Lecture 1. Review of Matrix Algebra

Yonghui Zhang

School of Economics, RUC

September 2020

Outline

1. Matrix and Vector: definitions
2. Partition matrix
3. Trace, Determinant
4. Rank, Inverse Matrix
5. Eigenvalues and Eigenvectors
6. Quadratic form
7. Idempotent matrix
8. Decomposition
9. Differentiation of Matrices

Matrix

Definition (Matrix)

An $m \times n$ matrix A is a rectangular array of elements in m rows and n columns. Write

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij}),$$

where a_{ij} is the ith row and jth column of A.

- Sometimes write $A : m \times n$ or $A_{m \times n}$ to indicate an $m \times n$ matrix A.
- When $n = m$, the matrix A is called a **square** *matrix*.
- A square matrix having zeros as elements below (above) the diagonal is called an upper **(lower) triangular matrix**.

Matrix

Example (Matrice in Econometrics)

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} \text{ and } Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

where n is the number of observations and p is the number of regressors (independent variables).
In general, $n > p$.
But for high dimensional data sets, it is common to have $p > n$ or $p >> n$.

Basic operations of matrices

- A is an $m \times n$ matrix:
  - For any scalar c, define $cA = (ca_{ij})$;
  - For any $m \times n$ matrix B, define $A \pm B = (a_{ij} \pm b_{ij})$;
  - For an $n \times p$ matrix $C = (c_{ij})$, define the product of two conformable matrices A and C
$$AC = \left(\sum_{k=1}^{n} a_{ik} c_{kj} \right)_{m \times p}.$$
- For the product of matrices, we have
$$\begin{aligned} ABC &= A(BC) = (AB)C; \\ AB &\neq BA. \end{aligned}$$

Transpose

Definition (Transpose)

The transpose A'(or A^T) : $m \times n$ of a matrix $A_{m \times n}$ is obtained by interchanging the rows and columns and columns of A. Thus

$$A' = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix} = (a_{ji}) .$$

- Transpose satisfies the following properties:

$$(A')' = A, \ (A+B)' = A' + B', \ (AB)' = B'A'$$

- A square matrix A is *symmetric* if $A' = A$.

Some special matrices

Definition (Vector)

An $m \times 1$ matrix a is said to be a column m-vector and written as

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}.$$

A $1 \times m$ matrix a' is said to be a **row** vector and written as $a' = (a_1, \cdots, a_m)$.

- All vectors in this course are *column vectors*.
- In Econometrics, we usually use X_i to denote the ith observation, then

$$X = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix} = (X_1, \cdots, X_n)'.$$

Vector

Definition (Inner product (dot prodcut))

$$a \cdot b \equiv a'b = a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

- Properties: (i) $a \cdot b = b \cdot a$, (ii) $a \cdot (b + c) = a \cdot b + a \cdot c$; (iii) $(ca) \cdot b = a \cdot (cb) = c\,(a \cdot b)$ for any scalar c
- Euclidean norm (length): $\|a\| = (a \cdot a)^{1/2} = (a'a)^{1/2}$
  - Cauchy-Schwarz inequality: $\|a \cdot b\| \leq \|a\|\,\|b\|$
  - Triangle inequality for vector norms: $\|a + b\| \leq \|a\| + \|b\|$ ()
- Angle θ between two nonzero vectors a and b is determined by

$$\cos\theta = \frac{a \cdot b}{\|a\|\,\|b\|}, \theta \in [0, \pi]$$

when $\theta = \pi/2$, a and b are orthogonal and $a \cdot b$, and it is denoted as $a \perp b$.

Some special matrices

- Other commonly-used special matrices:
  - Identity matrix: $I_n = \texttt{Diag}(1, \cdots, 1)$
  - Unity vector: $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)'$
  - Ones vector: $i_n = (1, , \cdots, 1)'$ (sometimes ι_n)
  - Zeros vector: $0_{n\times 1} = (0, \cdots, 0)'$
  - Zeros matrix: $0_{n\times p}$
  - Time regressors: $\mathbf{T} = (1, 2, \cdots, t, \cdots, T)'$ or $(1/T, 2/T, \cdots, 1)'$.

Definition (Orthogonality)

(i) Two vectors a and b are orthogonal if $a'b = 0$;
(ii) Two matrices $A_{n\times p}$ and $B_{p\times m}$ are orthogonal if $AB = 0_{n\times p}$.

Partition matrix

Definition (Partition)

A matrix A is said to be partitioned if its elements are arranged in submatrices.

