Advanced Econometrics

Lecture 2: Review of Matrix Algebra

Instructor: Ma, Jun

Renmin University of China

September 18, 2021

Notation

- A **scalar** a is a single number.
- A **vector** $\boldsymbol{a}$ is $k \times 1$ list of numbers, typically arranged in a column. We write this as
$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}.$$
- Equivalently, a vector $\boldsymbol{a}$ is an element of $\mathbb{R}^k$.
- A **matrix** $\boldsymbol{A}$ is a $k \times r$ rectangular array of numbers, written as
$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kr} \end{bmatrix}.$$

  By convention, a_{ij} refers to the element in the i^{th} row and j^{th} column of $\boldsymbol{A}$. Sometimes a matrix $\boldsymbol{A}$ is denoted by the symbol (a_{ij}).

- A matrix can be written as a set of column vectors or as a set of row vectors. That is,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_r \end{bmatrix} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_k' \end{bmatrix}$$

  where

$$\boldsymbol{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ki} \end{bmatrix}$$

are column vectors and

$$\alpha_j' = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jr} \end{bmatrix}$$

are row vectors.

- The **transpose** of a matrix $\boldsymbol{A}$, denoted as $\boldsymbol{A}'$, is obtained by flipping the matrix on its diagonal:

$$\boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1r} & a_{2r} & \cdots & a_{kr} \end{bmatrix}.$$

- If $\boldsymbol{A}$ is $k \times r$, then $\boldsymbol{A}'$ is $r \times k$. If $\boldsymbol{a}$ is a $k \times 1$ vector, then $\boldsymbol{a}'$ is an $1 \times k$ row vector.
- A matrix is **square** if $k = r$. A matrix is **symmetric** if $\boldsymbol{A} = \boldsymbol{A}'$. A square matrix is **diagonal** if the off-diagonal elements are all zero, so that $a_{ij} = a_{ji}$. A square matrix is **upper (lower) diagonal** if all elements below (above) the diagonal equal zero.

- An important diagonal matrix is the **identity** matrix, which has ones on the diagonal. The $k \times k$ identity matrix is denoted as

$$\boldsymbol{I}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

- A **partitioned matrix** takes the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1r} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2r} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}_{k1} & \boldsymbol{A}_{k2} & \cdots & \boldsymbol{A}_{kr} \end{bmatrix}.$$

Matrix addition

- If the matrices $\boldsymbol{A} = \left(a_{ij}\right)$ and $\boldsymbol{B} = \left(b_{ij}\right)$ are of the same order, we define the sum
$$\boldsymbol{A} + \boldsymbol{B} = \left(a_{ij} + b_{ij}\right).$$

- Matrix addition follows the commutative and associative laws:
$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A};\ \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}.$$

Matrix multiplication

- If $\boldsymbol{A}$ is $k \times r$ and c is scalar, we define the product as

$$\boldsymbol{A}c = c\boldsymbol{A} = \left(a_{ij}c\right).$$

- If $\boldsymbol{a}$ and $\boldsymbol{b}$ are both $k \times 1$, then their **inner product** is

$$\boldsymbol{a}'\boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots a_k b_k = \sum_{j=1}^{k} a_j b_j.$$

- We say that the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are **orthogonal** if $\boldsymbol{a}'\boldsymbol{b} = 0$.

- If $\boldsymbol{A}$ is $k \times r$ and $\boldsymbol{B}$ is $r \times s$ so that the number of columns of $\boldsymbol{A}$ equals the number of rows of $\boldsymbol{B}$ we say that $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable. In this event, the **matrix product** $\boldsymbol{AB}$ is defined.
- Writing $\boldsymbol{A}$ as a set of row vectors and $\boldsymbol{B}$ as a set of column vectors (each of length r), then the matrix product is defined as

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_k' \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_s \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1'\boldsymbol{b}_s \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2'\boldsymbol{b}_s \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_k'\boldsymbol{b}_1 & \boldsymbol{a}_k'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_k'\boldsymbol{b}_s \end{bmatrix}.$$

