# Lecture 0. Useful matrix theory

## Basic definitions

A $p \times 1$ vector $\mathbf{a}$ is a column vector of real (or complex) numbers $a_1, \ldots, a_p$,

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}$$

By default, a vector is understood as a column vector. The collection of such column vectors is called the vector space of dimension $p$ and is denoted by $\mathbb{R}^p$ (or $\mathbb{C}^p$). An $1 \times p$ row vector is obtained by the transpose of the column vector and is

$$\mathbf{a}' = (a_1, \ldots, a_p).$$

A $p \times q$ matrix $\mathbf{A}$ is a rectangular array of real (or complex) numbers $a_{11}, a_{12}, \ldots, a_{pq}$ written as

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{bmatrix} = (a_{ij})_{(p \times q)},$$

so that $a_{ij}$ is the element in the $i$th row and $j$th column of $\mathbf{A}$. The following definitions are basic:

1. $\mathbf{A}$ is a square matrix if $p = q$.
2. $\mathbf{A} = \mathbf{0}$ is a zero matrix if $a_{ij} = 0$ for all $i, j$.
3. $\mathbf{A} = \mathbb{I} = \mathbb{I}_p$ is the identity matrix of order $p$ if $\mathbf{A}$ is a square matrix of order $p$ and $a_{ij} = 0$ for $i \neq j$ and $a_{ij} = 1$ for $i = j$.
4. The transpose of a $p \times q$ matrix $\mathbf{A}$, denoted by $\mathbf{A}'$ or $\mathbf{A}^T$, is the $q \times p$ matrix by interchanging the rows and columns of $\mathbf{A}$, $\mathbf{A}' = (a_{ji})$. Note that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ and $(\mathbf{A}')' = \mathbf{A}$.
5. A square matrix $\mathbf{A}$ is called symmetric if $\mathbf{A} = \mathbf{A}'$.
6. A square matrix $\mathbf{A}$ is called skew-symmetric if $\mathbf{A} = -\mathbf{A}'$.
7. A square matrix $\mathbf{A}$ is called diagonal if $a_{ij} = 0$ for all $i \neq j$ and is denoted by $\mathrm{diag}(a_{11}, \ldots, a_{pp})$.
8. A $p \times q$ matrix $\mathbf{A}$ is sometimes written as $\mathbf{A}_{(p \times q)}$ to emphasize its size.
9. For $\mathbf{A}_{(p \times q)}$ and $\mathbf{B}_{(q \times m)}$,
$$\mathbf{C} = \mathbf{AB} = (c_{ij})_{(p \times m)},$$
where $c_{ij} = \sum_{k=1}^{q} a_{ik} b_{kj}$ is the inner product between the $i$th row of $\mathbf{A}$ and the $j$th column of $\mathbf{B}$.

10. A square matrix $\mathbf{A}$ is called orthogonal if $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbb{I}$.

11. A square matrix $\mathbf{A}$ is called idempotent if $\mathbf{A}^2 = \mathbf{A}$.

*Example* 1. The following matrices appear frequently in multivariate statistical analysis.

- A rotation matrix $\mathbf{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ is orthogonal.
- The centering matrix $\mathbf{C} = \mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ and the projection matrix $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ are idempotent and symmetric.
- The householder matrix $H = \mathbb{I} - 2\mathbf{u}\mathbf{u}'$, where $\|\mathbf{u}\| = 1$, is orthogonal and symmetric.

## Determinant and inverse matrix

The determinant of a square matrix $\mathbf{A}$ is denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$.

1. For a $2 \times 2$ square matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $|\mathbf{A}| = ad - bc$.
2. For $\mathbf{A}_{(p\times p)}$, $|\alpha\mathbf{A}| = \alpha^p|\mathbf{A}|$.
3. For $\mathbf{A}_{(p\times p)}$, $\mathbf{B}_{(p\times p)}$, $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.
4. For $\mathbf{A}_{(p\times q)}$, $\mathbf{B}_{(q\times p)}$, $\det(\mathbb{I}_p + \mathbf{AB}) = \det(\mathbb{I}_q + \mathbf{BA})$.
5. If $\mathbf{A}$ is orthogonal, then $\det(\mathbf{A}) = \pm 1$.
6. A square matrix $\mathbf{A}$ is called nonsingular if $\det(\mathbf{A}) \neq 0$.