- For example, there are 4 blocks for matrices A and B:

$$\underset{m\times n}{A} = \begin{pmatrix} \underset{m_1\times n_1}{A_{11}} & \underset{m_1\times n_2}{A_{12}} \\ \underset{m_2\times n_1}{A_{21}} & \underset{m_2\times n_2}{A_{22}} \end{pmatrix} \text{ and } \underset{m\times n}{B} = \begin{pmatrix} \underset{m_1\times n_1}{B_{11}} & \underset{m_1\times n_2}{B_{12}} \\ \underset{m_2\times n_1}{B_{21}} & \underset{m_2\times n_2}{B_{22}} \end{pmatrix}$$

where $m = m_1 + m_2$ and $n = n_1 + n_2$. Then we have

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix} \text{ and } A' = \begin{pmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{pmatrix}.$$

Partition of matrix

Example (Partition the regressor matrix: $X : n \times p$)

(i) Partition according to different sets of variables:

$$X_{n \times p} = (X_1, X_2),$$

where $X_1 : n \times p_1$, $X_2 : n \times p_2$ and $p = p_1 + p_2$.

(ii) Partition according to subsamples:

$$X = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$$

where $X_{(1)} : n_1 \times p$, $X_{(2)} : n_2 \times p$ and $n = n_1 + n_2$.

Trace

Definition (Trace)

For an $n \times n$ square matrix $A = (a_{ij})$, its trace is defined as the sum of its diagonal elements, ie.,

$$\mathtt{tr}\,(A) = \sum_{i=1}^{n} a_{ii}.$$

- Let A and B be both $n \times n$ square matrices, and let c be a scalar. Then we have the following rules:
  - $\mathtt{tr}\,(A+B) = \mathtt{tr}(A) + \mathtt{tr}(B)$; $\mathtt{tr}\,(A') = \mathtt{tr}(A)$; $\mathtt{tr}\,(cA) = c\mathtt{tr}\,(A)$;
  - (**Rotation property**)

    $$\mathtt{tr}\,(AB) = \mathtt{tr}(BA)$$

    (It also holds for the general case $A : n \times m$ and $B : m \times n$)
  - $\mathtt{tr}\,(AA') = \mathtt{tr}\,(A'A) = \sum_{i=1}^{n} \sum_{i=j}^{n} a_{ij}^2 = \|A\|_2^2$ ($\|A\|_2$-Euclidean norm, sometime $\|A\|$)
- For the regressor matrix X : $\mathtt{tr}\,(XX') = \mathtt{tr}\,(X'X)$.

Determinant

Definitions (Determinant I)

Let $A = (a_{ij})$ be an $n \times n$ matrix and $\pi = (j_1, \ldots, j_n)$ be a permutation of $(1, \ldots, n)$. There are $n!$ such permutations. For each permutation, there is a unique count of the number of inversions of the indices of such permutations (relative to the natural order $(1, \ldots, n)$), and let $\varepsilon_\pi = 1$ if this count is **even** and $\varepsilon_\pi = -1$ if the count is **odd.** Then the determinant of A is defined as

$$\det A = |A| = \sum_{\text{all } \pi'\text{s}} \varepsilon_\pi a_{1j_1} a_{2j_2} \cdots a_{nj_m}.$$

Example (2×2 matrix A)

Two permutations of $(1, 2)$ are $(1, 2)$ and $(2, 1)$ with $\varepsilon_{(1,2)} = 1$ and $\varepsilon_{(2,1)} = -1$. Thus $\det A = \varepsilon_{(1,2)} a_{11} a_{22} + \varepsilon_{(2,1)} a_{12} a_{21} = a_{11} a_{22} - a_{12} a_{21}$.

Determinant

Definition (Determinant II)

The determinant of an $n \times n$ square matrix A is defined by

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij} \text{ for any fixed } j$$

with M_{ij} being the **minor** of the elements a_{ij}, ie., the determinant of the remaining $(n-1) \times (n-1)$ matrix when the ith row and the jth column of A are deleted, and $A_{ij} = (-1)^{i+j} M_{ij}$ is the **cofactor** of a_{ij}.

Remark: The definition can also be extended to more general expansions with any fixed k rows.

Example (2×2 matrix A)

$|A| = a_{11}a_{22} - a_{12}a_{21}$, where $M_{11} = a_{22}$ and $M_{12} = a_{21}$.