- Matrix multiplication is not commutative: in general $\boldsymbol{AB} \neq \boldsymbol{BA}$. However, it is associative and distributive:

$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C};\ \boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}.$$

- An alternative way to write the matrix product is to use matrix partitions:

$$\begin{aligned}\boldsymbol{AB} &= \begin{bmatrix} \boldsymbol{A_1} & \boldsymbol{A_2} & \cdots & \boldsymbol{A_r} \end{bmatrix} \begin{bmatrix} \boldsymbol{B_1} \\ \boldsymbol{B_2} \\ \vdots \\ \boldsymbol{B_r} \end{bmatrix} \\ &= \boldsymbol{A_1B_1} + \boldsymbol{A_2B_2} + \cdots + \boldsymbol{A_rB_r} \\ &= \sum_{j=1}^{r} \boldsymbol{A_jB_j}.\end{aligned}$$

- An important property of the identity matrix is that if $\boldsymbol{A}$ is $k \times r$, then $\boldsymbol{A}\boldsymbol{I}_r = \boldsymbol{A}$ and $\boldsymbol{I}_k\boldsymbol{A} = \boldsymbol{A}$.
- We say two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are **orthogonal** if $\boldsymbol{A}'\boldsymbol{B} = 0$. This means that all columns of $\boldsymbol{A}$ are orthogonal with all columns of $\boldsymbol{B}$.
- The $k \times r$ matrix $\boldsymbol{H}$, $r \leq k$, is called **orthonormal** if $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_r$. This means that the columns of $\boldsymbol{H}$ are mutually orthogonal, and each column is normalized to have unit length.

Trace

- The **trace** of a $k \times k$ square matrix $\boldsymbol{A}$ is the sum of its diagonal elements

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}.$$

- Some straightforward properties for square matrices$\boldsymbol{A}$ and $\boldsymbol{B}$ and scalar c are

$$\mathrm{tr}(c\boldsymbol{A}) = c\mathrm{tr}(\boldsymbol{A});\ \mathrm{tr}(\boldsymbol{A}') = \mathrm{tr}(\boldsymbol{A});\ \mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B});\ \mathrm{tr}(\boldsymbol{I}_k) = k.$$

- For $k \times r$ $\boldsymbol{A}$ and $r \times k$ $\boldsymbol{B}$ we have

$$\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$$

since

$$\begin{aligned}
\text{tr}(\boldsymbol{AB}) &= \text{tr}\begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1'\boldsymbol{b}_k \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2'\boldsymbol{b}_k \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_k'\boldsymbol{b}_1 & \boldsymbol{a}_k'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_k'\boldsymbol{b}_k \end{bmatrix} \\
&= \sum_{i=1}^{k} \boldsymbol{a}_i'\boldsymbol{b}_i \\
&= \sum_{i=1}^{k} \sum_{j=1}^{r} a_{ij} b_{ji} \\
&= \sum_{j=1}^{r} \sum_{i=1}^{k} b_{ji} a_{ij} \\
&= \text{tr}(\boldsymbol{BA}).
\end{aligned}$$

Rank and inverse

- The **rank** of the $k \times r$ matrix $(r \leq k)$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_r} \end{bmatrix}$$

  is the number of linearly independent columns, written as $\text{rank}(\boldsymbol{A})$. We say that $\boldsymbol{A}$ has full rank if $\text{rank}(\boldsymbol{A}) = r$.
- A $k \times k$ square matrix $\boldsymbol{A}$ is said to be **nonsingular** if it is has full rank, e.g. $\text{rank}(\boldsymbol{A}) = k$. This means that there is no $k \times 1$ $\boldsymbol{c} \neq \boldsymbol{0}$ such that $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$.
- If $k \times k$ square matrix $\boldsymbol{A}$ is nonsingular then there exists a unique $k \times k$ matrix $\boldsymbol{A}^{-1}$ called the **inverse** of $\boldsymbol{A}$ which satisfies

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k.$$

- For non-singular $\boldsymbol{A}$ and $\boldsymbol{C}$, some important properties include

