Lecture 2 Matrix Algebra

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP533 Multivariate Analysis
Spring 2025

Table of Contents

Vector

- A *vector* $\mathbf{x}$ is an array of n numbers $x_1, x_2, \cdots, x_n$, i.e.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad or \quad \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

- A scaler factor of vector $\mathbf{x}$ is $c\mathbf{x} = [cx_1, cx_2, \cdots, cx_n]^T$.
- Two vectors $\mathbf{x}, \mathbf{y}$ can be added $\mathbf{x} + \mathbf{y} = [x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n]^T$.
- The *length* of a vector $\mathbf{x}$ is defined as $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$.
- We have $\|c\mathbf{x}\|_2 = |c|\|\mathbf{x}\|_2$.

Vector

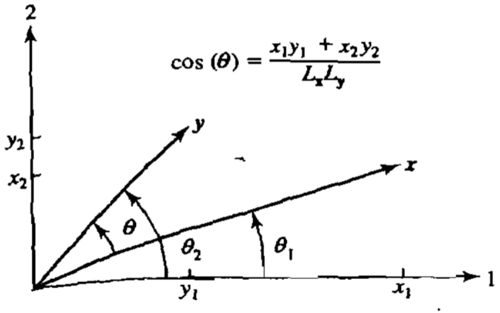

Figure 2.4 The angle θ between $\mathbf{x}' = [x_1, x_2]$ and $\mathbf{y}' = [y_1, y_2]$.

Vectors

- The *inner product* of two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

- The *angle* between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as $\theta = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right)$.
- When the angle $\theta = \pi/2$, i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, we say $\mathbf{x}$ and $\mathbf{y}$ are perpendicular, denoted as $\mathbf{x} \perp \mathbf{y}$.
- A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$ is said to be *linearly dependent* if there exist constants $c_1, c_2, \cdots, c_p$ not all zero, such that

$$\sum_{i=1}^{n} c_i \mathbf{x}_i = 0$$

Vectors

- What does *linearly independence* imply? Algebraically? Geometrically?
- The *projection* of a vector $\mathbf{x}$ onto another vector $\mathbf{y}$ is defined as

$$P_{\mathbf{y}}\mathbf{x} = \langle \mathbf{x}, \mathbf{y}^* \rangle \mathbf{y}^* = \left\langle \mathbf{x}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle \frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}$$

- Given a set of vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$, what is Gram–Schmidt process?
- What is it good for?

Cauchy-Schwarz Inequality

Theorem

Let $\mathbf{x}$ *and* $\mathbf{y}$ *be two* $p \times 1$ *vectors. Then*

$$\langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

with equality if and only if $\mathbf{x} = c\mathbf{y}$ *for some constant* c*.*

Matrix

- A *matrix*, $\mathbf{X}$, is any rectangle array of numbers with n rows and p columns

$$\mathbf{X} = [x_{ij}]_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}$$

- The *transpose* of $\mathbf{X}$ is an array by swapping rows and columns, denoted as $\mathbf{X}^T = [x_{ji}]_{p\times n}$.
- When $n = p$, $\mathbf{X}$ is a *square* matrix. Further $\mathbf{X}$ is *symmetric* if $\mathbf{X}^T = \mathbf{X}$.
- Diagonal matrix $\mathrm{diag}(\mathbf{x}) = [x_i\delta_{ij}]_{n\times n}$. Identity matrix $\mathbf{I} = [\delta_{ij}]_{n\times n}$.

Matrices

- The linear combination of matrices, $\mathbf{A}$ and $\mathbf{B}$: $a\mathbf{A} + b\mathbf{B} = [aa_{ij} + bb_{ij}]$.
- Two matrices $\mathbf{A}_{n\times k}$ and $\mathbf{B}_{k\times p}$ need to be size compatible to multiply

$$\mathbf{C} = [c_{ij}] = \mathbf{AB} = \left[\sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}\right]_{n\times p}, \quad c_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$$

- In general $\mathbf{AB}$ exists does not imply $\mathbf{BA}$ exists.
- If $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, then we say $\mathbf{B}$ is the *inverse* of square matrix $\mathbf{A}$, denoted as $\mathbf{B} = \mathbf{A}^{-1}$.
- A special case for square matrix is $\mathbf{Q}^{-1} = \mathbf{Q}^T$, i.e. $\mathbf{QQ}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$.
- We define *trace* of matrix $\mathbf{A}$ to be $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$. We have $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ whenever the matrix multiplication holds.

