

# Lecture 7

## Inference for Covariance Matrix

Readings: Johnson & Wichern 2007, Chapter 4.4; Rencher 2002, Chapter 7.1, 7.2

*DSA 8070 Multivariate Analysis*

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

Whitney Huang  
Clemson University

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## 1 Covariance Matrix

## 2 Statistical Inference for the Covariance Matrix

## 3 High-Dimensional Covariance Estimation

- **Understand the sample covariance matrix and its distributional properties:** compute the sample covariance matrix and describe the [Wishart](#) distribution and its key properties
- **Conduct statistical inference for covariance structures:** perform hypothesis testing and estimation for covariance matrices in both classical and modern settings
- **Address challenges in high-dimensional covariance estimation:** recognize issues in high-dimensional data and apply [shrinkage](#) and [sparse](#) estimation techniques to obtain reliable covariance estimates

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

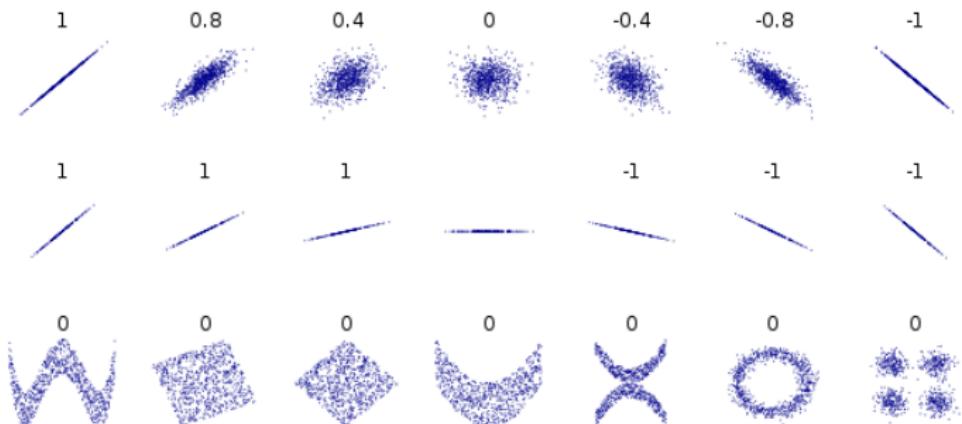
High-Dimensional  
Covariance Estimation

Covariance Matrix

Statistical Inference for  
the Covariance MatrixHigh-Dimensional  
Covariance Estimation

## Review: Interpretation of Covariance

- $\text{Cov}(X, Y) > 0$ : variables tend to increase together
- $\text{Cov}(X, Y) < 0$ : one increases as the other decreases
- $\text{Cov}(X, Y) = 0$ : no linear association



**Figure:** Image courtesy of Wikipedia at [https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)

# Population Covariance Matrix and Its Properties

For a random vector  $\mathbf{X} = (X_1, \dots, X_p)^\top$  with mean vector  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$ , the population covariance matrix is:

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top] \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}\end{aligned}$$

$\sigma_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = \Sigma_{ji}$  for  $i, j = 1, \dots, p \Rightarrow \boldsymbol{\Sigma}$  is a  $p \times p$  symmetric matrix  $\Rightarrow \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\top$

Covariance Matrix

Statistical Inference for  
the Covariance MatrixHigh-Dimensional  
Covariance Estimation

## Properties of Covariance Matrices (Cont'd)

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

- Positive semi-definite: A covariance matrix is positive semi-definite if, for any non-zero vector  $\mathbf{a} = (a_1, \dots, a_p)^T$ ,

$$\mathbf{a}^T \Sigma \mathbf{a} \geq 0$$

⇒ Any linear combination of the variables  $\mathbf{X}$  has non-negative variance