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^{-1} &= \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k \\
(\boldsymbol{A}^{-1})' &= (\boldsymbol{A}')^{-1} \\
(\boldsymbol{A}\boldsymbol{C})^{-1} &= \boldsymbol{C}^{-1}\boldsymbol{A}^{-1} \\
(\boldsymbol{A} + \boldsymbol{C})^{-1} &= \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{C}^{-1} \\
\boldsymbol{A}^{-1} - (\boldsymbol{A} + \boldsymbol{C})^{-1} &= \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{A}^{-1}.
\end{aligned}$$

- If a $k \times k$ matrix $\boldsymbol{H}$ is orthonormal (so that $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$), then $\boldsymbol{H}$ is non-singular and $\boldsymbol{H}^{-1} = \boldsymbol{H}'$. Furthermore, $\boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}_k$ and $\boldsymbol{H}'^{-1} = \boldsymbol{H}$.
- The following fact about inverting partitioned matrices is quite useful.

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11\cdot 2}^{-1} & -\boldsymbol{A}_{11\cdot 2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22\cdot 1}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{22\cdot 1}^{-1} \end{bmatrix},$$

where $\boldsymbol{A}_{11\cdot 2} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22\cdot 1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$.

- For any $k \times r$ matrix $\boldsymbol{A}$, the linear sub-space $\{\boldsymbol{x} \in \mathbb{R}^r : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$ is called the **null space**. The linear sub-space $\{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^r\}$ is called the **column space**, i.e., the set of vectors spanned by the columns of $\boldsymbol{A}$.

- Suppose we have a $n \times k$ matrix $\boldsymbol{X}$ with $n \geq k$. We have the following result
$$\operatorname{rank}(\boldsymbol{X}\boldsymbol{X}') = \operatorname{rank}(\boldsymbol{X}'\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{X}) \leq k.$$

- $\operatorname{rank}(\boldsymbol{X})$ is equal to the difference between k and the dimension of its null space. The null spaces of $\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{X}$ are the same: if $\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{0}$, then $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{0}$; if $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{0}$, then $\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha} = \|\boldsymbol{X}\boldsymbol{\alpha}\|^2 = 0$ and therefore $\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{0}$. Therefore, $\operatorname{rank}(\boldsymbol{X}'\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{X})$. Similarly, $\operatorname{rank}(\boldsymbol{X}\boldsymbol{X}') = \operatorname{rank}(\boldsymbol{X}')$. Transposing a matrix does not change its rank: $\operatorname{rank}(\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{X}')$.
- Similarly, we can show the following result: let $\boldsymbol{Q}$, $\boldsymbol{P}$ be non-singular matrices and $\boldsymbol{A}$ be a $k \times r$ matrix with rank $\operatorname{rank}(\boldsymbol{A})$, then
$$\operatorname{rank}(\boldsymbol{P}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{Q}) = \operatorname{rank}(\boldsymbol{P}\boldsymbol{A}\boldsymbol{Q}) = \operatorname{rank}(\boldsymbol{A}).$$

Determinant

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

$$\det \boldsymbol{A} = \sum_{\pi} \varepsilon_\pi a_{1j_1} a_{2j_2} \cdots a_{kj_k}.$$

- For example, if $\boldsymbol{A}$ is 2×2 then the two permutations of $(1, 2)$ are $(1, 2)$ and $(2, 1)$ for which $\varepsilon_{(1,2)} = 1$ and $\varepsilon_{(2,1)} = -1$. Thus

$$\begin{aligned} \det \boldsymbol{A} &= \varepsilon_{(1,2)} a_{11} a_{22} + \varepsilon_{(2,1)} a_{21} a_{12} \\ &= a_{11} a_{22} - a_{12} a_{21}. \end{aligned}$$

Theorem (A.7.1, Hansen)

Let $\boldsymbol{A} = (a_{ij})$ *be a* $k \times k$ *matrix. Properties of the determinant*