Determinant

Example (3×3 matrix A)

If $A = (a_{ij})$ is a 3×3 matrix, we can hold column $j = 1$ fixed and have

$$\begin{aligned} A_{11} &= (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = M_{11}, \\ A_{21} &= (-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = -M_{21}, \\ A_{31} &= (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = M_{31}. \end{aligned}$$

Consequently, $|A| = a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31}$.

Determinant

- A square matrix A is said to be **nonsingular** if $|A| \neq 0$; otherwise, A is said to be **singular**.
- Let A and B be $n \times n$ square matrices, and c be a scalar. Then we have the following properties:
  - $|A'| = |A|$;
  - $|cA| = c^n |A|$;
  - $|AB| = |A| |B|$;
  - $\left|A^2\right| = |A|^2$;
  - $\det A = \pm 1$ if A is orthonormal $(A'A = I)$
  - If A is diagonal or triangular, then $|A| = \prod_{i=1}^{n} a_{ii}$;

Determinant

Other important facts for block matrices:

- $\begin{vmatrix} A_{11} & 0 \\ 0 & A_{22} \end{vmatrix} = |A_{11}|\,|A_{22}|.$
- $\begin{vmatrix} A & 0 \\ C & B \end{vmatrix} = |A|\,|B|;$
- $\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D|\,\left|A - BD^{-1}C\right|$ if $|D| \neq 0$

Rank

Definition (Linear dependence/independence)

A set of vectors $v_1, \cdots, v_n$ are linearly dependent if only if (iff) one of the vectors in the set can be expressed as a linear combination of the others. They are linearly independent iff the only solution to

$$c_1 v_1 + \cdots + c_n v_n = 0$$

is $c_1 = \cdots = c_n = 0$.

Definition (Rank)

The rank of $A : m \times n$ is the maximum number of linearly independent rows (or columns) of A denoted as $\text{rank}(A)$.

- $\texttt{rank}\begin{pmatrix} -1 & 1 \\ 2 & -4 \end{pmatrix} =? \ \texttt{rank}\begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} =?$

Rank

Let A be an $m \times n$ matrix. Then

- $0 \leq \texttt{rank}(A) \leq \min\{m, n\}$;
- $\texttt{rank}(A) = \texttt{rank}(A')$;
- $\texttt{rank}(A + B) \leq \texttt{rank}(A) + \texttt{rank}(B)$;
- $\texttt{rank}(AB) \leq \min\{\texttt{rank}(A), \texttt{rank}(B)\}$;
- $\texttt{rank}(A) = \texttt{rank}(A'A) = \texttt{rank}(AA')$;
- For any nonsingular $B : m \times m$ and $C : n \times n$, $\texttt{rank}(BAC) = \texttt{rank}(A)$;
- If $A : n \times n$ is diagonal, then $\texttt{rank}(A)$ equals the number of a_{ii} that is nonzero.

The inverse matrix

Definition (Inverse matrix)

Let A be an $n \times n$ square matrix with full rank ($\texttt{rank}(A) = n$), the inverse A^{-1} of A is defined to be a matrix B satisfying

$$AB = BA = I_n.$$

The inverse A^{-1} exists iff A is full rank, or equivalently, A is nonsingular.

Example

Let A is a 2×2 nonsingular matrix, then $A^{-1} = \frac{1}{|A|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$, where $|A| = a_{11}a_{22} - a_{12}a_{21}$.

The inverse matrix

- More general, if $A = \left(a_{ij}\right)_{n\times n}$ and $|A| \neq 0$, the unique inverse of A is given by

$$A^{-1} = \frac{1}{|A|}\texttt{adj}\left(A\right), \text{ where } \texttt{adj}\left(A\right) = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}.$$

The inverse matrix

- Let A and B be $n \times n$ square matrices with full rank, and c be a nonzero scalar. Then we have the following properties.
  - $(cA)^{-1} = c^{-1}A^{-1}$;
  - $(AB)^{-1} = B^{-1}A^{-1}$;
  - $\left(A^{-1}\right)^{-1} = A$;
  - $(A')^{-1} = \left(A^{-1}\right)'$;
  - $\left|A^{-1}\right| = |A|^{-1}$;
  - $(A + C)^{-1} = A^{-1}\left(A^{-1} + C^{-1}\right)^{-1} C^{-1}$;
  - $A^{-1} - (A + C)^{-1} = A^{-1}\left(A^{-1} + C^{-1}\right)^{-1} A^{-1}$.