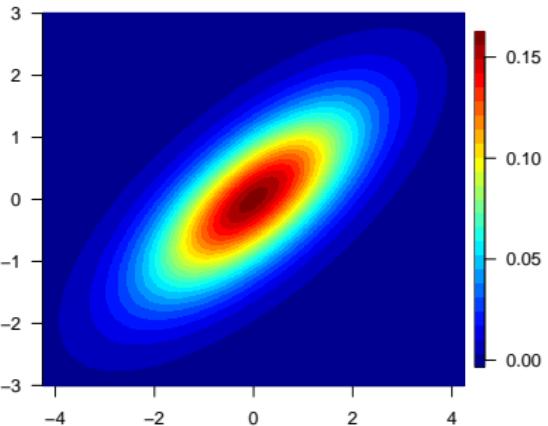
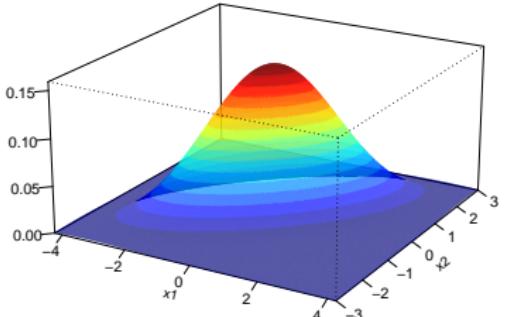
- Eigenvalues of  $\Sigma$  are nonnegative; eigenvectors give the principal directions
- Determinant  $|\Sigma|$  measures the generalized variance

## Geometric Interpretation

- Covariance matrix defines **ellipses of constant Mahalanobis distance**:

$$D^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

- Contours visualize variance and correlation structure



$\Sigma$  typically unknown and estimated using the sample covariance matrix

# Sample Covariance Matrix (SCM)

Given  $n$  observations  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ :

$$\hat{\Sigma}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

- Symmetric and positive semi-definite
- Used in PCA, discriminant analysis, Mahalanobis distance

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## Review: Chi-Square and Variance Estimation

Let  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  be i.i.d. random variables.

- The sample variance is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- Then:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$\Rightarrow \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}$$

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## Review: Chi-Square and Variance Estimation

Let  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  be i.i.d. random variables.

- The sample variance is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- Then:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$\Rightarrow \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}$$

**Motivation:** Just as the *chi-square* distribution governs *sample variance* in the **univariate** case, the *Wishart* distribution governs the *sample covariance matrix* in the **multivariate** case.

Covariance Matrix

Statistical Inference for the Covariance Matrix

High-Dimensional Covariance Estimation

# Wishart Distribution and Its Properties

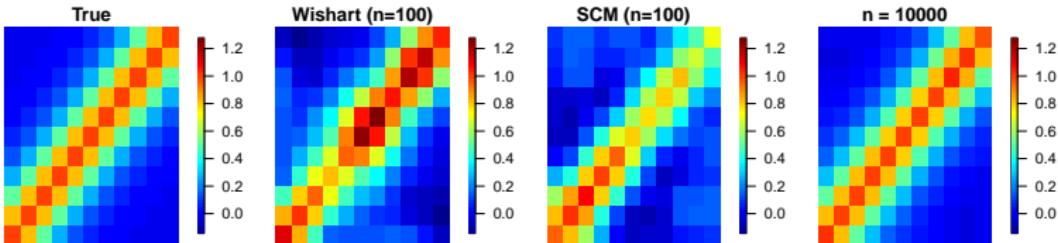
If  $\mathbf{X}_1, \dots, \mathbf{X}_n \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , then:

$$\mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top \sim W_p(n-1, \boldsymbol{\Sigma})$$

- $\mathbb{E}[\mathbf{S}] = (n-1)\boldsymbol{\Sigma}$
- $\text{Var}[\mathbf{S}_{ij}] = (n-1)[\boldsymbol{\Sigma}_{ij}^2 + \boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj}]$

$$\hat{\boldsymbol{\Sigma}}_n = \frac{\mathbf{S}}{n-1} \sim \frac{W_p(n-1, \boldsymbol{\Sigma})}{n-1} \Rightarrow \mathbb{E}[\hat{\boldsymbol{\Sigma}}_n] = \boldsymbol{\Sigma}$$

Fundamental for statistical inference on  $\boldsymbol{\Sigma}$



## Confidence Intervals for Elements of $\Sigma$

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

- Variance element:

$$\frac{(n-1)\hat{\sigma}_{ii}}{\sigma_{ii}} \sim \chi^2_{n-1}$$

⇒ CI for  $\sigma_{ii}$

- For covariance or correlation: asymptotic normality or bootstrap [Efron, 1979]
- Bootstrap: resample rows of data matrix, recompute  $\hat{\Sigma}$