1. $\det \boldsymbol{A} = \det(\boldsymbol{A}')$
2. $\det(c\boldsymbol{A}) = c^k \det \boldsymbol{A}$
3. $\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{B}\boldsymbol{A}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$
4. $\det(\boldsymbol{A}^{-1}) = (\det \boldsymbol{A})^{-1}$
5. $\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A})(\det \boldsymbol{D})$ and $\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A})(\det \boldsymbol{D})$
6. $\det \boldsymbol{A} \neq 0$ if and only if $\boldsymbol{A}$ is nonsingular
7. If $\boldsymbol{A}$ is triangular (upper or lower), then $\det \boldsymbol{A} = \prod_{i=1}^{k} a_{ii}$
8. If $\boldsymbol{A}$ is orthonormal, then $\det \boldsymbol{A} = \pm 1$.

Eigenvalues

- The characteristic equation of a $k \times k$ square matrix $\boldsymbol{A}$ is

$$\det(\lambda \boldsymbol{I}_k - \boldsymbol{A}) = 0.$$

  The left side is a polynomial of degree k in λ so it has exactly k roots, which are not necessarily distinct and may be real or complex. They are called the **characteristic roots** or **eigenvalues** of $\boldsymbol{A}$.
- If is an eigenvalue of $\boldsymbol{A}$ then $\lambda \boldsymbol{I}_k - \boldsymbol{A}$ is singular so there exists a non-zero vector $\boldsymbol{h}$ such that $(\lambda \boldsymbol{I}_k - \boldsymbol{A}) \boldsymbol{h} = \boldsymbol{0}$ or $\boldsymbol{A}\boldsymbol{h} = \boldsymbol{h}\lambda$. The vector $\boldsymbol{h}$ is called a **characteristic vector** or **eigenvector** of $\boldsymbol{A}$ corresponding to λ. They are typically normalized so that $\boldsymbol{h}'\boldsymbol{h} = 1$ and thus $\lambda = \boldsymbol{h}'\boldsymbol{A}\boldsymbol{h}$.
- Set $\boldsymbol{H} = [\boldsymbol{h}_1 \cdots \boldsymbol{h}_k]$ and $\boldsymbol{\Lambda} = \text{diag}\left\{\lambda_1, \ldots, \lambda_k\right\}$. A matrix expression is $\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}$.

Theorem (A.8.1, Hansen)

Properties of eigenvalues. Let λ_i*and* $\boldsymbol{h}_i, i = 1, \ldots, k$, *denote the k eigenvalues and eigenvectors of a square matrix* $\boldsymbol{A}$.

1. $\det(\boldsymbol{A}) = \prod_{i=1}^{k} \lambda_i$
2. $\text{tr}(\boldsymbol{A})=\sum_{i=1}^{k} \lambda_i$
3. $\boldsymbol{A}$ is non-singular if and only if all its eigenvalues are non-zero.
4. If $\boldsymbol{A}$ has distinct eigenvalues, there exists a nonsingular matrix $\boldsymbol{P}$, such that $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{\Lambda}\boldsymbol{P}$ and $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} = \boldsymbol{\Lambda}$.
5. The non-zero eigenvalues of $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are identical.
6. If $\boldsymbol{B}$ is non-singular then $\boldsymbol{A}$ and $\boldsymbol{B}^{-1}\boldsymbol{AB}$ have the same eigenvalues.
7. If $\boldsymbol{Ah} = \boldsymbol{h}\lambda$ then $(\boldsymbol{I} - \boldsymbol{A}) = \boldsymbol{h}(1 - \lambda)$. So $\boldsymbol{I} - \boldsymbol{A}$ has the eigenvalue $1 - \lambda$ and associated eigenvector $\boldsymbol{h}$.

