STT 843: Multivariate Analysis

3. Random Vectors and Matrices (Chapter 2.3-2.7)

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

1. A Square-Root Matrix
2. Random Vectors and Matrices
3. Matrix Inequalities and Maximization

Distance

- Straight-line, or Euclidean distance between $P = (x_1, x_2)$ and $O = (0, 0)$

$$d(O, P) = \sqrt{x_1^2 + x_2^2}$$

- In general, if the point $P = (x_1, x_2, \ldots, x_p)$ and $O = (0, 0, \ldots, 0)$

$$d(O, P) = \sqrt{x_1^2 + x_2^2 + \ldots + x_p^2}$$

- Straight-line or Euclidean distance is unsatisfactory for most statistical purposes. This is because each coordinates contributes equally to the calculation of Eucildean distance.

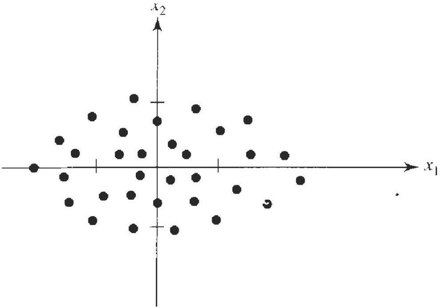

Figure 1.20 A scatter plot with greater variability in the x_1 direction than in the x_2 direction.

Statistical distance

- When the coordinates represent measurements that are subject to random fluctuations of differing magnitudes, it is often to desirable to weight coordinates subject to a great deal of variability less heavily than those that are not highly variable.
- Standardize coordinates.

Suppose we have n pairs of measurements on two variables x_1, x_2 each having mean zero.

$$x_1^* = x_1/\sqrt{s_{11}} \quad \text{and} \quad x_1^* = x_1/\sqrt{s_{22}}.$$

Hence a statistical distance of the point $P = (x_1, x_2)$ from the origin $O = (0, 0)$ can be defined as

$$d(O, P) = \sqrt{(x_1^*)^2 + (x_2^*)^2} = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}}}.$$

- All points which have coordinates (x_1, x_2) and are constant square distance c^2 from origin satisfy

$$\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2$$

and lie on an ellipse.

- The statistical distance from an arbitrary point $P = (x_1, x_2)$ to any fixed point $Q = (y_1, y_2)$.

$$d(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}}}.$$

- The extension of statistical distance to more than two dimensions $P = (x_1, x_2, \dots, x_p)$ and $Q = (y_1, y_2, \dots, y_p)$

$$d(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \cdots + \frac{(x_p - y_p)^2}{s_{pp}}}.$$

- All point P that are a constant squared distance from Q lie on a hyperellipsoid centered at Q whose major and minor axes are parallel to the coordinate axes.

The distances defined above does not include most of the important cases we shall encounter, because of the assumption of independent coordinates. See the following scatter plot

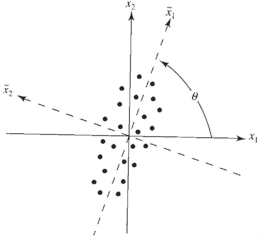

Figure 1.23 A scatter plot for positively correlated measurements and a rotated coordinate system.

- Rotate x_1 and x_2 directions to directions $\tilde{x}_1$ and $\tilde{x}_2$.
- Define the distance from the point $P = (\tilde{x}_1, \tilde{x}_2)$ to the origin $O = (0, 0)$ as

$$d(O, P) = \sqrt{\frac{\tilde{x}_1^2}{\tilde{s}_{11}} + \frac{\tilde{x}_2^2}{\tilde{s}_{22}}}$$

where $\tilde{s}_{11}$ and $\tilde{s}_{22}$ denote the sample variances computed with the $\tilde{x}_1$ and $\tilde{x}_2$ measurements.

- The relation between the original coordinates (x_1, x_2) and the rotated coordinates $(\tilde{x}_1, \tilde{x}_2)$ is provided by

$$\begin{aligned}\tilde{x}_1 &= x_1 \cos(\theta) + x_2 \sin(\theta)\\ \tilde{x}_2 &= -x_1 \sin(\theta) + x_2 \cos(\theta)\end{aligned}$$

- After some straightforward algebraic manipulations, the distance from $P = (\tilde{x}_1, \tilde{x}_2)$ to origin $O = (0, 0)$ can be written in term of the original coordinates x_1 and x_2

$$d(O, P) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}$$

where the a's are numbers such that the distance is nonnegative for all possible variables of x_1 and x_2.

