{23} \end{bmatrix}$

Matrix Operations (Add, Hadamard Prod, Scalar Mult.)

The following should be obvious:

- $(A+B)+C = A+(B+C)\,,\ (A\odot B)\odot C = A\odot(B\odot C)$
- $A+B = B+A\,,\ A\odot B = B\odot A$
- $A\odot(B+C) = A\odot B + A\odot C$
- $\alpha(A+B) = \alpha A + \alpha B\,,\ (\alpha+\beta)A = \alpha A + \beta A$

Matrix Operations (Transposition)

- Matrix Transpose of A, denoted A^{T}, is defined by $(A^{\mathrm{T}})_{ij} = A_{ji}$

e.g., if $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$, then $A^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$

e.g., if $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then $x^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

We often write column vectors as $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ to save space

Sometimes a matrix transpose is written as A' instead of A^{T}

Matrix Operations (Transposition)

Clearly

- $(A+B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$
- $(A \odot B)^{\mathrm{T}} = A^{\mathrm{T}} \odot B^{\mathrm{T}}$
- $(\alpha A)^{\mathrm{T}} = \alpha A^{\mathrm{T}}$
- $(A^{\mathrm{T}})^{\mathrm{T}} = A$

Definition: A square matrix is **symmetric** if $(A)_{ij} = (A)_{ji}$, i.e., $A^{\mathrm{T}} = A$

e.g., $\begin{bmatrix} 1 & 3 & 2 \\ 3 & 4 & 6 \\ 2 & 6 & 3 \end{bmatrix}$ is symmetric, $\begin{bmatrix} 1 & 3 & 2 \\ 7 & 4 & 6 \\ 2 & 6 & 3 \end{bmatrix}$ is not

Matrix Operations (Matrix Multiplication)

Matrix Multiplication/Product: For any $m \times n$ matrix A and $n \times p$ matrix B, we have

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \ .$$

i.e., (i, j)th element of AB is the sum of the product of the elements of the ith row of A with the corresponding elements in the jth column of B

For example,

- $(AB)_{11} = \sum_{k=1}^{n} a_{1k} b_{k1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + \cdots + a_{1n}b_{n1}$
- $(AB)_{23} = \sum_{k=1}^{n} a_{2k} b_{k3} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} + \cdots + a_{2n}b_{n3}$

Matrix Operations

For the product of a 3×3 matrix and a 3×2 matrix, we have

$$\begin{bmatrix} \boxed{a_{11} \quad a_{12} \quad a_{13}} \\ a_{21} \quad a_{22} \quad a_{23} \\ a_{31} \quad a_{32} \quad a_{33} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} b_{11} \\ b_{21} \\ b_{31} \end{matrix}} & \begin{matrix} b_{12} \\ b_{22} \\ b_{32} \end{matrix} \end{bmatrix} = \begin{bmatrix} \boxed{a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}} & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} \boxed{a_{11} \quad a_{12} \quad a_{13}} \\ a_{21} \quad a_{22} \quad a_{23} \\ a_{31} \quad a_{32} \quad a_{33} \end{bmatrix} \begin{bmatrix} \begin{matrix} b_{11} \\ b_{21} \\ b_{31} \end{matrix} & \boxed{\begin{matrix} b_{12} \\ b_{22} \\ b_{32} \end{matrix}} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{3} a_{1k}b_{k1} & \boxed{a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}} \\ \bullet & \bullet \\ \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} a_{11} \quad a_{12} \quad a_{13} \\ \boxed{a_{21} \quad a_{22} \quad a_{23}} \\ a_{31} \quad a_{32} \quad a_{33} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} b_{11} \\ b_{21} \\ b_{31} \end{matrix}} & \begin{matrix} b_{12} \\ b_{22} \\ b_{32} \end{matrix} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{3} a_{1k}b_{k1} & \sum_{k=1}^{3} a_{1k}b_{k2} \\ \boxed{a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31}} & \bullet \\ \bullet & \bullet \end{bmatrix}$$

and so on.

Matrix Operations (Matrix Multiplication)

If $A = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 7 \\ 6 & 9 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 3 & 4 \\ 6 & 2 & 5 \end{bmatrix}$ then

$$AB = \begin{bmatrix} 2 & 8 \\ 3 & 0 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 6 & 9 \end{bmatrix} = \begin{bmatrix} 2 \cdot 4 + 8 \cdot 6 & 2 \cdot 7 + 8 \cdot 9 \\ 3 \cdot 4 + 0 \cdot 6 & 3 \cdot 7 + 0 \cdot 9 \\ 5 \cdot 4 + 1 \cdot 6 & 5 \cdot 7 + 1 \cdot 9 \end{bmatrix} = \begin{bmatrix} 56 & 86 \\ 12 & 21 \\ 26 & 44 \end{bmatrix}.$$

- A and B "conformable" for the product AB requires no.cols(A) = no.rows(B)
- Even if A and B are conformable for AB, the product BA might not be possible
- Even if AB and BA are possible, they may not be equal (might not even be the same dimensions)

Matrix Operation (Matrix Multiplication)

```r
A = matrix(c(2,8,3,0,5,1),
           nrow=3, byrow=T)
A
```

```
     [,1] [,2]
[1,]    2    8
[2,]    3    0
[3,]    5    1
```

```r
B = matrix(c(4,7,6,9),
           nrow=2, byrow=T)
B
```

```
     [,1] [,2]
[1,]    4    7
[2,]    6    9
```

```r
C = matrix(c(1,3,4,6,2,5),
           nrow=2, byrow=T)
C
```

```
     [,1] [,2] [,3]
[1,]    1    3    4
[2,]    6    2    5
```

```r
A %*% B
```

```
     [,1] [,2]
[1,]   56   86
[2,]   12   21
[3,]   26   44
```

```r
A %*% C
```

```
     [,1] [,2] [,3]
[1,]   50   22   48
[2,]    3    9   12
[3,]   11   17   25
```

```r
B %*% A
```

```
Error in B %*% A: non-conformable arguments
```

```r
C %*% A
```

```
     [,1] [,2]
[1,]   31   12
[2,]   43   53
```

