Math3806 Lecture Note 2 Appendix

Heng Peng

- Example 2.10:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2 = [x_1 \;\; x_2]\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

But the root of

$$\begin{vmatrix} 3-\lambda & -\sqrt{2} \\ -\sqrt{2} & 2-\lambda \end{vmatrix}$$

or

$$(3-\lambda)(2-\lambda) - 2 = \lambda^2 - 5\lambda + 4 = (\lambda - 4)(\lambda - 1)$$

is 4 and 1. Hence

$$\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

is positive definite matrix, and so is

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

P26. For example, if $\mathbf{X}$ is an $n \times 1$ random vector, and the random elements of $\mathbf{X}$ independent with same distribution and same variance σ^2. Then

$$
\begin{aligned}
& \mathrm{E}\{\mathbf{A}(\mathbf{X}-\mathrm{E}\mathbf{X})(\mathbf{X}-\mathrm{E}\mathbf{X})^T\mathbf{B}\} \\
= \;& \mathbf{A}\mathrm{E}\{(\mathbf{X}-\mathrm{E}\mathbf{X})(\mathbf{X}-\mathrm{E}\mathbf{X})^T\}\mathbf{B} \\
= \;& \mathbf{A}\sigma^2 I_{n\times n}\mathbf{B} = \sigma^2\mathbf{AB}
\end{aligned}
$$

P27. For this example, $\mathbf{X}$ is a 2×1 random vector

$$
\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \text{and then} \quad \mathrm{E}\mathbf{X} = \begin{bmatrix} \mathrm{E}X_1 \\ \mathrm{E}X_2 \end{bmatrix}
$$

$$
\mathrm{E}X_1 = -1\cdot 0.3 + 0\cdot 0.3 + 1\cdot 0.4 = 0.1, \quad \mathrm{E}X_2 = 0\cdot 0.8 + 1\cdot 0.2 = 0.2
$$

So $\mathrm{E}\mathbf{X} = [0.1, 0.2]^T$.

P28.

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}X_1 \\ \mathrm{E}X_2 \\ \vdots \\ \mathrm{E}X_p \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathrm{E}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_p-\mu_p \end{pmatrix}(X_1-\mu_1, X_2-\mu_2, \ldots, X_p-\mu_p) \\ &= \begin{pmatrix} \mathrm{E}(X_1-\mu_1)^2 & \cdots & \mathrm{E}(X_1-\mu_1)(X_p-\mu_p) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(X_p-\mu_p)(X_1-\mu_1) & \cdots & \mathrm{E}(X_p-\mu_p)^2 \end{pmatrix} \end{aligned}$$

P28. Hence

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma 1p \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} \quad \text{and} \quad \sigma_{ij} = \sigma_{ji}, \sigma_{ii} = \sigma_i^2, i, j = 1, \dots, p.$$

- Furthermore, Σ is nonnegative definite matrix since

$$\begin{aligned} Y^T \Sigma Y &= \mathrm{E}\left\{ Y^T \begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{pmatrix} (X_1 - \mu_1, \dots, X_p - \mu_p) Y \right\} \\ &= \mathrm{E}[y_1(X_1 - \mu_1) + \cdots + y_p(X_p - \mu_p)]^2 \geq 0 \end{aligned}$$

where $Y = (y_1, y_2, \dots, y_p)^T$.

P29. It is obvious that if X_i, X_j are statistical independent, then

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \mathrm{E}(X_i - \mu_i)(X_j - \mu_j) \\ &= \int (X_i - \mu_i)(X_j - \mu_j) f(X_i, X_j) dX_i dX_j \\ &= \int (X_i - \mu_i)(X_j - \mu_j) f(X_i) f(X_j) dX_i dX_j \\ &= \mathrm{E}(X_i - \mu_i)\mathrm{E}(X_j - \mu_j) = 0 \end{aligned}$$

So statistical independence is stronger than linear dependence.