- In general, the statistical distance of the point $P = (x_1, x_2)$ from the fixed point $Q = (y_1, y_2)$ for the situation in which the variables are correlated has the general form

$$d(P,Q) = \sqrt{a_{11}(x_1 - y_1)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + a_{22}(x_2 - y_2)^2}$$

- The generalization of the distance formulas to p dimensions

$$d(P,Q) = \sqrt{\sum_{i=1}^{p} a_{ii}(x_i - y_i)^2 + \sum_{i \neq j}^{p} 2a_{ij}(x_i - y_i)(x_j - y_j)}$$

Any distance measure $d(P, Q)$ between two points P and Q is valid provided that it satisfies the following properties, where R is any other intermediate point:

- $d(P, Q) = d(Q, P)$
- $d(P, Q) > 0$ if $P \neq Q$
- $d(P, Q) = 0$ if $P = Q$
- $d(P, Q) \leq d(P, R) + d(R, Q)$

Example 2.10 (A positive definite matrix quadratic form)

Show that the matrix for the following quadratic form is positive definite:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2.$$

- the "distance" of the point $[x_1, x_2, \ldots, x_p]^T$ to origin

$$\begin{aligned}(\text{ distance })^2 =& a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{pp}^2 \\ &+ 2\left(a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots + a_{p-1,p}x_{p-1}x_p\right)\end{aligned}$$

- the square of the distance $\mathbf{x}$ to an arbitrary fixed point $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p]$.

- A geometric interpretation based on the eigenvalues and eigenvectors of the matrix $\mathbf{A}$.

For example, suppose $p = 2$. Then the points $\mathbf{x}^{\mathsf{T}} = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}^2 + 2a_{12}x_1x_2 = c^2$$

By the spectral decomposition,

$$\mathbf{A} = \lambda_1 \mathbf{e}_1\mathbf{e}_1^{\mathsf{T}} + \lambda_2 \mathbf{e}_2\mathbf{e}_2^{\mathsf{T}}$$

so

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \lambda_1 \left(\mathbf{x}^{\mathsf{T}}\mathbf{e}_1\right)^2 + \lambda_2 \left(\mathbf{x}^{\mathsf{T}}\mathbf{e}_2\right)^2$$

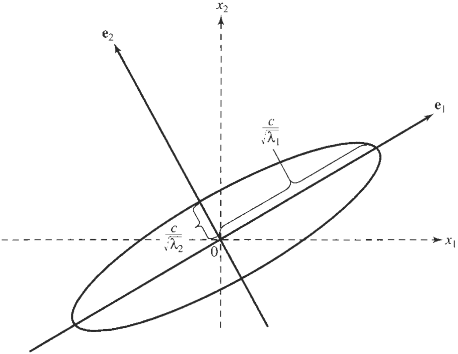

Figure 2.6 Points a constant distance c from the origin ($p = 2, 1 \le \lambda_1 < \lambda_2$).

A Square-Root Matrix

Let $\mathbf{A}$ be a $k\times$ positive definite matrix with spectral decomposition $\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$. Let the normalized eigenvectors be the columns of another matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]$. Then

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} = \mathbf{P}\Lambda\mathbf{P}^{\mathsf{T}}$$

where $\mathbf{P}\mathbf{P}^{\mathsf{T}} = \mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{I}$ and Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \qquad \text{with } \lambda_i > 0$$

Thus

$$\mathbf{A}^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}^{\mathsf{T}} = \sum_{i=1}^{k} \frac{1}{\lambda_i}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}$$

The square-root matrix, of a positive definite matrix $\mathbf{A}$,

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} = \mathbf{P}\Lambda^{1/2}\mathbf{P}^{\mathsf{T}}$$

- symmetric: $\mathbf{A}^{1/2^{\mathsf{T}}} = \mathbf{A}^{1/2}$
- $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$
- $\left(\mathbf{A}^{1/2}\right)^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} = \mathbf{P}\Lambda^{-1/2}\mathbf{P}^{\mathsf{T}}$
- $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$ and $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$, where $\mathbf{A}^{-1/2} = \left(\mathbf{A}^{1/2}\right)^{-1}$.

Random Vectors and Matrices

A random vector is a vector whose elements are random variables. Similarly a random matrix is a matrix whose elements are random variables.