Matrix Operations (Matrix Multiplication)

Easy to show

- $(AB)C = A(BC)$ if A is $m \times n$, B is $n \times p$ and C is $p \times q$
- $A(B + C) = AB + AC$ if A is $m \times n$, B and C are $n \times p$
- $(A + B)C = AC + BC$ if A and B are $m \times n$ and C is $n \times p$

Proof of $(AB)C = A(BC)$:

$$\begin{aligned}((AB)C)_{ij} &= \sum_{k=1}^{p} ((AB))_{ik}(C)_{kj} = \sum_{k=1}^{p} \left(\sum_{r=1}^{n} (A)_{ir}(B)_{rk} \right) (C)_{kj} \\ &= \sum_{r=1}^{n} (A)_{ir} \left(\sum_{k=1}^{p} (B)_{rk}(C)_{kj} \right) = \sum_{r=1}^{n} (A)_{ir}(BC)_{rj} = (A(BC))_{ij}\end{aligned}$$

Matrix Operations

In some ways, matrix multiplication behaves differently from multiplication of numbers:

e.g., For numbers, $ab = 0 \Rightarrow a = 0$ or $b = 0$

Not necessarily for matrices!

- $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
- $\begin{bmatrix} 1 & b \\ -\frac{1}{b} & -1 \end{bmatrix} \begin{bmatrix} 1 & b \\ -\frac{1}{b} & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

A matrix with all elements 0 is the "zero matrix" $0_{m\times n}$

Sometimes subscripts left out

- $A0 = 0$
- $0B = 0$
- But $AB = 0$ does not imply $A = 0$ or $B = 0$
- Possible for $A \neq 0$, yet $A^2 = AA = 0$

Matrix Operations (Matrix Multiplication)

For numbers: $ab = ac \Rightarrow b = c$ for all $a \neq 0$

For matrices, it is possible for $Ab = Ac$, yet $b \neq c$

e.g.,

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

There is an important special case where $Ab = Ac \Longrightarrow b = c$

We'll come to it later

Relationship between Matrix Multiplication and Transpose

Suppose A is $m \times n$ and B is $n \times p$, then

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$$

Proof: We have

$$\begin{aligned}((AB)^{\mathrm{T}})_{ij} = (AB)_{ji} &= \sum_{k=1}^{n} (A)_{jk} (B)_{ki} \\ &= \sum_{k=1}^{n} (A^{\mathrm{T}})_{kj} (B^{\mathrm{T}})_{ik} \\ &= \sum_{k=1}^{n} (B^{\mathrm{T}})_{ik} (A^{\mathrm{T}})_{kj} = (B^{\mathrm{T}} A^{\mathrm{T}})_{ij}\end{aligned}$$

Trace Operator

Given a square $n \times n$ matrix A, the **trace** of A is defined as

$$\text{trace}(A) = \sum_{i=1}^{n} a_{ii}$$

- The trace of a scalar is the scalar itself
- If A is $n \times p$ and B is $p \times n$, then $\text{trace}(AB) = \text{trace}(BA)$

Proof:

$$\text{trace}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{n} b_{ji} a_{ij} = \sum_{j=1}^{n} (BA)_{jj} = \text{trace}(BA)$$

- Example: $\text{trace}(x^{\mathrm{T}} x) = \text{trace}(x x^{\mathrm{T}})$

The Identity Matrix

The **identity matrix** I_n is the $n \times n$ matrix such that

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for} \quad i, j = 1, \dots, n$$

That is,

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

If A is $m \times n$ then

$$AI_n = A \quad \text{and} \quad I_m A = A$$

Diagonal, Upper and Lower Triangular Matrices

The identity matrix is an example of a diagonal matrix

A **diagonal matrix** D is a square matrix such that $(D)_{ij} = 0$ for all $i \neq j$

- It doesn't matter what the diagonal elements $(D)_{ii}$ are
- Diagonal matrices are often written $\text{diag}(d_1, d_2, \dots, d_n)$
- The identity matrix is $\text{diag}(1, 1, \dots, 1)$

A **lower triangular matrix** L is a square matrix such that $(L)_{ij} = 0$ for all $i < j$

An **upper triangular matrix** U is a square matrix such that $(U)_{ij} = 0$ for all $i > j$

Diagonal, Upper and Lower Triangular Matrices

diagonal

$$D = \begin{bmatrix} * & 0 & \cdots & 0 & 0 \\ 0 & * & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & 0 \\ 0 & 0 & \cdots & 0 & * \end{bmatrix}$$

lower triangular

$$L = \begin{bmatrix} * & 0 & \cdots & 0 & 0 \\ * & * & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & * & 0 \\ * & * & \cdots & * & * \end{bmatrix}$$

upper triangular

$$U = \begin{bmatrix} * & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & * \\ 0 & 0 & \cdots & 0 & * \end{bmatrix}$$

where * means any value, including 0

Important examples of matrix products

Example: The general linear system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as $\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ or $Ax = b$

Often the problem is: given A and b, want to find x so that equation holds

Important examples of matrix products

If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, then

$$\text{Inner Product:} \quad x^{\mathrm{T}}x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i^2$$