- Most eigenvalue applications in econometrics concern the case where the matrix $\boldsymbol{A}$ is real and symmetric. In this case all eigenvalues of $\boldsymbol{A}$ are real and its eigenvectors are mutually orthogonal. Thus $\boldsymbol{H}$ is orthonormal so $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$ and $\boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}_k$. When the eigenvalues are all real it is conventional to write them in descending order $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k$.
- **Spectral Decomposition**. If $\boldsymbol{A}$is a $k \times k$ real symmetric matrix, then $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{H}$ contains the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues on the diagonal. The eigenvalues are all real and the eigenvector matrix satisfies $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$. The decomposition can be alternatively written as $\boldsymbol{H}'\boldsymbol{A}\boldsymbol{H} = \boldsymbol{\Lambda}$.
- If $\boldsymbol{A}$ is real, symmetric, and invertible, then by the spectral decomposition and the properties of orthonormal matrices, $\boldsymbol{A}^{-1} = \boldsymbol{H}'^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{-1} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}'$. Thus the columns of $\boldsymbol{H}$ are also the eigenvectors of $\boldsymbol{A}^{-1}$, and its eigenvalues are $\lambda_1^{-1}, \ldots, \lambda_k^{-1}$.

Positive definite matrices

- We say that a $k \times k$ real symmetric square matrix $\boldsymbol{A}$ is positive semi-definite if for all $\boldsymbol{c} \neq \boldsymbol{0}$, $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \geq 0$. This is written as $\boldsymbol{A} \geq 0$.
- We say that $\boldsymbol{A}$ is positive definite if for all $\boldsymbol{c} \neq \boldsymbol{0}$, $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} > 0$. This is written as $\boldsymbol{A} > 0$.

Theorem (A.9.1, Hansen)

Properties of positive semi-definite matrices

1. If $\boldsymbol{A} = \boldsymbol{G}'\boldsymbol{B}\boldsymbol{G}$ with $\boldsymbol{B} \geq 0$ and some matrix $\boldsymbol{G}$, then $\boldsymbol{A}$ is positive semi-definite.
2. If $\boldsymbol{A}$is positive definite, then $\boldsymbol{A}$ is non-singular. Furthermore, $\boldsymbol{A}^{-1} > 0$.
3. $\boldsymbol{A} > 0$ if and only if it is symmetric and all its eigenvalues are positive.
4. By the spectral decomposition, $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$ and $\boldsymbol{\Lambda}$ is diagonal with non-negative diagonal elements. All diagonal elements of $\boldsymbol{\Lambda}$ are strictly positive if and only if $\boldsymbol{A} > 0$.
5. The rank of $\boldsymbol{A}$ equals the number of strictly positive eigenvalues.
6. If $A > 0$ then $A^{-1} = \boldsymbol{H}\boldsymbol{\Lambda}^{\boldsymbol{-1}}\boldsymbol{H}'$.
7. If $A \geq 0$ we can find a matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}'$.We call $\boldsymbol{B}$ a matrix square root of $\boldsymbol{A}$ and is typically written as $\boldsymbol{B} = \boldsymbol{A}^{1/2}$. The matrix $\boldsymbol{B}$ need not be unique. One matrix square root is obtained using the spectral decomposition $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$. Then $\boldsymbol{B} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{H}'$ is itself symmetric and positive definite and satisfies $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}$.

Idempotent matrices

- A $k \times k$ square matrix $\boldsymbol{A}$ is idempotent if $\boldsymbol{AA} = \boldsymbol{A}$. For example, the following matrix is idempotent

$$\boldsymbol{A} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

- If $\boldsymbol{A}$ is idempotent and symmetric with rank r, then it has r eigenvalues which equal 1 and $k - r$ eigenvalues which equal 0. To see this, by the spectral decomposition we can write $\boldsymbol{A} = \boldsymbol{H\Lambda H'}$ where $\boldsymbol{H}$ is orthonormal and $\boldsymbol{\Lambda}$ contains the eigenvalues. Then

$$\boldsymbol{A} = \boldsymbol{AA} = \boldsymbol{H\Lambda H'H\Lambda H'} = \boldsymbol{H\Lambda}^2\boldsymbol{H'}.$$

- We deduce that $\boldsymbol{\Lambda}^2 = \boldsymbol{\Lambda}$ and $\lambda_i^2 = \lambda_i$ for $i = 1, ..., k$. Hence λ_i must equal either 0 or 1. Since the rank of $\boldsymbol{A}$ is r, and the rank equals the number of positive eigenvalues, it follows that

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_{k-r} \end{bmatrix}.$$

- $\operatorname{tr}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$ and $\boldsymbol{A}$ is positive semi-definite.