Eigen-decomposition: `eigen`

- A square matrix $\mathbf{A}_{n\times n}$ is said to have an *eigenvalue* λ with corresponding *eigenvector* $\mathbf{v}$ if
$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad \textit{usually } \|\mathbf{v}\|_2 = 1.$$
- If we organize eigenvectors as $\mathbf{P} = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ and eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_n]^T$. Then we have the eigendecomposition (a.k.a. spectral decomposition)
$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^{-1}, \quad \Lambda = \mathrm{diag}(\boldsymbol{\lambda})$$

Example 2.9 (Verifying eigenvalues and eigenvectors) Let

$$\mathbf{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$$

Then, since

$$\begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = 6\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$\lambda_1 = 6$ is an eigenvalue, and

$$\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

is its corresponding normalized eigenvector. You may wish to show that a second eigenvalue–eigenvector pair is $\lambda_2 = -4$, $\mathbf{e}_2' = [1/\sqrt{2}, 1/\sqrt{2}]$.

Table of Contents

Positive Definite Matrices (PSD)

- Many probability densities, e.g. normal, in multivariate analysis involve *quadric form* defined with a square matrix $\mathbf{A}$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i,j=1}^{n} x_i a_{ij} x_j$$

- The statistic distance is a quadratic form with $\mathbf{A} = \mathrm{diag}(\mathbf{s}^{-2})$.
- A symmetric matrix $\mathbf{A}$ is said to be *nonnegative* is $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$, denoted as $\mathbf{A} \geq 0$.
- If particularly $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ holds for any $\mathbf{x} \neq 0$, then we say $\mathbf{A}$ is *positive definite (PSD)*, denoted as $\mathbf{A} > 0$.
- Consider the spectral decomposition $\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^T$. $\mathbf{A}$ is PSD if and only if all eigenvalues are positive, i.e. $\Lambda > 0$.

Example 2.11 (A positive definite matrix and quadratic form) Show that the matrix for the following quadratic form is positive definite:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}\, x_1 x_2$$

To illustrate the general approach, we first write the quadratic form in matrix notation as

$$[x_1 \quad x_2]\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}'\mathbf{A}\,\mathbf{x}$$

By Definition 2A.30, the eigenvalues of $\mathbf{A}$ are the solutions of the equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$, or $(3 - \lambda)(2 - \lambda) - 2 = 0$. The solutions are $\lambda_1 = 4$ and $\lambda_2 = 1$. Using the spectral decomposition in (2-16), we can write

$$\underset{(2\times2)}{\mathbf{A}} = \lambda_1 \underset{(2\times1)}{\mathbf{e}_1}\underset{(1\times2)}{\mathbf{e}_1'} + \lambda_2 \underset{(2\times1)}{\mathbf{e}_2}\underset{(1\times2)}{\mathbf{e}_2'}$$

$$= 4\underset{(2\times1)}{\mathbf{e}_1}\underset{(1\times2)}{\mathbf{e}_1'} + \underset{(2\times1)}{\mathbf{e}_2}\underset{(1\times2)}{\mathbf{e}_2'}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the normalized and orthogonal eigenvectors associated with the eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$, respectively. Because 4 and 1 are scalars, premultiplication and postmultiplication of $\mathbf{A}$ by $\mathbf{x}'$ and $\mathbf{x}$, respectively, where $\mathbf{x}' = [x_1, x_2]$ is any *nonzero* vector, give

$$\underset{(1\times2)}{\mathbf{x}'}\underset{(2\times2)}{\mathbf{A}}\underset{(2\times1)}{\mathbf{x}} = 4\underset{(1\times2)}{\mathbf{x}'}\underset{(2\times1)}{\mathbf{e}_1}\underset{(1\times2)}{\mathbf{e}_1'}\underset{(2\times1)}{\mathbf{x}} + \underset{(1\times2)}{\mathbf{x}'}\underset{(2\times1)}{\mathbf{e}_2}\underset{(1\times2)}{\mathbf{e}_2'}\underset{(2\times1)}{\mathbf{x}}$$

$$= 4y_1^2 + y_2^2 \geq 0$$

with

$$y_1 = \mathbf{x}'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{x} \quad \text{and} \quad y_2 = \mathbf{x}'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{x}$$

Weighted Distance

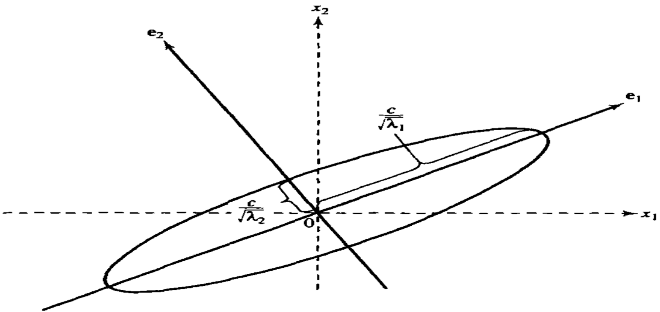

Figure 2.6 Points a constant distance c from the origin $(p = 2, 1 \leq \lambda_1 < \lambda_2)$.