The norm of a vector x is defined as $\|x\| = \sqrt{x^{\mathrm{T}}x}$

$$\text{Outer Product:} \quad xx^{\mathrm{T}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{bmatrix}$$

Important examples of matrix products

If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ and $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$, then

$$x^{\mathrm{T}} y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

If $x^{\mathrm{T}} y = 0$, the x and y are said to be **orthogonal**

Geometrically, x and y are "perpendicular"

If $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 0$ or $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} x_i = 0$, then

$$x^{\mathrm{T}} y = 0 \implies \textit{Sample Cov}(x_i, y_i) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1} \sum_{i=1}^{n} x_i y_i = 0$$

Important examples of matrix products

If $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$, and $A = (a_{ij})_{n\times n}$, then $x^T A x = \sum_{i=1}^n \sum_{j=1}^n x_i x_j a_{ij}$

For the case when $n = 3$:

$$\begin{aligned} x^T A x &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1 a_{11} + x_2 a_{21} + x_3 a_{31} & x_1 a_{12} + x_2 a_{22} + x_3 a_{32} & x_1 a_{13} + x_2 a_{23} + x_3 a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= x_1^2 a_{11} + x_2^2 a_{22} + x_3^2 a_{33} + x_1 x_2 (a_{12} + a_{21}) + x_1 x_3 (a_{13} + a_{31}) + x_2 x_3 (a_{23} + a_{32}) \end{aligned}$$

When A is symmetric, $x^{\mathrm{T}} A x$ is called a **quadratic form**

Partitioned Matrices

We can partition contents of an $m \times n$ matrix into blocks of submatrices, e.g.,

$$A = \begin{bmatrix} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ \hline 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where $A_{11} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 3 & 2 & 6 \\ 8 & 2 & 1 \end{bmatrix}$ and $A_{22} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 3 \\ 1 & 1 & 7 \end{bmatrix}$

Partitioned Matrices

- Partitioned matrices are often called **block matrices**
- Many ways of partitioning any given matrix, e.g.,

$$A = \begin{bmatrix} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 3 & 2 & 6 \\ 2 & 8 & 2 & 1 \\ 3 & 1 & 2 & 4 \\ \hline 4 & 2 & 1 & 3 \\ 3 & 1 & 1 & 7 \end{array}\right].$$

Main point of this section: as long as the matrices are appropriately partitioned, we can add / multiply partitioned matrices as though the blocks were elements

Partitioned Matrices

Addition of Partitioned Matrices If A and B are two $m \times n$ matrices A and B partitioned as:

$$A = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times n_1} & \underbrace{A_{12}}_{m_1 \times n_2} \\ \underbrace{A_{21}}_{m_2 \times n_1} & \underbrace{A_{22}}_{m_2 \times n_2} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \underbrace{B_{11}}_{m_1 \times n_1} & \underbrace{B_{12}}_{m_1 \times n_2} \\ \underbrace{B_{21}}_{m_2 \times n_1} & \underbrace{B_{22}}_{m_2 \times n_2} \end{bmatrix}$$

where $n_1 + n_2 = n$ and $m_1 + m_2 = m$, then

$$A + B = \begin{bmatrix} \underbrace{A_{11} + B_{11}}_{m_1 \times n_1} & \underbrace{A_{12} + B_{12}}_{m_1 \times n_2} \\ \underbrace{A_{21} + B_{21}}_{m_2 \times n_1} & \underbrace{A_{22} + B_{22}}_{m_2 \times n_2} \end{bmatrix}$$

Partitioned Matrices

Multiplication of Partitioned Matrices. If A and B are $m \times p$ and $p \times n$ respectively, and partitioned as:

$$A = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times p_1} & \underbrace{A_{12}}_{m_1 \times p_2} \\ \underbrace{A_{21}}_{m_2 \times p_1} & \underbrace{A_{22}}_{m_2 \times p_2} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \underbrace{B_{11}}_{p_1 \times n_1} & \underbrace{B_{12}}_{p_1 \times n_2} \\ \underbrace{B_{21}}_{p_2 \times n_1} & \underbrace{B_{22}}_{p_2 \times n_2} \end{bmatrix}$$

then

$$AB = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times p_1} & \underbrace{A_{12}}_{m_1 \times p_2} \\ \underbrace{A_{21}}_{m_2 \times p_1} & \underbrace{A_{22}}_{m_2 \times p_2} \end{bmatrix} \begin{bmatrix} \underbrace{B_{11}}_{p_1 \times n_1} & \underbrace{B_{12}}_{p_1 \times n_2} \\ \underbrace{B_{21}}_{p_2 \times n_1} & \underbrace{B_{22}}_{p_2 \times n_2} \end{bmatrix} = \begin{bmatrix} \underbrace{A_{11}B_{11} + A_{12}B_{21}}_{m_1 \times n_1} & \underbrace{A_{11}B_{12} + A_{12}B_{22}}_{m_1 \times n_2} \\ \underbrace{A_{21}B_{11} + A_{22}B_{21}}_{m_2 \times n_1} & \underbrace{A_{21}B_{12} + A_{22}B_{22}}_{m_2 \times n_2} \end{bmatrix}$$

Partitioned Matrices

Transposition of Partitioned Matrices: We have

$$A = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times n_1} & \underbrace{A_{12}}_{m_1 \times n_2} \\ \underbrace{A_{21}}_{m_2 \times n_1} & \underbrace{A_{22}}_{m_2 \times n_2} \end{bmatrix} \quad \Rightarrow \quad A^{\mathrm{T}} = \begin{bmatrix} \underbrace{A_{11}^{\mathrm{T}}}_{n_1 \times m_1} & \underbrace{A_{21}^{\mathrm{T}}}_{n_1 \times m_2} \\ \underbrace{A_{12}^{\mathrm{T}}}_{n_2 \times m_1} & \underbrace{A_{22}^{\mathrm{T}}}_{n_2 \times m_2} \end{bmatrix}$$

e.g., If X is an $n \times k$ data matrix partitioned into columns, then

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} \Longrightarrow X^{\mathrm{T}} = \begin{bmatrix} X_{*1}^{\mathrm{T}} \\ X_{*2}^{\mathrm{T}} \\ \vdots \\ X_{*k}^{\mathrm{T}} \end{bmatrix}$$