- If $\boldsymbol{A}$ is idempotent and symmetric with rank $r < k$ then it does not possess an inverse, but its Moore-Penrose generalized inverse takes the simple form $\boldsymbol{A}^{-} = \boldsymbol{A}$.
- If $\boldsymbol{A}$ is idempotent then $\boldsymbol{I} - \boldsymbol{A}$ is also idempotent.
- One useful fact is that if $\boldsymbol{A}$ is idempotent then

$$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \leq \boldsymbol{c}'\boldsymbol{c} \text{ and } \boldsymbol{c}'(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{c} \leq \boldsymbol{c}'\boldsymbol{c}.$$

To see this, note that both $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}$ and $\boldsymbol{c}'(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{c}$ are non-negative and

$$\boldsymbol{c}'\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} + \boldsymbol{c}'(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{c}.$$

Matrix calculus

- Let $\boldsymbol{x} = (x_1, \ldots, x_k)'$ be $k \times 1$ and $g : \mathbb{R}^k \to \mathbb{R}$. We adopt the following notational convention: the vector derivative is

$$\frac{\partial}{\partial \boldsymbol{x}} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) \\ \vdots \\ \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix}$$

and

- $$\frac{\partial}{\partial \boldsymbol{x}'} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix}.$$

- Let $\boldsymbol{A} = \left(a_{ij}\right)_{m \times n}$ be a $m \times n$ matrix and $g : \mathbb{R}^{m \times n} \to \mathbb{R}$. The derivative of $g\left(\boldsymbol{A}\right)$ with respect to $\boldsymbol{A}$ is (by convention)

$$\frac{\partial}{\partial \boldsymbol{A}} g\left(\boldsymbol{A}\right) = \begin{pmatrix} \frac{\partial g(\boldsymbol{A})}{\partial a_{11}} & \cdots & \frac{\partial g(\boldsymbol{A})}{\partial a_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g(\boldsymbol{A})}{\partial a_{m1}} & \cdots & \frac{\partial g(\boldsymbol{A})}{\partial a_{mn}} \end{pmatrix}.$$

Theorem (A.15.1 Hansen)

Properties of matrix derivatives

1. $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{a}) = \boldsymbol{a}$
2. $\frac{\partial}{\partial \boldsymbol{x}'}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$ and $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}') = \boldsymbol{A}'$
3. $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}$
4. $\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}'}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A} + \boldsymbol{A}'$
5. $\frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \boldsymbol{B}'$
6. $\frac{\partial}{\partial \boldsymbol{A}}\log\det(\boldsymbol{A}) = (\boldsymbol{A}^{-1})'$

Let $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$. Then

$$\frac{\partial\left(\boldsymbol{a}'\boldsymbol{x}\right)}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial(\boldsymbol{a}'\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial(\boldsymbol{a}'\boldsymbol{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial(a_1x_1+\ldots+a_nx_n)}{\partial x_1} \\ \vdots \\ \frac{\partial(a_1x_1+\ldots+a_nx_n)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \boldsymbol{a}.$$

$$\frac{\partial\left(\boldsymbol{a}'\boldsymbol{x}\right)}{\partial \boldsymbol{x}'} = \begin{pmatrix} \frac{\partial(\boldsymbol{a}'\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial(\boldsymbol{a}'\boldsymbol{x})}{\partial x_n} \end{pmatrix} = (a_1, \ldots, a_n) = \boldsymbol{a}'.$$

- Let $\boldsymbol{A}$ be an $m \times n$ matrix,

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_m' \end{pmatrix},$$

  where $\boldsymbol{a}_j \in \mathbb{R}^n$ for $j = 1, \ldots, m$.