P30.

$$\rho = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \cdots & \rho_{pp} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \cdots & 1 \end{bmatrix}$$

Furthermore,

$$\rho = V^{-1/2} \Sigma V^{-1/2}, \quad , \Sigma = V^{1/2} \rho V^{1/2}$$

where

$$V = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

P31. Example 2.12. We have known $\mu_1 = \mathrm{E}X_1 = .1$ and $\mu_2 = \mathrm{E}X_2 = .2$. Then

$$\sigma_{11} = \mathrm{E}(X_1 - \mu_1)^2 = \sum_{\text{all} x_1} (X_1 - .1)^2 \cdot p(X_1)$$
$$= (-1 - .1)^2(.3) + (0 - .1)^2(.3) + (1 - .1)^2(.4) = .69$$

$$\sigma_{22} = \mathrm{E}(X_1 - \mu_1)^2 = \sum_{\text{all} x_2} (X_2 - .2)^2 \cdot p(X_2)$$
$$= (0 - .2)^2(.8) + (1 - .2)^2(.2) = .16$$

$$\sigma_{12} = \mathrm{E}(X_1 - \mu_1)(X_2 - \mu_2) = \sum_{\text{all pair} x_1, x_2} (X_1 - .1)(X_2 - .2) \cdot p($$
$$= (-1 - .1)(0 - 0.2)(.24) + (-1 - .1)(1 - 0.2)(.06)$$
$$+ \cdots + (1 - .1)(1 - 0.2)(.00) = -.08 = \sigma_{21}$$

Hence

$$\mathrm{E} = \begin{pmatrix} \mathrm{E}X_1 \\ \mathrm{E}X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} .1 \\ .2 \end{pmatrix}$$

$$\begin{aligned} \Sigma &= \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T \\ &= \mathrm{E}\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .69 & -.08 \\ -.08 & .16 \end{bmatrix} \end{aligned}$$

P31. Example 2.13

$$V^{1/2} = \text{diag}(\sqrt{\sigma}_{11}, \sqrt{\sigma}_{22}, \sqrt{\sigma}_{33}) = \text{diag}(2, 3, 5)$$

$$V^{-1/2} = \text{diag}(\frac{1}{2}, \frac{1}{3}, \frac{1}{5})$$

$$\begin{aligned} \rho &= V^{-1/2}\Sigma V^{-1/2} \\ &= \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \frac{1}{6} & \frac{1}{5} \\ \frac{1}{6} & 1 & -\frac{1}{5} \\ \frac{1}{5} & -\frac{1}{5} & 1 \end{bmatrix} \end{aligned}$$

P34. For example, $X = (X_1, X_2)^T$, we wish to know if $\mu_1 = \mu_2$, then we will consider $X_1 - X_2$ where $c_1 = 1, c_2 = -1$.

$$
\begin{aligned}
\mathrm{E}\mathbf{c}^T\boldsymbol{X} &= \mathrm{E}(c_1X_1 + \cdots + c_pX_p) = c_1\mathrm{E}X_1 + \cdots + c_p\mathrm{E}X_p \\
&= (c_1, \ldots, c_p)\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \mathbf{c}^T\boldsymbol{\mu}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(\mathbf{c}^T\boldsymbol{X}) &= \mathrm{E}(\mathbf{c}^T\boldsymbol{X} - \mathrm{E}\mathbf{c}^T\boldsymbol{X})^2 = \mathrm{E}(\mathbf{c}^T\boldsymbol{X} - \mathbf{c}^T\boldsymbol{\mu})^2 \\
&= \mathrm{E}\mathbf{c}^T(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\mathbf{c} \\
&= \mathbf{c}^T\mathrm{E}\{(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\}\mathbf{c} \\
&= \mathbf{c}^T\Sigma\mathbf{c}
\end{aligned}
$$

$$\mathrm{E}\mathbf{Z} = \mathrm{E}\mathbf{C}\boldsymbol{X} = \mathbf{C}\mathrm{E}\boldsymbol{X} = \mathbf{C}\mu_X$$

$$\begin{aligned}
\Sigma_Z &= \mathrm{Cov}(Z) = \mathrm{E}(\mathbf{Z} - \mathrm{E}\mathbf{Z})(\mathbf{Z} - \mathrm{E}\mathbf{Z})^T \\
&= \mathrm{Cov}(Z) = \mathrm{E}(\mathbf{C}\boldsymbol{X} - \mathrm{E}\mathbf{C}\boldsymbol{X})(\mathbf{C}\boldsymbol{X} - \mathrm{E}\mathbf{C}\boldsymbol{X})^T \\
&= \mathbf{C}\mathrm{E}(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})^T\mathbf{C}^T = \mathbf{C}\Sigma_X\mathbf{C}^T
\end{aligned}$$

- Example: Let $\boldsymbol{X} = (X_1, X_2)^T$ be a random vector with men vector $\boldsymbol{\mu}_X = (\mu_1, \mu_2)^T$ and variance-covariance matrix

$$\Sigma_x = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Find mean and variance of
$\mathbf{Z} = (Z_1, Z_2)^T = (X_1 - X_2, X_1 + X_2)^T$.