X_{*i} is the column vector of all n observations of variable i

Partitioned Matrices

$$Xc = \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = c_1 X_{*1} + c_2 X_{*2} + \cdots + c_k X_{*k}$$

$$X^{\mathrm{T}}X = \begin{bmatrix} X_{*1}^{\mathrm{T}} \\ X_{*2}^{\mathrm{T}} \\ \vdots \\ X_{*k}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} X_{*1} & X_{*2} & \cdots & X_{*k} \end{bmatrix} = \begin{bmatrix} X_{*1}^{\mathrm{T}} X_{*1} & X_{*1}^{\mathrm{T}} X_{*2} & \cdots & X_{*1}^{\mathrm{T}} X_{*k} \\ X_{*2}^{\mathrm{T}} X_{*1} & X_{*2}^{\mathrm{T}} X_{*2} & \cdots & X_{*2}^{\mathrm{T}} X_{*k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{*k}^{\mathrm{T}} X_{*1} & X_{*k}^{\mathrm{T}} X_{*2} & \cdots & X_{*k}^{\mathrm{T}} X_{*k} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \sum_{i=1}^n x_{i2}x_{i1} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix}$$

Partitioned Matrices

If we partition the data matrix X into rows, i.e.,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} X_{1*} \\ X_{2*} \\ \vdots \\ X_{n*} \end{bmatrix}$$

where X_{j*} is the row vector containing the jth obs of all variables

then
$$X^{\mathrm{T}}X = \begin{bmatrix} X_{1*}^{\mathrm{T}} & X_{2*}^{\mathrm{T}} & \cdots & X_{n*}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} X_{1*} \\ X_{2*} \\ \vdots \\ X_{n*} \end{bmatrix}$$

$$= X_{1*}^{\mathrm{T}}X_{1*} + X_{2*}^{\mathrm{T}}X_{2*} + \cdots + X_{n*}^{\mathrm{T}}X_{n*} = \underbrace{\sum_{i=1}^{n} X_{i*}^{\mathrm{T}}X_{i*}}_{\text{sum of } n \ k \times k \text{ matrices}}$$

Partitioned Matrices

Yet another type of matrix product is the *Kronecker product*

Kronecker product, denoted $\otimes$, of an $m \times n$ matrix A with a $p \times q$ matrix B is the $mp \times nq$ block matrix formed by multiplying each element of A by the entire B matrix

For example

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \left[\begin{array}{cc|cc|cc} a_{11} & 0 & a_{12} & 0 & a_{13} & 0 \\ 0 & a_{11} & 0 & a_{12} & 0 & a_{13} \\ \hline a_{21} & 0 & a_{22} & 0 & a_{23} & 0 \\ 0 & a_{21} & 0 & a_{22} & 0 & a_{23} \end{array}\right]$$

The Inverse Matrix

Let A be an $m \times n$ matrix

- The $n \times m$ matrix B is a **left-inverse** of A if $BA = I_n$
- The $n \times m$ matrix C is a **right-inverse** of A if $AC = I_m$.

Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 0.2 & 0.4 \\ 2 & -0.2 & -0.4 \end{bmatrix}$

B is a left-inverse of A (or A is the right-inverse of B) since

$$BA = \begin{bmatrix} -1 & 0.2 & 0.4 \\ 2 & -0.2 & -0.4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} -1 + 0.4 + 1.6 & -1 + 0.2 + 0.8 \\ 2 - 0.4 - 1.6 & 2 - 0.2 - 0.8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Inverse Matrix

B is not a right-inverse of A since $AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.2 & 0.4 \\ 0 & 0.4 & 0.8 \end{bmatrix}$

In fact A has no right-inverse

Suppose $AC = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ then we have

$$2b + e = 1 \quad \text{and} \quad 4b + 2e = 0$$

which is a contradiction

The Inverse Matrix

A has left- and right-inverse only if it is square, and they will be the same matrix, i.e.,

- If A is $n \times n$, and $BA = AC = I_n$, then it must be that $B = C$

$$BA = I_n \implies BAC = I_nC \implies BI_n = C \implies B = C$$

Then $B = C$ is the "two-sided inverse", or simply the **inverse** of A, denoted A^{-1}

The inverse of a $n \times n$ matrix A, *if it exists*, is the unique matrix A^{-1} such that

$$A^{-1}A = I_n = AA^{-1}$$

The Inverse Matrix

We emphasize

- A has a (two-sided) inverse only if it is square
- But not all square matrices have an inverse

Examples:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \text{is} \quad A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$$

Verify by direct multiplication:

$$A^{-1}A = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Inverse Matrix

The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ has no inverse

Proof: Suppose

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+2c & b+2d \\ 2a+4c & 2b+4d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This implies $a + 2c = 1$ but $2a + 4c = 2(a + 2c) = 0$ which gives a contradiction

The Inverse Matrix

Consider the system of n equations in n unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \implies \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \text{ or } Ax = b$$

If A has an inverse, then the unique solution to this system is

$$Ax = b \iff x = A^{-1}b$$

- $Ax = b \Rightarrow A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$
- $x = A^{-1}b \Rightarrow Ax = AA^{-1}b = b$ so $x = A^{-1}b$ is indeed a solution