$$\begin{aligned} \frac{\partial \left(\boldsymbol{A}\boldsymbol{x} \right)}{\partial \boldsymbol{x}'} =& \begin{pmatrix} \frac{\partial \left(\boldsymbol{a}_1' \boldsymbol{x} \right)}{\partial \boldsymbol{x}'} \\ \vdots \\ \frac{\partial \left(\boldsymbol{a}_m' \boldsymbol{x} \right)}{\partial \boldsymbol{x}'} \end{pmatrix} \\ =& \begin{pmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_m' \end{pmatrix} \\ =& \boldsymbol{A}. \end{aligned}$$

- Similarly, $\frac{\partial}{\partial \boldsymbol{x}} \left(\boldsymbol{x}' \boldsymbol{A}' \right) = \boldsymbol{A}'$.
- Using "multiplication rule",

$$\frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{x}' \boldsymbol{A} \boldsymbol{x}) = \frac{\partial \boldsymbol{x}'}{\partial \boldsymbol{x}} \boldsymbol{A} \boldsymbol{x} + \frac{\partial \boldsymbol{x}' \boldsymbol{A}'}{\partial \boldsymbol{x}} \boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{A}') \boldsymbol{x}.$$

Vector Norms

Given any vector space V (such as Euclidean space $\mathbb{R}^m$) a **norm** on V is a function $\rho : \mathrm{V} \longrightarrow \mathbb{R}$ with the properties

1. $\rho(c\boldsymbol{a}) = |c|\rho(\boldsymbol{a})$ for any real number c and $\boldsymbol{a} \in \mathrm{V}$
2. $\rho(\boldsymbol{a} + \boldsymbol{b}) \leq \rho(\boldsymbol{a}) + \rho(\boldsymbol{b})$
3. If $\rho(\boldsymbol{a}) = 0$ then $\boldsymbol{a} = \mathbf{0}$

- The typical norm used for $\mathbb{R}^m$ is the **Euclidean norm**

$$\begin{aligned} \| \boldsymbol{a} \| &= (\boldsymbol{a}'\boldsymbol{a})^{1/2} \\ &= (\sum_{i=1}^{m} a_i^2)^{1/2}. \end{aligned}$$

- The p−norm ($p \geq 1$)

$$\| \boldsymbol{a} \|_p = (\sum_{i=1}^{m} |a_i|^p)^{1/p}.$$

- Special cases are the Euclidean norm and the 1-norm:

$$\| \boldsymbol{a} \|_1 = \sum_{i=1}^{m} |a_i|.$$

- The “max-norm”

$$\| \boldsymbol{a} \|_\infty = \max(|a_1|, \ldots, |a_m|).$$

- **Jensen's Inequality.** If $g(\cdot) : \mathbb{R} \longrightarrow \mathbb{R}$ is convex, then for any non-negative weights a_j such that $\sum_{j=1}^{m} a_j = 1$, and any real numbers $\boldsymbol{x}_j$

$$g(\sum_{j=1}^{m} a_j x_j) \leq \sum_{j=1}^{m} a_j g(x_j)$$

- In particular, setting $a_j = 1/m$, then

$$g(\frac{1}{m} \sum_{j=1}^{m} x_j) \leq \frac{1}{m} \sum_{j=1}^{m} g(x_j)$$

- If $g(\cdot) : \mathbb{R} \longrightarrow \mathbb{R}$ is concave then the inequalities are reversed.

- **Hölder's Inequality.** If $p > 1, q > 1$, and $1/p + 1/q = 1$, then for any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,
$$\sum_{j=1}^{m} | a_j b_j | \leq \| \boldsymbol{a} \|_p \| \boldsymbol{b} \|_q .$$

- **Minkowski's Inequality.** For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, if $p \geq 1$, then
$$\| \boldsymbol{a} + \boldsymbol{b} \|_p \leq \| \boldsymbol{a} \|_p + \| \boldsymbol{b} \|_p .$$

- **Schwarz Inequality.** For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,
$$| \boldsymbol{a}'\boldsymbol{b} | \leq \| \boldsymbol{a} \| \| \boldsymbol{b} \| .$$