- Solution:

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

$$\boldsymbol{\mu}_Z = (\mu_1 - \mu_2, \mu_1 + \mu_2)^T$$

$$\Sigma_Z = \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix}$$

- If $\sigma_{11} = \sigma_{22}$, then $X_1 - X_2$ and $X_1 + X_2$ are linear independent.

P35. Partition the sample mean vector and covariance matrix

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_q \\ \cdots \\ \bar{X}_{q+1} \\ \vdots \\ \bar{X}_p \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{X}}_1 \\ \bar{\boldsymbol{X}}_2 \end{bmatrix}$$

- $$S_n = \left[\begin{array}{ccc|ccc} s_{11} & \cdots & s_{1q} & s_{1,q+1} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{q1} & \cdots & s_{qq} & s_{q,q+1} & \cdots & s_{qp} \\ \hline s_{q+1,1} & \cdots & s_{q+1,q} & s_{q+1,q+1} & \cdots & s_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pq} & s_{p,q+1} & \cdots & s_{pp} \end{array}\right] = \left[\begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22} \end{array}\right]$$

- $$S_{12} = S_{21}^T$$

P36. Cauchy-Schwarz Inequality is a famous inequality,

$$(b_1^2+b_2^2+\cdots+b_p^2)(d_1^2+d_2^2+\cdots+d_p^2) \geq (b_1d_1+b_2d_2+\cdots+b_pd_p)^2$$

Let $\mathbf{b} = (b_1, b_2, \ldots, b_p)^T, \mathbf{d} = (d_1, d_2, \ldots, d_p)^T$, then

$$(\mathbf{b}^T\mathbf{b})(\mathbf{d}^T\mathbf{d}) \geq (\mathbf{b}^T\mathbf{d})^2$$

If $\mathbf{b} - x\mathbf{d} \neq 0$,

$$0 < (\mathbf{b} - x\mathbf{d})^T(\mathbf{b} - x\mathbf{d}) = \mathbf{b}^T\mathbf{b} - 2x(\mathbf{b}^T\mathbf{d}) + x^2(\mathbf{d}^T\mathbf{d})$$

So

$$\begin{aligned} 0 &< \mathbf{b}^T\mathbf{b} - \frac{(\mathbf{b}^T\mathbf{d})^2}{\mathbf{d}^T\mathbf{d}} + \frac{(\mathbf{b}^T\mathbf{d})^2}{\mathbf{d}^T\mathbf{d}} - 2x(\mathbf{b}^T\mathbf{d}) + x^2(\mathbf{d}^T\mathbf{d}) \\ &= \mathbf{b}^T\mathbf{b} - \frac{(\mathbf{b}^T\mathbf{d})^2}{\mathbf{d}^T\mathbf{d}} + (\mathbf{d}^T\mathbf{d})(x - \frac{(\mathbf{b}^T\mathbf{d})}{\mathbf{d}^T\mathbf{d}})^2 \end{aligned}$$

- Let $x = \frac{(\mathbf{b}^T\mathbf{d})}{\mathbf{d}^T\mathbf{d}}$, we have

$$\mathbf{b}^T\mathbf{b} - \frac{(\mathbf{b}^T\mathbf{d})^2}{\mathbf{d}^T\mathbf{d}} > 0.$$

Hence

$$(\mathbf{b}^T\mathbf{b})(\mathbf{d}^T\mathbf{d}) \geq (\mathbf{b}^T\mathbf{d})^2$$

If $\mathbf{b} - x\mathbf{d} = 0$, $\mathbf{b}$ and $\mathbf{d}$ are proportional or with same direction, then

$$(\mathbf{b}^T\mathbf{b})(\mathbf{d}^T\mathbf{d}) = (\mathbf{b}^T\mathbf{d})^2$$

P37.