The Inverse Matrix

Example: Consider the system

$$\begin{aligned} x_1 + 3x_2 &= 1 \\ 2x_1 + 4x_2 &= 3 \end{aligned} \quad \text{or} \quad \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{or} \quad Ax = b$$

We saw earlier that the inverse of $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ is $A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$

The unique solution is

$$A^{-1}b = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5/2 \\ -1/2 \end{bmatrix}$$

The Inverse Matrix

(Warning) The same argument doesn't quite hold for left-inverses

Suppose the system is $Ax = b$ where A is $m \times n$, $m \geq n$, with left-inverse A^{-1}_{left}

Pre-multiplying both side of $Ax = b$ by A^{-1}_{left} gives

$$A^{-1}_{left}Ax = A^{-1}_{left}b \implies x = A^{-1}_{left}b$$

However, when we check if $x = A^{-1}_{left}b$ is a solution, we get

$$Ax = AA^{-1}_{left}b$$

which *may or may not* be equal to b, since $AA^{-1}_{left} \neq I_m$

- If $AA^{-1}_{left}b = b$, there is a unique solution and you have found it
- If $AA^{-1}_{left}b \neq b$, there is no solution

The Inverse Matrix

Example: consider the systems

$$\text{(i)} \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \text{(ii)} \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

which we write as (i) $Ax = b$ and (ii) $Ax = c$ respectively

The left inverse of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}$ is $A_{left}^{-1} = \begin{bmatrix} -4/9 & 5/9 & 1/9 \\ 5/9 & -4/9 & 1/9 \end{bmatrix}$ (verify!)

You can verify that

- $AA_{left}^{-1}b = b$ (despite $AA_{left}^{-1} \neq I_3$) so $A_{left}^{-1}b$ is a unique solution to (i)
- $AA_{left}^{-1}c \neq c$ so $A_{left}^{-1}c$ is not a solution to (ii)

The Inverse Matrix

Consider the general two-equation two-unknown system

$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 = b_1 \\ a_{21}x_1 &+ a_{22}x_2 = b_2 \end{aligned} \quad \text{or} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad \text{or} \quad Ax = b$$

Solution, if it exists, is $x = A^{-1}b$

If we solve the system “manually”, we get

$$x_1 = \frac{a_{22}\,b_1 - a_{12}\,b_2}{a_{11}a_{22} - a_{12}a_{21}} \quad \text{and} \quad x_2 = \frac{a_{11}\,b_2 - a_{21}\,b_1}{a_{11}a_{22} - a_{12}a_{21}}$$

Solution requires that the (common) denominator in both expressions is not zero

The Inverse Matrix

Define common denominator as the **determinant** of A: $\det(A) = a_{11}a_{22} - a_{12}a_{21}$

We have

$$x = A^{-1}b = \begin{bmatrix} \dfrac{a_{22}\,b_1 - a_{12}\,b_2}{a_{11}a_{22} - a_{12}a_{21}} \\ \dfrac{a_{11}\,b_2 - a_{21}\,b_1}{a_{11}a_{22} - a_{12}a_{21}} \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} a_{22}\,b_1 - a_{12}\,b_2 \\ a_{11}\,b_2 - a_{21}\,b_1 \end{bmatrix} = \underbrace{\frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}}_{A^{-1}} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

The inverse of an arbitrary 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, if it exists, is

$$\textit{Memorize this!} \rightarrow \quad A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad \text{where } \det(A) = |A| = a_{11}a_{22} - a_{12}a_{21}$$

The Inverse Matrix

Notice also that

$$x_1 = \frac{a_{22}\,b_1 - a_{12}\,b_2}{a_{11}a_{22} - a_{12}a_{21}} = \frac{\det(A_1(b))}{\det(A)} \quad \text{and} \quad x_2 = \frac{a_{11}\,b_2 - a_{21}\,b_1}{a_{11}a_{22} - a_{12}a_{21}} = \frac{\det(A_2(b))}{\det(A)}$$

where

$$A_1(b) = \begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix} \quad \text{and} \quad A_2(b) = \begin{bmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{bmatrix}.$$

This is **Cramer's Rule** for a two-equation two-unknown system

The Inverse Matrix

If inverse of A does not exist, we say that A is **singular**

If inverse exists, we say that A is **non-singular**

E.g. The inverse of $A = \begin{bmatrix} 1 & 4 \\ 5 & 6 \end{bmatrix}$ is

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} 6 & -4 \\ -5 & 1 \end{bmatrix} = -\frac{1}{14} \begin{bmatrix} 6 & -4 \\ -5 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{7} & \frac{2}{7} \\ \frac{5}{14} & -\frac{1}{14} \end{bmatrix}$$

A is non-singular

The determinant of $B = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ is $\det(B) = 1 \cdot 6 - 2 \cdot 3 = 0$, so B is singular

The Inverse Matrix

Similar argument for system of n equations in n unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \implies \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \text{ or } Ax = b$$

The solution (if you did it manually) can be written as ratios with common denominator, which we define as the determinant of A

The solution can then also be written

$$x_1 = \frac{\det(A_1(b))}{\det(A)}, \; \dots \;, \; x_n = \frac{\det(A_n(b))}{\det(A)}$$

Solution requires $\det(A) \neq 0$

The Inverse Matrix

We will not need to know the formula for det(A) or A^{-1} for general $n \times n$ case, $n > 2$

See TPB(2025) Chapter 8 if interested

- Gaussian elimination
- Cofactor formula for A^{-1}
- Laplace Expansion for det(A)