Matrix Norms

- The **Frobenius norm** of an $m \times k$ matrix $\boldsymbol{A}$ is the Euclidean norm applied to its elements:

$$\begin{aligned} \| \boldsymbol{A} \|_F &= (\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A}))^{1/2} \\ &= \left(\sum_{i=1}^{m} \sum_{j=1}^{k} a_{ij}^2 \right)^{1/2} . \end{aligned}$$

- When $m \times m$ $\boldsymbol{A}$ is real symmetric, then

$$\| \boldsymbol{A} \|_F = \left(\sum_{l=1}^{m} \lambda_l^2 \right)^{1/2} ,$$

where λ_l, $l = 1, \ldots, m$ are the eigenvalues of $\boldsymbol{A}$. To see this,

$$\| \boldsymbol{A} \|_F = (\mathrm{tr}\,(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'))^{1/2} = (\mathrm{tr}\,(\boldsymbol{\Lambda}\boldsymbol{\Lambda}))^{1/2} = \left(\sum_{l=1}^{m} \lambda_l^2 \right)^{1/2} .$$

- For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$\| \boldsymbol{a}\boldsymbol{b}' \|_F = \mathrm{tr}\,(\boldsymbol{b}\boldsymbol{a}'\boldsymbol{a}\boldsymbol{b}')^{1/2} = (\boldsymbol{b}'\boldsymbol{b}\boldsymbol{a}'\boldsymbol{a})^{1/2} = \| \boldsymbol{a} \| \| \boldsymbol{b} \|$$

and $\| \boldsymbol{a}\boldsymbol{a}' \|_F = \| \boldsymbol{a} \|^2$.

- The **spectral norm** of an $m \times k$ matrix is

$$\| \boldsymbol{A} \|_2 = (\lambda_{max}(\boldsymbol{A}'\boldsymbol{A}))^{1/2},$$

where $\lambda_{\max}(\boldsymbol{B})$ denotes the largest eigenvalue of the symmetric matrix $\boldsymbol{B}$.

- If $\boldsymbol{A}$ is $m \times m$ and symmetric with eigenvalues λ_j then

$$\| \boldsymbol{A} \|_2 = \max_{j \le m} | \lambda_j | .$$

- Suppose $\boldsymbol{A}$ is $m \times k$ with rank r,

$$\| \boldsymbol{A} \|_2 \le \| \boldsymbol{A} \|_F \text{ and } \| \boldsymbol{A} \|_F \le \sqrt{r} \, \| \boldsymbol{A} \|_2 .$$

- Given any vector norm $\|\cdot\|$, the **induced matrix norm** is
-

$$\| \boldsymbol{A} \| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{Ax} \| = \sup_{\boldsymbol{x}\neq 0} \frac{\| \boldsymbol{Ax} \|}{\| \boldsymbol{x} \|}.$$

- The triangle inequality is satisfied:

$$\| \boldsymbol{A}+\boldsymbol{B} \| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{Ax}+\boldsymbol{Bx} \| \leq \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{Ax} \| + \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{Bx} \| = \| \boldsymbol{A} \| + \| \boldsymbol{B} \| .$$

- For any vector $\boldsymbol{x}$, $\| \boldsymbol{Ax} \| \leq \| \boldsymbol{A} \| \| \boldsymbol{x} \|$. The induced matrix norm satisfies this property which is a matrix form of the Schwarz inequality:

$$\| \boldsymbol{AB} \| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{ABx} \| \leq \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{A} \| \| \boldsymbol{Bx} \| = \| \boldsymbol{A} \| \| \boldsymbol{B} \| .$$

- The matrix norm induced by the Euclidean vector norm is the spectral norm

$$\sup_{\boldsymbol{x}'\boldsymbol{x}=1} \| \boldsymbol{Ax} \|^2 = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{Ax} = \lambda_{max}(\boldsymbol{A}'\boldsymbol{A}) = \| \boldsymbol{A} \|_2^2$$