$$
\begin{aligned}
(\mathbf{b}^T\mathbf{d})^2 &= (\mathbf{b}^T\mathbf{B}^{\frac{1}{2}}\mathbf{B}^{-\frac{1}{2}}\mathbf{d})^2 \le (\mathbf{b}^T\mathbf{B}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}}\mathbf{b})(\mathbf{d}^T\mathbf{B}^{-\frac{1}{2}}\mathbf{B}^{-\frac{1}{2}}\mathbf{d}) \\
&= (\mathbf{b}^T\mathbf{B}\mathbf{b})(\mathbf{d}^T\mathbf{B}^{-1}\mathbf{d})
\end{aligned}
$$

or

$$
(\mathbf{x}^T\mathbf{d})^2 \le (\mathbf{x}^T\mathbf{B}\mathbf{x})(\mathbf{d}^T\mathbf{B}^{-1}\mathbf{d})
$$

If $\mathbf{B}^{\frac{1}{2}}x = c\mathbf{B}^{-\frac{1}{2}}d$, then then $(\mathbf{x}^T\mathbf{d})^2 = (\mathbf{x}^T\mathbf{B}\mathbf{x})(\mathbf{d}^T\mathbf{B}^{-1}\mathbf{d})$.

- Notice
$$\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$$
, Define $\mathbf{x}^* = \mathbf{x}/(\mathbf{x}^T\mathbf{x})^{\frac{1}{2}}$, then $\|\mathbf{x}^*\| = 1$, and the optimal problem changes to
$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T\mathbf{B}\mathbf{x}$$
- By spectral decomposition of $\mathbf{B}$ that
$$\mathbf{B} = \lambda_1\mathbf{e}_1\mathbf{e}_1^T + \cdots + \lambda_p\mathbf{e}_p\mathbf{e}_p^T = P\Lambda P'$$
Then
$$\begin{aligned}
\mathbf{x}^T\mathbf{B}\mathbf{x} &= \mathbf{x}^T(\lambda_1\mathbf{e}_1\mathbf{e}_1^T + \cdots + \lambda_p\mathbf{e}_p\mathbf{e}_p^T)\mathbf{x} \\
&= \lambda_1(\mathbf{x}^T\mathbf{e}_1)(\mathbf{e}_1^T\mathbf{x}) + \cdots + \lambda_p(\mathbf{x}^T\mathbf{e}_p)(\mathbf{e}_p^T\mathbf{x}) \\
&\leq \lambda_1\{(\mathbf{x}^T\mathbf{e}_1)^2 + \cdots + (\mathbf{x}^T\mathbf{e}_p)^2\} \\
&= \lambda_1\mathbf{x}^T PP'\mathbf{x} = \lambda_1\|\mathbf{x}\|^2 = \lambda_1
\end{aligned}$$

- Similar way

$$\mathbf{x}^T\mathbf{B}\mathbf{x} \geq \lambda_p\{(\mathbf{x}^T\mathbf{e}_1)^2+\cdots+(\mathbf{x}^T\mathbf{e}_p)^2\} = \lambda_p\mathbf{x}^T PP'\mathbf{x} = \lambda_p\|\mathbf{x}\|^2 = \lambda_p$$

- Furthermore, if $\mathbf{x}$ is perpendicular to $\mathbf{e}_1, \ldots, \mathbf{e}_1$, i.e. $\mathbf{x}^T\mathbf{e}_i = 0, i = 1, \ldots, k,$

$$\begin{aligned}
\mathbf{x}^T\mathbf{B}\mathbf{x} &= \sum_{i=1}^{k}\lambda_i(\mathbf{x}^T\mathbf{e}_i)^2 + \sum_{i=k+1}^{p}\lambda_i(\mathbf{x}^T\mathbf{e}_i)^2 \\
&\leq \lambda_i\{(\mathbf{x}^T\mathbf{e}_{k+1})^2 + \cdots + (\mathbf{x}^T\mathbf{e}_p)^2\} \\
&\leq \lambda_{k+1}\{(\mathbf{x}^T\mathbf{e}_1)^2 + \cdots + (\mathbf{x}^T\mathbf{e}_p)^2\} \\
&= \lambda_{k+1}\mathbf{x}^T PP'\mathbf{x} = \lambda_{k+1}
\end{aligned}$$