The inverse matrix

- If a $k \times k$ matrix H is orthonormal $(H'H = I_k)$, then H is nonsingular and $H^{-1} = H'$. Furthermore, $HH' = I_k$ and $(H')^{-1} = H$.
- **Woodbury matrix identity:** For a non-singular A,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}BC\left(C + CDA^{-1}BC\right)^{-1}CDA^{-1}.$$

- **Sherman–Morrison formula**: When $C = 1$, $B = b$ and $D = b'$ for a vector b

$$\left(A + bb'\right)^{-1} = A^{-1} - \left(1 + b'A^{-1}b\right)^{-1}A^{-1}bb'A^{-1}.$$

Similarly, when $C = -1$

$$\left(A - bb'\right)^{-1} = A^{-1} + \left(1 - b'A^{-1}b\right)^{-1}A^{-1}bb'A^{-1}.$$

The inverse matrix

- If $A=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}$, then $A^{-1}=\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix}$, where A_{ii} is also a square invertible matrix.
- If $A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, then

$$A^{-1}=\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}=\begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix}$$

where $A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$, $A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$. By **Woodbury matrix identity,** we have

$$\begin{aligned} A^{11} &= A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} \\ A^{22} &= A_{22}^{-1}+A_{22}^{1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ A^{12} &= -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ A^{21} &= -A_{22}^{-1}A_{21}A_{11.2}^{-1} \end{aligned}$$

Generalized inverse

Frequently we use the generalized inverse of a matrix when it is singular.

Definition (Generalized inverse)

A generalized inverse A^- of a matrix A satisfies the property

$$A^- A A^- = A^-.$$

Note that A^- is generally not unique and it reduces to the usual inverse A^{-1} if A is a nonsingular square matrix.

Definition (Moore-Penrose generalized inverse)

The Moore-Penrose generalized inverse A^- exists, is unique, and satisfies the following three properties:
(i) $A^- A A^- = A^-$;
(ii) AA^- is symmetric;
(iii) $A^- A$ is symmetric.

Generalized inverse

Example

$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, then

$$A^{-} = \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \end{pmatrix}.$$

More general, if $A = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}$, then

$$A^{-} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Eigenvalues and Eigenvectors

Definition (Eigenvalues or Characteristic Roots)

Let A be an $n \times n$ square matrix, then

$$Q(\lambda) = |A - \lambda I_n|$$

is an nth order polynomial in λ. The n roots $\lambda_1, \cdots, \lambda_n$ of the characteristic function $Q(\lambda) = |A - \lambda I_n| = 0$ are called eigenvalues or characteristic roots of A.

Let A be an $n \times n$ square matrix and x be a n-vector. Consider

$$Ax = \lambda x.$$

Then we have $(A - \lambda I_n)x = 0$. The nontrivial solution of x (i.e., $x \neq 0$) to the above problem exists only if

$$|A - \lambda I_n| = 0.$$

Otherwise, $(A - \lambda I_n)^{-1}$ exists such that $x = (A - \lambda I_n)^{-1} 0 = 0$.

Eigenvalues and Eigenvectors

Definition (Eigenvalue & Eigenvector)

Let λ^* be an eigenvalue of A. Corresponding to λ^* the value of x^* that satisfies

$$Ax^* = \lambda^* x^*$$

is called the eigenvector of A. Often, we impose the normalization rule $x^{*\prime}x^* = 1$.

Eigenvalues and Eigenvectors

- Let A be a *real symmetric* $n \times n$ matrix.
  - The eigenvalues of A are *real*.
  - The eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.
  - A can be diagonalized. That is, there exists an orthogonal matrix X (i.e., $X'X = XX' = I_n$ or equivalently, $X' = X^{-1}$) and a diagonal matrix
  $$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$
  such that $X'AX = \Lambda$.
  - $|A| = \prod_{i=1}^n \lambda_i$.
  - $\texttt{tr}\,(A) = \sum_{i=1}^n \lambda_i$

Definite Matrices and Quadratic Forms

Definition (Quadratic form)

Let A be an $n \times n$ symmetric matrix and x an $n \times 1$ vector. Then the quadratic form in x is defined as the function

$$Q(x) = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Then
(i) A is positive definite (p.d.) if $x'Ax > 0$ for all $x \neq 0$;
(ii) It is negative definite (n.d.) if $x'Ax < 0$ for all $x \neq 0$;
(iiii) It is positive semidefinite. (p.s.d.) if $x'Ax \geq 0$ for all x;
(iv) It is negative semidefinite. (n.s.d.) if $x'Ax \leq 0$ for all x.