- The expected value of a random matrix

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}(X_{11}) & \mathrm{E}(X_{12}) & \cdots & \mathrm{E}(X_{1p}) \\ \mathrm{E}(X_{21}) & \mathrm{E}(X_{22}) & \cdots & \mathrm{E}(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(X_{n1}) & \mathrm{E}(X_{n2}) & \cdots & \mathrm{E}(X_{np}) \end{bmatrix}$$

- $\mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$
- $\mathrm{E}(\mathbf{A}\boldsymbol{X}\mathbf{B}) = \mathbf{A}\mathrm{E}(\boldsymbol{X})\mathbf{B}$

Example 2.11 (Computing expected values for discrete random variables)

Suppose $p = 2$ and $n = 1$, and consider the random vector $\boldsymbol{X}^{\mathsf{T}} = [X_1, X_2]$. Let the discrete random variable X_1 have the following probability function

X_1	-1	0	1
$p_1(X_1)$	0.3	0.3	0.4

Similarly, let the discrete random varibale X_2 have the probability function

X_2	0	1
$p_2(X_2)$	0.8	0.2

Calculate $\mathrm{E}(\boldsymbol{X})$.

Mean Vectors and Covariance Matrices

Suppose $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ is a $p \times 1$ random vectors. Then each element of $\boldsymbol{X}$ is a random variables with its own marginal probability distribution.

- The marginal mean $\mu_i = \mathrm{E}\left(X_i\right), i = 1, 2, \ldots, p$.
- The marginal variance $\sigma_i^2 = \mathrm{E}\left(X_i - \mu_i\right)^2, i = 1, 2, \ldots, p$.
- The behavior of any pair of random variables, such as X_i and X_k, is described by their joint probability function, and a measure of the linear association between them is provided by the covariance

$$\sigma_{ik} = \mathrm{E}\left(X_i - \mu_i\right)\left(X_k - \mu_k\right)$$

- The means and covariances of $p \times 1$ random vector $\boldsymbol{X}$ can be set out as matrices named population variance-covariance (matrices).

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}), \quad \Sigma = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^{\mathsf{T}}.$$

- Statistical independent X_i and X_k if

$$P\left(X_i \leq x_i \text{ and } X_k \leq x_k\right) = P\left(X_i \leq x_i\right) P\left(X_k \leq x_k\right)$$

or

$$f_{ik}\left(x_i, x_k\right) = f_i\left(x_i\right) f_k\left(x_k\right).$$

- Mutually statistically independent of the p continuous random variables $X_1, X_2, \ldots, X_p$ if

$$f_{1,2,\ldots,p}(x_1, x_2, \ldots, x_p) = f_1(x_1) f_2(x_2) \cdots f_p(x_p)$$

- linear independent of X_i, X_k if

$$\operatorname{Cov}(X_i, X_k) = 0$$

- Population correlation coefficient ρ_{ik}

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

  The correlation coefficient measures the amount of linear association between the random variable X_i and X_k.

- The population correlation matrix ρ

Example 2.12 (Computing the covariance matrix) Find the covariance matrix for the two random variables X_1 and X_2 introduced in Example 2.11 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

$x_1 \backslash x_2$	0	1	$p_1(x_1)$
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.4	0.00	0.4
$p_2(x_2)$	0.8	0.2	1

Example 2.13 (Computing the correlation matrix from the covariance matrix)

Suppose

$$\Sigma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

Obtain the population correlation matrix ρ

Partitioning the Covariance Matrix

Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \cdots \\ X_{q+1} \\ \cdots \\ X_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \cdots \\ \boldsymbol{X}^{(2)} \end{bmatrix} \text{ and then } \boldsymbol{\mu} = \mathrm{E}\boldsymbol{X} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \cdots \\ \mu_{q+1} \\ \cdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \cdots \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

Define

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathsf{T}} \\
=&\mathrm{E}\left[\begin{array}{cc}
\left(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}\right)\left(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}\right)^{\mathsf{T}} & \left(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}\right)\left(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}\right)^{\mathsf{T}} \\
\left(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}\right)\left(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}\right)^{\mathsf{T}} & \left(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}\right)\left(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}\right)
\end{array}\right] \\
=&\left[\begin{array}{cc}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{array}\right]
\end{aligned}
$$

- It is sometimes convenient to use $\text{Cov}\left(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}\right)$ note where

$$\text{Cov}\left(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}\right) = \Sigma_{12} = \Sigma_{21}^{T}$$

is a matrix containing all of the covariance between a component of $\boldsymbol{X}^{(1)}$ and a component of $\boldsymbol{X}^{(2)}$.