The Inverse Matrix

We just need to know when $\det(A) \neq 0$ for an $n \times n$ matrix A

- No zero columns
- No column can be written as a multiple of another column
- No column can be written as a linear combination of another other columns

This will be the case if

$$c_1 A_{*1} + c_2 A_{*2} + \cdots = c_n A_{*n} = 0_n \iff c_1 = c_2 = \cdots = c_n = 0$$

where A_{*i} is the ith column of A, $i = 1, \ldots, n$

i.e., $Ac = 0 \iff c = 0_n$

The Inverse Matrix

If $Ac = 0 \iff c = 0_n$, we say

- columns of A “linearly independent”
- A is “full rank”, and the inverse exists

If there exists $c \neq 0_n$ such that $Ac = 0$, we say

- columns of A are “linearly dependent

The Inverse Matrix

There are many results involving determinants and inverses

We will only need the following:

- The inverse of a diagonal matrix diag$(d_1, \ldots, d_n)$ is diag$(d_1^{-1}, \ldots d_n^{-1})$
- If A is $n \times n$ and non-singular, then $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$

$$\begin{aligned} AA^{-1} = I_n &\Rightarrow (A^{-1})^{\mathrm{T}} A^{\mathrm{T}} = I_n \\ &\Rightarrow (A^{-1})^{\mathrm{T}} A^{\mathrm{T}} (A^{\mathrm{T}})^{-1} = I (A^{\mathrm{T}})^{-1} \\ &\Rightarrow (A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1} \end{aligned}$$

The Inverse Matrix

- A and B are both $n \times n$ and non-singular, then $(AB)^{-1} = B^{-1}A^{-1}$.

$$B^{-1}A^{-1}AB = ABB^{-1}A^{-1} = I \text{ implies } B^{-1}A^{-1} = (AB)^{-1}$$

- The inverse of a non-singular symmetric matrix is symmetric (exercise)

- If X is $n \times k$ with $n > k$ and $Xc = 0_{n\times 1} \iff c = 0_{k\times 1}$, then

$$X^{\mathrm{T}}X \text{ is non-singular}$$

i.e., as long as the columns of the $n \times k$ matrix X, where $n > k$, are linearly independent, $X^{\mathrm{T}}X$ will have an inverse

Inverse of Partitioned Matrices

If $A = \begin{bmatrix} \underbrace{A_{11}}_{m_1 \times m_1} & \underbrace{A_{12}}_{m_1 \times m_2} \\ \underbrace{A_{21}}_{m_2 \times m_1} & \underbrace{A_{22}}_{m_2 \times m_2} \end{bmatrix}$ and non-singular, then

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

You can verify this by direct multiplication, to show that

$$A^{-1}A = \begin{bmatrix} I_{m_1} & 0_{m_1 \times m_2} \\ 0_{m_2 \times m_1} & I_{m_2} \end{bmatrix}$$

Vectors and Matrices of Random Variables

Matrix algebra helps in organizing large numbers of random variable, especially their expectations and variances and covariances

If x is a $m \times 1$ vector of random variables $x = \begin{bmatrix} X_1 & X_2 & \dots & X_m \end{bmatrix}^{\mathrm{T}}$, then we **define**

$$E(x) = \begin{bmatrix} E(X_1) & E(X_2) & \dots & E(X_m) \end{bmatrix}^{\mathrm{T}}$$

If X is a matrix $m \times n$ matrix of random variables, then

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \Leftrightarrow E(X) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \dots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \dots & E(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{m1}) & E(X_{m2}) & \dots & E(X_{mn}) \end{bmatrix}$$

Vectors and Matrices of Random Variables

Let x be a $m \times 1$ vector of random variables. Let

$$\tilde{x} = x - E(x) = \begin{bmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \\ \vdots \\ X_m - E(X_m) \end{bmatrix} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_m \end{bmatrix}$$

Then the **variance-covariance matrix** of x, denoted $Var(x)$, is defined as

$$Var(x) = E((x - E(x))(x - E(x))^{\mathrm{T}}) = E(\tilde{x}\tilde{x}^{\mathrm{T}})$$

Vectors and Matrices of Random Variables

$$
\begin{aligned}
Var(x) = E(\tilde{x}\tilde{x}^{\mathrm{T}}) &= E((x - E[x])(x - E[x])^{\mathrm{T}}) \\
&= E\begin{bmatrix} \tilde{X}_1^2 & \tilde{X}_1\tilde{X}_2 & \dots & \tilde{X}_1\tilde{X}_m \\ \tilde{X}_2\tilde{X}_1 & \tilde{X}_2^2 & \dots & \tilde{X}_2\tilde{X}_m \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{X}_m\tilde{X}_1 & \tilde{X}_m\tilde{X}_2 & \dots & \tilde{X}_m\tilde{X}_m \end{bmatrix} \\
&= \begin{bmatrix} E(\tilde{X}_1^2) & E(\tilde{X}_1\tilde{X}_2) & \dots & E(\tilde{X}_1\tilde{X}_m) \\ E(\tilde{X}_2\tilde{X}_1) & E(\tilde{X}_2^2) & \dots & E(\tilde{X}_2\tilde{X}_m) \\ \vdots & \vdots & \ddots & \vdots \\ E(\tilde{X}_m\tilde{X}_1) & E(\tilde{X}_m\tilde{X}_2) & \dots & E(\tilde{X}_m\tilde{X}_m) \end{bmatrix} \\
&= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_m) \\ Cov(X_1, X_2) & Var(X_2) & \dots & Cov(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_m) & Cov(X_2, X_m) & \dots & Var(X_m) \end{bmatrix}
\end{aligned}
$$