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

- Hypothesis testing:  $H_0 : \Sigma = \Sigma_0$

- Likelihood ratio test statistic:

$$LRT = n \left( \log |\Sigma_0| - \log |\hat{\Sigma}| + \text{tr}(\Sigma_0^{-1} \hat{\Sigma}) - p \right)$$

- Asymptotic distribution:  $\chi^2_{p(p+1)/2}$

- Bootstrap methods for confidence intervals

## Example: Hypothesis Test for $\Sigma$

Test  $H_0 : \Sigma = I$  with  $p = 2, n = 50$ :

$$LRT = n \left( \log |I| - \log |\hat{\Sigma}| + \text{tr}(\hat{\Sigma}) - p \right)$$

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## Example: Hypothesis Test for $\Sigma$

Test  $H_0 : \Sigma = I$  with  $p = 2, n = 50$ :

$$LRT = n \left( \log |I| - \log |\hat{\Sigma}| + \text{tr}(\hat{\Sigma}) - p \right)$$

Suppose  $\hat{\Sigma} = \begin{bmatrix} 1.5 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$

$$LRT \approx 50 \times (0 - \log(1.05) + 2.2 - 2) = 9.5$$

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## Example: Hypothesis Test for $\Sigma$

Test  $H_0 : \Sigma = I$  with  $p = 2$ ,  $n = 50$ :

$$LRT = n \left( \log |I| - \log |\hat{\Sigma}| + \text{tr}(\hat{\Sigma}) - p \right)$$

Suppose  $\hat{\Sigma} = \begin{bmatrix} 1.5 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$

$$LRT \approx 50 \times (0 - \log(1.05) + 2.2 - 2) = 9.5$$

Compare with the  $\chi^2_3$  critical value of  $\approx 7.81$  at  $\alpha = 0.05$ . Since the test statistic (9.5) exceeds this critical value, we reject the null hypothesis that  $\Sigma = I$ .

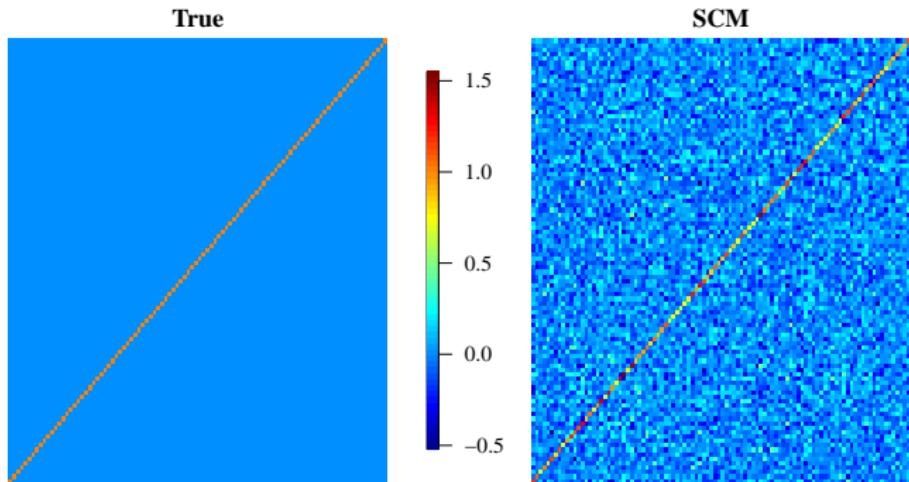
Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## SCM Breakdown in High Dimensions

- When  $p \approx n$  or  $p > n$ ,  $\hat{\Sigma}$  becomes ill-conditioned or singular
- Eigenvalues may be unstable; inverse is unreliable
- Motivates need for regularized estimation