The Mean Vector and Covariance Matrix for Linear Combinations of Random Variables

- The linear combination $\mathbf{c}^{\top}\boldsymbol{X} = c_1X_1 + \cdots + c_pX_p$ has

$$\begin{aligned} \text{mean} &= \mathrm{E}\left(\mathbf{c}^{\top}\boldsymbol{X}\right) = \mathbf{c}^{\top}\boldsymbol{\mu} \\ \text{variance} &= \mathrm{Var}\left(\mathbf{c}^{\top}\boldsymbol{X}\right) = \mathbf{c}^{\top}\Sigma\mathbf{c} \end{aligned}$$

where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\Sigma = \mathrm{Cov}(\boldsymbol{X})$.

- Let $\mathbf{C}$ be a matrix, then the linear combinations of $\mathbf{Z} = \mathbf{C}\boldsymbol{X}$ have

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{Z}} &= \mathrm{E}(\mathbf{Z}) = \mathrm{E}(\mathbf{C}\boldsymbol{X}) = \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}} \\ \Sigma_{\mathbf{Z}} &= \mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{C}\mathbf{X}) = \mathbf{C}\Sigma_{\mathbf{X}}\mathbf{C}^{\top} \end{aligned}$$

- Sample Mean

$$\bar{\mathbf{x}}^T = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p]$$

- Sample Covariance Matrix

$$S_n = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}\left(x_{j1} - \bar{x}_1\right)^2 & \cdots & \frac{1}{n}\sum_{j=1}^{n}\left(x_{j1} - \bar{x}_1\right)\left(x_{jp} - \bar{x}_p\right) \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{j=1}^{n}\left(x_{j1} - \bar{x}_1\right)\left(x_{jp} - \bar{x}_p\right) & \cdots & \frac{1}{n}\sum_{j=1}^{n}\left(x_{jp} - \bar{x}_p\right)^2 \end{bmatrix}$$

Matrix Inequalities and Maximization

- Cauchy-Schwarz Inequality

  Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors. Then

$$\left(\mathbf{b}^\top \mathbf{d}\right)^2 \leq \left(\mathbf{b}^\top \mathbf{b}\right)\left(\mathbf{d}^\top \mathbf{d}\right)$$

  with equality if and only if $\mathbf{b} = c\mathbf{d}$ or $\mathbf{d} = c\mathbf{b}$ for some constant c.

- Extended Cauchy-Schwarz Inequality

  Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors, and $\mathbf{B}$ be a positive definite matrix. Then

$$\left(\mathbf{b}^\top \mathbf{d}\right)^2 \leq \left(\mathbf{b}^\top \mathbf{B}\mathbf{b}\right)\left(\mathbf{d}^\top \mathbf{B}^{-1}\mathbf{d}\right)$$

  with equality if and only if $\mathbf{b} = c\mathbf{B}^{-1}\mathbf{d}$ or $\mathbf{d} = c\mathbf{B}\mathbf{b}$ for some constant c.

- **Maximization Lemma**

  Let $\mathbf{B}_{p\times p}$ be positive definite and $\mathbf{d}_{p\times 1}$ be a given vector. Then, for arbitrary nonzero vector $\mathbf{x}$,

$$\max_{\mathbf{x}\neq 0} \frac{\left(\mathbf{x}^\mathsf{T}\mathbf{d}\right)^2}{\mathbf{x}^\mathsf{T}\mathbf{B}\mathbf{x}} = \mathbf{d}^\mathsf{T}\mathbf{B}^{-1}\mathbf{d}$$

  with the maximum attained when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$ for any constant $c \neq 0$.

- Maximization of Quadratic Forms for Points on the Unit Sphere

Let $\mathbf{B}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\max_{x \neq 0} \frac{\mathbf{x}^T \mathbf{B} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_1 \quad (\text{ attained when } \mathbf{x} = \mathbf{e}_1)$$

$$\min_{x \neq 0} \frac{\mathbf{x}^T \mathbf{B} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_p \quad (\text{ attained when } \mathbf{x} = \mathbf{e}_p)$$

Moreover,

$$\max_{x \perp \mathbf{e}_1, \ldots \mathbf{e}_k} \frac{\mathbf{x}^T \mathbf{B} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_{k+1} \quad (\text{ attained when } \mathbf{x} = \mathbf{e}_{k+1}, k = 1, 2, \ldots, p$$

where the symbol $\perp$ is read "perpendicular to."