Definite Matrices and Quadratic Forms

- Here are some properties about definite matrices.
  - Let a be an $n \times 1$ vector, then $A = aa'$ is always p.s.d;
  - If A is p.s.d. (p.d.), then the eigenvalues of A are all not less than 0 (greater than 0);
  - Let A be a real symmetric p.s.d. $n \times n$ matrix. Then there exists a matrix C such that
  $$A = C'C.$$
  Note that C is not unique. Since A is real symmetric, there exists an orthogonal matrix X and a nonnegative diagonal matrix Λ such that
  $$A = X\Lambda X' = X\Lambda^{1/2}\Lambda^{1/2}X' = C'C,$$
  with $\Lambda^{1/2} = \mathtt{diag}\left(\lambda_1^{1/2}, \cdots, \lambda_n^{1/2}\right)$ and $C = \Lambda^{1/2}X'$. But one can also choose $C = X\Lambda^{1/2}X'$. (Please check it by yourself.) If we require C to be real symmetric, then it is unique

Idempotent Matrix

Definition (Idempotent matrix)

An $n \times n$ square matrix A is idempotent iff

$$A^2 \equiv AA = A.$$

An idempotent matrix A is called an **orthogonal projector** or a projection matrix if $A = A'$.

- Note that a matrix can be idempotent but not symmetric, e.g., $A = \begin{pmatrix} -2 & 1 \\ -6 & 3 \end{pmatrix}$.
- $P_X = X\left(X'X\right)^{-1}X'$

Idempotent Matrix

- Let A be an $n \times n$ idempotent matrix. Then we have:
  - $\texttt{rank}\,(A) = \texttt{tr}(A)$;
  - $I_n - A$ is idempotent;
  - If A is symmetric, then its eigenvalues are 0 or 1, and it is p.s.d.(Check it!)
  - If A is of full rank n, then $A = I_n$.
  - If A and B are idempotent and if $AB = BA$, then AB is also idempotent.
  - If A is idempotent and B is orthogonal, then BAB^T is also idempotent.

Singular Values

Definition (Singular value)

The singular values of a $k \times r$ real matrix A are the positive square roots of the eigenvalues of $A'A$. Thus for $j = 1, \ldots, r$

$$s_j = \sqrt{\lambda_j \left(A'A\right)}$$

- Since $A'A$ is p.s.d., its eigenvalues are non-negative. Thus singular values are always real and non-negative.
- The non-zero singular values of A and A' are the same.
- When A is p.s.d., then the singular values of A correspond to its eigenvalues.
- It is convention to write the singular values in descending order $s_1 \geq s_2 \geq \cdots \geq s_r$.

Matrix Decompositions (I)

Definition (Spectral Decomposition)

If A is $n \times n$ and **real symmetric** then

$$A = H\Lambda H'$$

where H contains the eigenvectors and $H'H = I_n$, Λ is a diagonal matrix with the (real) eigenvalues on the diagonal.

Definition (Eigendecomposition)

If A is $n \times n$ and has **distinct** eigenvalues, there exists a nonsingular matrix P such that $A = P\Lambda P^{-1}$ and $P^{-1}AP = \Lambda$. The columns of P are the eigenvectors and Λ is diagonal with the eigenvalues on the diagonal.

Matrix Decompositions (I)

Definition (Matrix Square Root)

If A is $n \times n$ and positive definite we can find a matrix B such that $A = BB'$. We call B a matrix square root of A and is typically written as $B = A^{1/2}$.

Definition (Singular Value Decomposition, SVD)

If A is $k \times r$ then $A = U\Lambda V'$ where U is $k \times k$, Λ is $k \times r$ and V is $r \times r$. U and V are orthonormal $(U^T U = I_k$ and $V^T V = I_r)$. Λ is a diagonal matrix with the singular values of A on the diagonal.

Definition (Cholesky Decomposition)

If $k \times k$ matrix A is p.d., then $A = LL'$, where L is **lower triangular** and **full rank**.

Matrix Decompositions (II)

Definition (QR Decomposition)

If A is $k \times r$ with $k \geq r$ and rank r then $A = QR$,where Q is a $k \times r$ and **orthonormal** matrix ($Q^T Q = I_r$), R is an $r \times r$ full rank **upper triangular** matrix.

Definition (Jordan Decomposition)

If A is $k \times k$ with r unique eigenvalues then $A = PJP^{-1}$ where J takes the Jordan normal form. The latter is a block diagonal matrix $J = \mathtt{diag}\left(J_1, \ldots, J_r\right)$. The Jordan blocks J_i are $m_i \times m_i$ where m_i is the multiplicity of λ_i (number of eigenvalues equalling λ_i) and take the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}$$

illustrated here for $m_i = 3$.