Covariance Matrix

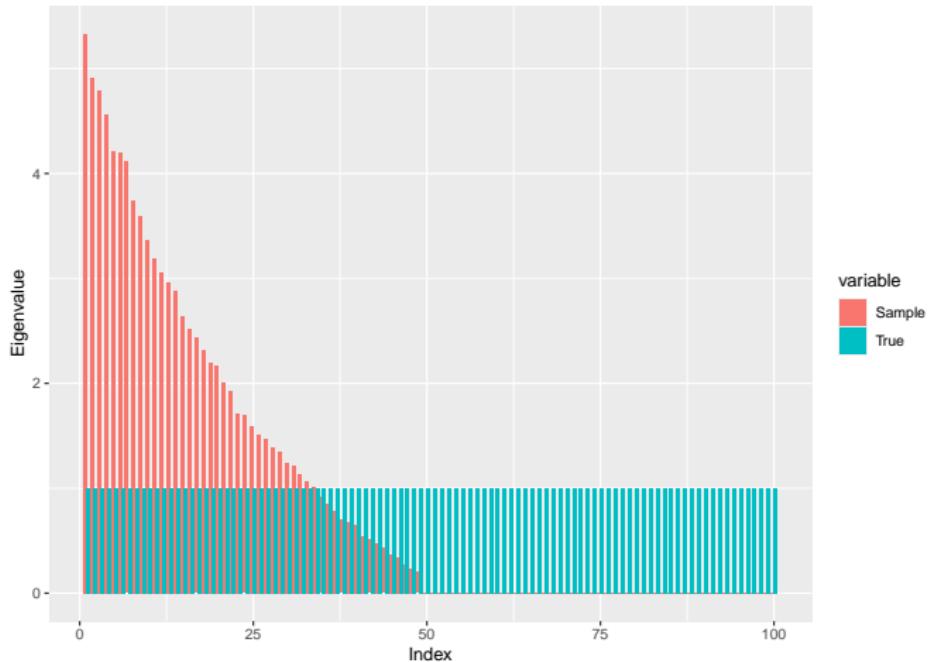
Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

## SCM Breakdown in High Dimensions

For the high-dimensional setting with  $n = 50 < p = 100$  and  $\Sigma = I$ , the estimated eigenvalues deviate substantially from the true values and exhibit degeneracy

Eigenvalues: Sample vs. True



Covariance Matrix

Statistical Inference for  
the Covariance Matrix

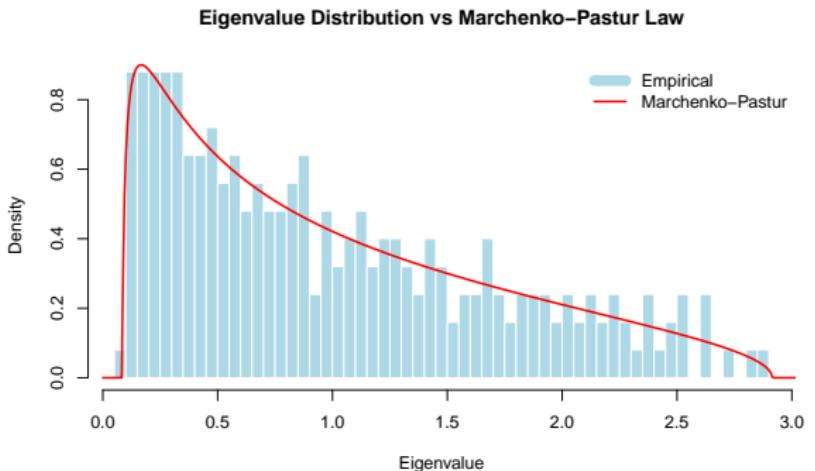
High-Dimensional  
Covariance Estimation

# Eigenvalue Distribution in High Dimensions

- When  $p, n \rightarrow \infty$  with  $p/n \rightarrow c \in (0, 1)$ :

Empirical eigenvalue distribution  $\rightarrow$  Marchenko-Pastur law

- Bulk eigenvalues spread, largest eigenvalue deviates
- Explains instability of SCM in high-d settings