Vectors and Matrices of Random Variables

Recall that if X is a (univariate) random variable, then

- $E(aX + b) = aE(X) + b$
- $Var(aX + b) = a^2 Var(X)$
- $Var(X) = E(X^2) - E(X)^2$

We have matrix analogues of these results: Suppose x is an $m \times 1$ vector of random variables, $A = (a_{ij})_{km}$ is a $k \times m$ matrix of constants and b is a $k \times 1$ vector of constants. Then

- $E(Ax + b) = AE(x) + b$
- $Var(Ax + b) = A\, Var(x) A^{\mathrm{T}}$
- $Var(x) = E(xx^{\mathrm{T}}) - E(x)E(x)^{\mathrm{T}}$

Vectors and Matrices of Random Variables

Proof of $E(Ax + b) = AE(x) + b$:

The ith element of the $k \times 1$ vector $Ax + b$ is $\sum_{j=1}^{m} a_{ij}X_j + b_i$. The expectation of this term is

$$E\left(\sum_{j=1}^{m} a_{ij}X_j + b_i\right) = \sum_{j=1}^{m} a_{ij}E(X_j) + b_i$$

which is the ith element of the vector $AE(x) + b$.

Proof of $Var(Ax + b) = A\,Var(x)A^{\mathrm{T}}$:

Since $Ax + b - E(Ax + b) = A(x - E(x)) = A\tilde{x}$, we have

$$\begin{aligned} Var(Ax + b) &= E((A\tilde{x})(A\tilde{x})^{\mathrm{T}}) = E(A\tilde{x}\tilde{x}^{\mathrm{T}}A^{\mathrm{T}}) = AE(\tilde{x}\tilde{x}^{\mathrm{T}})A^{\mathrm{T}} \\ &= A\ Var(x)A^{\mathrm{T}}\,. \end{aligned}$$

Proof of $Var(x) = E(xx^{\mathrm{T}}) - E(x)E(x)^{\mathrm{T}}$: Exercise!

Vectors and Matrices of Random Variables

Let x be $m \times 1$ vector of random variables, c be $m \times 1$ non-zero vector of constants

- Obviously the variance-covariance matrix of x is symmetric
- Consider the linear combination $c^{\mathrm{T}}x$ (this is a now a single random variable). We have

$$Var(c^{\mathrm{T}}x) = c^{\mathrm{T}}\, Var(x)c \geq 0 \ \text{ for all } \ c \neq 0_{m\times 1}$$

- $Var(c^{\mathrm{T}}x)$ cannot be negative since it is a variance
- If $Var(c^{\mathrm{T}}x) = 0$ then either
  - one of the random variables is not actually random, or
  - one of the random variables is just a multiple of the other
  - one of the random variables is a linear combination of two or more of the other random variables

Digression on Symmetric Matrices

A $m \times m$ symmetric (and square) matrix A is **positive definite** if

$$c^{\mathrm{T}} A c > 0 \text{ for all } c \neq 0_{m \times 1}$$

It is **positive semidefinite** if $c^{\mathrm{T}} A c \geq 0$ for all $c \neq 0_{m \times 1}$. Similar definitions for **negative definiteness** and **negative semidefiniteness**

- Variance covariance matrices $Var(x)$ are **positive semidefinite**
- If the random variables in x are not linearly dependent, then $Var(x)$ is **positive definite**
- Another example: suppose the columns of a $n \times k$ data matrix X are linearly independent, i.e.,

  $$Xc \neq 0_{n \times 1} \text{ for all } c \neq 0_{k \times 1}$$

  Then $c^{\mathrm{T}} X^{\mathrm{T}} X c = (Xc)^{\mathrm{T}} X c > 0$ for all $c \neq 0_{k \times 1}$, i.e., $X^{\mathrm{T}} X$ is positive definite

Sample Variance-Covariance Matrix

Let X be a $n \times k$ data matrix, each column contains n observations of some variable, mean removed. Then

$$X = \begin{bmatrix} x_{11} - \overline{x_1} & x_{12} - \overline{x_2} & \cdots & x_{1k} - \overline{x_k} \\ x_{21} - \overline{x_1} & x_{22} - \overline{x_2} & \cdots & x_{2k} - \overline{x_k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \overline{x_1} & x_{n2} - \overline{x_2} & \cdots & x_{nk} - \overline{x_k} \end{bmatrix} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1k} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdots & \tilde{x}_{nk} \end{bmatrix}$$

and $\dfrac{1}{n-1}X^{\mathrm{T}}X$ is the symmetric sample variance-covariance matrix

$$\frac{1}{n-1}X^{\mathrm{T}}X = \begin{bmatrix} \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}^2 & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}\tilde{x}_{i2} & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i1}\tilde{x}_{ik} \\ \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}\tilde{x}_{i1} & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}^2 & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{i2}\tilde{x}_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}\tilde{x}_{i1} & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}\tilde{x}_{i2} & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}\tilde{x}_{ik}^2 \end{bmatrix}$$

Eigendecomposition of Symmetric Matrices

Another very important fact about symmetric matrices (Eigendecomposition)

Every $k \times k$ symmetric matrix A can be decomposed in the following way

$$A = Q\Lambda Q^{\mathrm{T}} = \begin{bmatrix} q_1 & q_2 & \dots & q_k \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \vdots \\ q_k^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{k} \lambda_i q_i q_i^{\mathrm{T}}$$

- λ_i, $i = 1, \dots, k$ are real numbers called **eigenvalues** (usually ranked $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$)
- The $k \times 1$ vectors q_i, $i = 1, \dots, k$ are the corresponding **eigenvectors**
- Q satisfies the property $Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I_k$