Differentiation of Matrices

Definition

If $f(X)$ is a real function of an $m \times n$ matrix $X = (x_{ij})$, then the partial differential of f with respect to X is defined as the $m \times n$ matrix of partial differentials $\partial f(X)/\partial x_{ij}$

$$\frac{\partial f(X)}{\partial X} = \begin{pmatrix} \frac{\partial f(X)}{\partial x_{11}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(X)}{\partial x_{mn}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{pmatrix}$$

When $x = (x_1, \cdots, x_n)^T$ and $y = f(x)$ be a real function. Then

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

Differentiation of Matrices

Definition

Let $x = (x_1, \cdots, x_n)'$ and $g(x) = (g_1(x), \cdots, g_m(x))'$. Define

$$\frac{\partial g(x)}{\partial x'} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x'} \\ \vdots \\ \frac{\partial g_m(x)}{\partial x'} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} & \cdots & \frac{\partial g_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x_1} & \cdots & \frac{\partial g_m(x)}{\partial x_n} \end{pmatrix}$$

and $\frac{\partial g'(x)}{\partial x} = \left(\frac{\partial g(x)}{\partial x'}\right)'$.

Example

Let $a = (a_1, \cdots, a_n)'$, $x = (x_1, \cdots, x_n)'$, and $y = a'x = \sum_{i=1}^n a_i x_i$. Then

$$\frac{\partial (a'x)}{\partial x'} = \begin{pmatrix} \frac{\partial (a'x)}{\partial x_1} \\ \vdots \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \end{pmatrix} = a.$$

Differentiation of Matrices

Example

Let $A = (a_{ij})$ be a $m \times n$ matrix and $y = Ax$. Then

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} x = \begin{pmatrix} a_1'x \\ \vdots \\ a_m'x \end{pmatrix}$$

where a_i' is the ith row of A. Then $\partial y_i / \partial x' = a_i'$ and

$$\frac{\partial (Ax)}{\partial x'} = \begin{pmatrix} \frac{\partial y_1}{\partial x'} \\ \vdots \\ \frac{\partial y_m}{\partial x'} \end{pmatrix} = \begin{pmatrix} a_1' \\ \vdots \\ a_n' \end{pmatrix} = A.$$

Similarly, $\frac{\partial (x'A')}{\partial x} = A'$.

Differentiation of Matrices

Example

Let $x = (x_1, \cdots, x_n)'$, $y = (y_1, \cdots, y_n)'$ and $A = (a_{ij})$ be an $n \times n$ matrix. Let $z = x'Ay = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j$. Then

$$\frac{\partial (x'Ay)}{\partial x} = Ay.$$

If $x = y$,

$$\begin{aligned} \frac{\partial (x'Ax)}{\partial x} &= (A + A')\, x \\ &= 2Ax \text{ is } A \text{ is symmetric,} \end{aligned}$$

and $\frac{\partial^2 (x'Ax)}{\partial x' \partial x} = A + A'$. Noting that $\frac{\partial (x'Ax)}{\partial a_{ij}} = x_i x_j$, we have $\frac{\partial (x'Ax)}{\partial A} = xx'$.

Differentiation of Matrices

- Summary

$$
\begin{array}{cc}
\frac{\partial(a'x)}{\partial x'} = a & \frac{\partial(Ax)}{\partial x'} = A \\
\frac{\partial(x'Ay)}{\partial x} = Ay & \frac{\partial^2(x'Ay)}{\partial y'\partial x'} = A \\
\frac{\partial(x'Ax)}{\partial x'} = (A + A')\,x & \frac{\partial^2(x'Ax)}{\partial x\partial x'} = A + A'
\end{array}
$$

Example (OLS)

$$
\begin{aligned}
Q(\beta) &= \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2 = (Y - X\beta)'(Y - X\beta) \\
\frac{\partial Q(\beta)}{\partial \beta} &= -X'(Y - X\beta) + (Y - X\beta)'(-X)
\end{aligned}
$$

Reference

- Abadir, K. M., and Magnus, J. R., 2005. *Matrix Algebra*. Cambridge University Press, Cambridge.
- Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear System Theory*. Princeton University Press, Princeton.
- Magnus, J. R. & Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Willey & Sons, New York.
- Hansen, B., *Econometrics*, August 2020. Appendix A Matrix Algebra. Online lecture note.