Square-Root Matrix

- Consider the spectral decomposition for $\mathbf{A} > 0$

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^T, \quad \mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}, \; \Lambda = \mathrm{diag}(\{\lambda_\ell\}), \; \lambda_\ell > 0$$

- Then we have $\mathbf{A}^k = \mathbf{P}\Lambda^k\mathbf{P}^T$ for $k \in \mathbb{Z}$ with $\Lambda^k = \mathrm{diag}(\{\lambda_\ell^k\})$.
- We define *square-root* matrix of $\mathbf{A}$ as

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{P}\Lambda^{\frac{1}{2}}\mathbf{P}^T$$

- What is $(\mathbf{A}^{\frac{1}{2}})^T$? $(\mathbf{A}^{\frac{1}{2}})^{-1}$?
- Cholesky decomposition $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.

Maximization of Quadratic Forms on Unit Sphere

Theorem

Let $\mathbf{A}_{p\times p}$ be a PSD with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ with associated normalized eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$. Then

$$\max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_1, \quad \min_{\mathbf{x}\neq 0} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_p$$

Moreover,

$$\max_{\mathbf{x}\perp\mathbf{v}_1,\cdots,\mathbf{v}_k} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_{k+1}$$

Table of Contents

Random Vectors and Matrices

- Random vectors and matrices have elements as random variables (r.v.).
- We define the *expectation* of random matrix $\mathbf{X}$ as $\mathrm{E}[\mathbf{X}] = [\mathrm{E}X_{ij}]$ where

$$\mathrm{E}X_{ij} = \begin{cases} \int x_{ij} f_{ij}(x_{ij}) dx_{ij}, & \textit{if } X_{ij} \textit{ is continuous r.v. with pdf } f_{ij} \\ \sum x_{ij} p_{ij}(x_{ij}), & \textit{if } X_{ij} \textit{ is discrete r.v. with pmf } f_{ij} \end{cases}$$

- Expectation is a linear operator. For constants $c_1, \cdots, c_k$ and random matrices $\mathbf{X}_1, \cdots, \mathbf{X}_k$

$$\mathrm{E}\left[\sum_{\ell=1}^{k} c_\ell \mathbf{X}_\ell\right] = \sum_{\ell=1}^{k} c_\ell \mathrm{E}[\mathbf{X}_\ell]$$

- For deterministic matrices $\mathbf{A}, \mathbf{B}$, $\mathrm{E}[\mathbf{AXB}] = \mathbf{A}\mathrm{E}[\mathbf{X}]\mathbf{B}$.

Mean Vectors and Covariance Matrices

- For random vector $\mathbf{X} = [X_1, \cdots, X_p]^T$, we can define the *mean* vector $\mathrm{E}\mathbf{X} = [\mathrm{E}X_1, \cdots, \mathrm{E}X_p]^T$, often denoted as μ.
- We also define the covariance matrix of $\mathbf{X}$ as (often denoted as Σ)

$$\mathrm{Cov}[\mathbf{X}] = \mathrm{Cov}[\mathbf{X}, \mathbf{X}] = [\mathrm{Cov}(X_i, X_j)] = [\mathrm{E}((X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j))]$$

- Alternatively, we can write $\mathrm{Cov}[\mathbf{X}] = \mathrm{E}[(\mathbf{X} - \mathrm{E}\mathbf{X})(\mathbf{X} - \mathrm{E}\mathbf{X})^T]$.
- We say random variables X_i and X_j are *statistically independent* if the joint density is decomposable, i.e. $f_{ij}(x_i, x_j) = f_i(x_i) f_j(x_j)$.
- Consequently, we have $\mathrm{Cov}(X_i, X_j) = 0$ if X_i and X_j are independent.
- How about the converse?
- Generally, we have $\mathrm{Cov}[\sum_{\ell=1}^k a_\ell \mathbf{X}_\ell, \sum_{\ell'=1}^k b_{\ell'} \mathbf{Y}_{\ell'}] = \sum_{\ell=1}^k \sum_{\ell'=1}^k a_\ell b_{\ell'} \mathrm{Cov}[\mathbf{X}_\ell, \mathbf{Y}_{\ell'}]$.

Mean and Covariance

Example 2.13 (Computing the covariance matrix) Find the covariance matrix for the two random variables X_1 and X_2 introduced in Example 2.12 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

x_1 \ x_2	0	1	$p_1(x_1)$
-1	.24	.06	.3
0	.16	.14	.3
1	.40	.00	.4
$p_2(x_2)$	.8	.2	1

Expectation of Random Quadratic Form

Let $X \sim (\mu, \Sigma)$, i.e. $\mathrm{E}[X] = \mu$, $\mathrm{Cov}(X) = \Sigma$. Consider a symmetric matrix Λ and the corresponding random quadratic form $X^T \Lambda X$. What is its expectation $\mathrm{E}[X^T \Lambda X]$?