For full discussion, see TPB (2025) Chapter 10

Eigendecomposition of Symmetric Matrices

Suppose the $k \times 1$ random vector x has variance-covariance matrix Σ

Σ is positive definite, but not necessarily diagonal

The random variables in x may be correlated

We can write

$$\Sigma = Q\Lambda Q^{\mathrm{T}} \quad \text{such that} \quad Q^{\mathrm{T}}Q = I_k$$

Eigendecomposition of Symmetric Matrices

Consider the $k \times 1$ vector of random variables

$$Q^{\mathrm{T}}x = \begin{bmatrix} q_1^{\mathrm{T}}x \\ q_2^{\mathrm{T}}x \\ \vdots \\ q_k^{\mathrm{T}}x \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = y$$

Each y_i is a weighted average of the random variables in x

Variance-Covariance matrix of y is

$$Var(y) = Var(Q^{\mathrm{T}}x) = Q^{\mathrm{T}}\Sigma Q = Q^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}Q = \Lambda$$

Since Λ is diagonal, the random variables in y are uncorrelated (despite the fact that they are all weighted averages of the random variables in x)

Differentiation of Matrix Forms

Let $f(x)$ be some function of the $n \times 1$ vector x

e.g., $f(x) = x^{\mathrm{T}} A x$ where A is some $n \times n$ matrix of constants

This is just a multivariable function $f(x_1, \ldots, x_n)$

We *define*

$$\nabla f = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \quad \text{or} \quad Df = \frac{\partial f}{\partial x^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Differentiation of Matrix Forms

Example: If $y = x^{\mathrm{T}}Ax$ where A is $n \times n$ and x is $n \times 1$, then

$$\frac{\partial y}{\partial x} = \frac{\partial}{\partial x}(x^{\mathrm{T}}Ax) = (A + A^{\mathrm{T}})x$$

$n = 3$ example:

$$x^T Ax = x_1^2 a_{11} + x_2^2 a_{22} + x_3^2 a_{33} + x_1 x_2 (a_{12} + a_{21}) + x_1 x_3 (a_{13} + a_{31}) + x_2 x_3 (a_{23} + a_{32})$$

$$\left.\begin{aligned} \frac{\partial x^T Ax}{\partial x_1} &= 2x_1 a_{11} + x_2(a_{12} + a_{21}) + x_3(a_{13} + a_{31}) \\ \frac{\partial x^T Ax}{\partial x_2} &= x_1(a_{12} + a_{21}) + 2x_2 a_{22} + x_3(a_{23} + a_{32}) \\ \frac{\partial x^T Ax}{\partial x_3} &= x_1(a_{13} + a_{31}) + x_2(a_{23} + a_{32}) + 2x_3 a_{33} \end{aligned}\right\} \Rightarrow \quad \frac{\partial x^{\mathrm{T}} Ax}{\partial x} = (A + A^{\mathrm{T}})x$$

If A is symmetric, then $\dfrac{\partial x^{\mathrm{T}}Ax}{\partial x} = (A + A^{\mathrm{T}})x = 2Ax$ (compare $f(x) = ax^2 \Rightarrow f'(x) = 2ax$)

Differentiation of Matrix Forms

Suppose x is $n \times 1$ and $y = f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix}$ (i.e., y is a vector of m different functions)

Then we define

$$\frac{\partial y}{\partial x^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad \text{or} \quad \frac{\partial y^{\mathrm{T}}}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Differentiation of Matrix Forms

E.g.: if $y = f(x) = Ax$ where $A = (a_{ij})_{m\times n}$ and $x^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$, then

$$\frac{\partial y}{\partial x^{\mathrm{T}}} = \frac{\partial (Ax)}{\partial x^{\mathrm{T}}} = A \quad \text{or} \quad \frac{\partial y^{\mathrm{T}}}{\partial x} = \frac{\partial (x^{\mathrm{T}} A^{\mathrm{T}})}{\partial x} = A^{\mathrm{T}}$$

Matrix analogue of the univariate differentiation rule $f(x) = ax \Rightarrow f'(x) = a$

Proof:

- The product Ax is an $m \times 1$ vector whose i-th element is $\sum_{k=1}^{n} a_{ik}x_k$
- Therefore the $(i,j)th$ element of $\partial y/\partial x^{\mathrm{T}}$ is $(\partial/\partial x_j)\sum_{k=1}^{n} a_{ik}x_k = a_{ij}$
- This says that $\partial y/\partial x^{\mathrm{T}} = A$

Differentiation of Matrix Forms

E.g.: If $y = f(x)$ is a scalar-valued function of an $n \times 1$ vector of variables x, then

$$\frac{\partial}{\partial x^{\mathrm{T}}}\left(\frac{\partial y}{\partial x}\right) = \frac{\partial}{\partial x^{\mathrm{T}}}\begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 y}{\partial x_1^2} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\ \frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2^2} & \cdots & \frac{\partial^2 y}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 y}{\partial x_n \partial x_1} & \frac{\partial^2 y}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_n^2} \end{bmatrix}$$

- This is the Hessian matrix of $y = f(x)$. We usually write $\frac{\partial}{\partial x^{\mathrm{T}}}\left(\frac{\partial y}{\partial x}\right)$ as $\frac{\partial^2 y}{\partial x \partial x^{\mathrm{T}}}$

Optimization

The interior optimum point x^* of a function of n variables $f(x)$ satisfies the first order conditions

$$\nabla f(x^*) = \left.\frac{\partial f}{\partial x}\right|_{x=x^*} = \begin{bmatrix} \dfrac{\partial f(x^*)}{\partial x_1} \\ \dfrac{\partial f(x^*)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x^*)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If the Hessian is positive definite, then x^* is a global minimum point

If the Hessian is negative definite, then x^* is a globaal maximum point