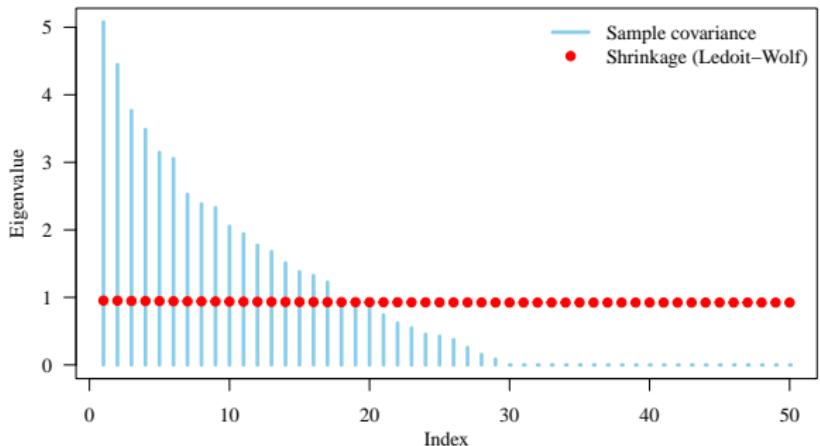
## Shrinkage Estimation (Ledoit-Wolf)

$$\hat{\Sigma}_{\text{shrink}} = (1 - \lambda)\hat{\Sigma} + \lambda T$$

- $T$ : target (e.g., identity, constant correlation)
- $\lambda$ : shrinkage intensity (data-driven)
- Reduces variance, improves conditioning

Simulation:  $p = 50, n = 30$

- Sample eigenvalues: unstable
- Shrinkage pulls toward identity
- Condition number improves



Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation

Covariance Matrix

Statistical Inference for  
the Covariance MatrixHigh-Dimensional  
Covariance Estimation

Estimate inverse covariance (precision) matrix  $\Theta = \Sigma^{-1}$ :

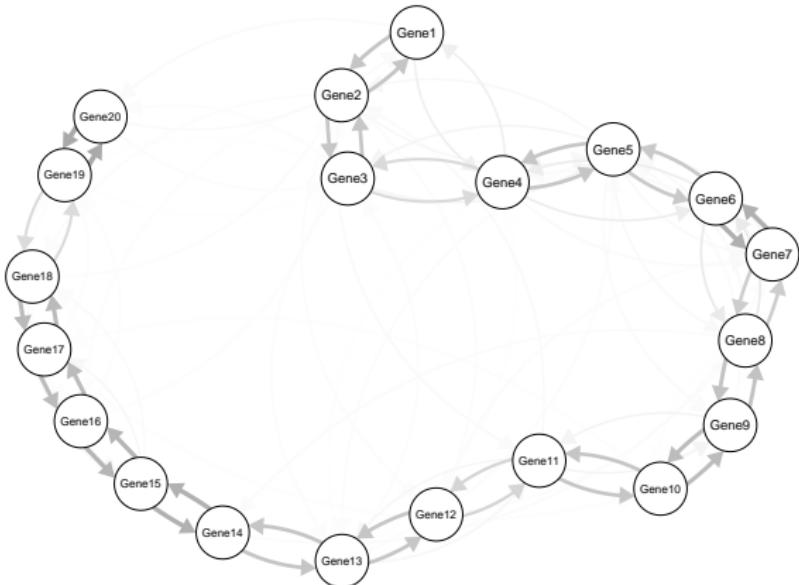
$$\hat{\Theta} = \arg \max_{\Theta > 0} \log \det(\Theta) - \text{tr}(\hat{\Sigma}\Theta) - \rho \|\Theta\|_1$$

- Promotes sparsity in  $\Theta$
- Interpretable as conditional independence graph
- Useful in high-dimensional settings (e.g., genomics, finance)

# Graphical Lasso Application

- Data: AR(1) covariance matrix  $\Rightarrow$  sparse precision matrix
- Graphical Lasso estimates the sparse precision matrix
- Nonzero entries  $\Rightarrow$  edges in the conditional independence graph

Graphical Lasso Network (nonzeros in precision)



## Summary

- The sample covariance matrix (SCM) is the classical estimator, with properties governed by the Wishart distribution
- Inference tools include confidence intervals, hypothesis tests, and bootstrap methods
- In high-dimensional settings, the SCM becomes unstable or singular, motivating regularization
- Shrinkage estimators (e.g., Ledoit-Wolf) improve conditioning; sparse methods (e.g., graphical lasso) enable interpretable structure learning
- Practical implementations are available in R: `cov`, `cov.shrink`, `glasso`, `boot`

Covariance Matrix

Statistical Inference for  
the Covariance Matrix

High-Dimensional  
Covariance Estimation