If $\mathbf{A}$ is nonsingular then there exists the inverse matrix of $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, which satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbb{I}$. Some basic results include:

1. $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbb{I}$.
2. $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$.
3. If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular, $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
4. If $\mathbf{A}$ is an orthogonal matrix, $\mathbf{A}^{-1} = \mathbf{A}'$.

**Lemma 1** (Woodbury's formula)**.** *Suppose* $\mathbf{A}_{(m\times m)}$ *is nonsingular. For* $\mathbf{U}_{(m\times n)}$ *and* $\mathbf{V}_{(m\times n)}$ *where* $m > n$*, the inverse of* $\mathbf{A} + \mathbf{U}\mathbf{V}'$ *is*

$$(\mathbf{A} + \mathbf{U}\mathbf{V}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbb{I}_n + \mathbf{V}'\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}'\mathbf{A}^{-1}.$$

## Trace of a matrix

The trace of a square matrix $\mathbf{A}_{(p\times p)}$ is the sum of all diagonal elements and is

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

Several properties of trace that are worth mentioning include:

1. Trace is a linear map, i.e., for $\mathbf{A}_{(p\times p)}, \mathbf{B}_{(p\times p)}$ and $\alpha, \beta \in \mathbb{R}$, $\text{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\text{trace}(\mathbf{A}) + \beta\text{trace}(\mathbf{B})$.
2. $\text{trace}(\mathbf{A}) = \text{trace}(\mathbf{A}')$
3. For $\mathbf{A}_{(p\times m)}, \mathbf{B}_{(p\times m)}$, $\text{trace}(\mathbf{A}\mathbf{B}') = \text{trace}(\mathbf{B}'\mathbf{A})$

## Eigenvalues and eigenvectors

A square matrix $\mathbf{A}_{(p\times p)}$ has the characteristic equation given by $q_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbb{I}_p)$, which is the $p$th order polynomial function of $\lambda$. The roots of $q_{\mathbf{A}}(\lambda)$ are called latent roots (or eigenvalues) of $\mathbf{A}$. Each eigenvalue $\lambda_i$ has its corresponding non-zero vector $\mathbf{u}_i$, called an eigenvector, satisfying

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i.$$

Properties include:

1. $\det(\mathbf{A}) = \prod_{i=1}^{p} \lambda_i$.
2. $\text{trace}(\mathbf{A}) = \sum_{i=1}^{p} \lambda_i$.
3. If $\mathbf{A}$ is symmetric, then its eigenvalues are all real.

The eigen-decomposition of a symmetric matrix $\mathbf{A}_{(p\times p)}$ leads to at most $p$ distinct eigenvalues $\lambda_1, \dots, \lambda_p$ and their corresponding eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$, and is denoted by

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}',$$

where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ is the orthogonal matrix whose $j$th column is the $j$th eigenvector, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ is the diagonal matrix whose $j$th diagonal element is the $j$th eigenvalue. This decomposition is also called the spectral decomposition.

**Definition 1.** A symmetric matrix $\mathbf{A}$ is

- non-negative definite ($\mathbf{A} \geq 0$) if $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} x_i a_{ij} x_j \geq 0$ for all $\mathbf{x} \neq 0$.
- positive definite ($\mathbf{A} > 0$) if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

For a non-negative definite matrix $\mathbf{A}$ we can uniquely define $\mathbf{A}^{\frac{1}{2}}$ so that $\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \mathbf{A}$. Using the spectral decomposition $\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}'$, we have

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\Lambda^{\frac{1}{2}}\mathbf{U}',$$

where $\Lambda^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})$.

**Lemma 2.** *i. If $\mathbf{A}$ is positive definite then $\mathbf{A}$ is nonsingular, $\mathbf{A}^{-1}$ exists and $\mathbf{A}^{-1} > 0$.*

*ii. $\mathbf{A}$ is positive definite (non-negative definite) if and only if the eigenvalues of $\mathbf{A}$ are all positive (non-negative).*

*iii. If $\mathbf{A}$ is positive definite and has its eigen-decomposition $\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}'$ then*

$$\mathbf{A}^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}',$$

*where $\Lambda^{-1} = diag(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_p})$.*