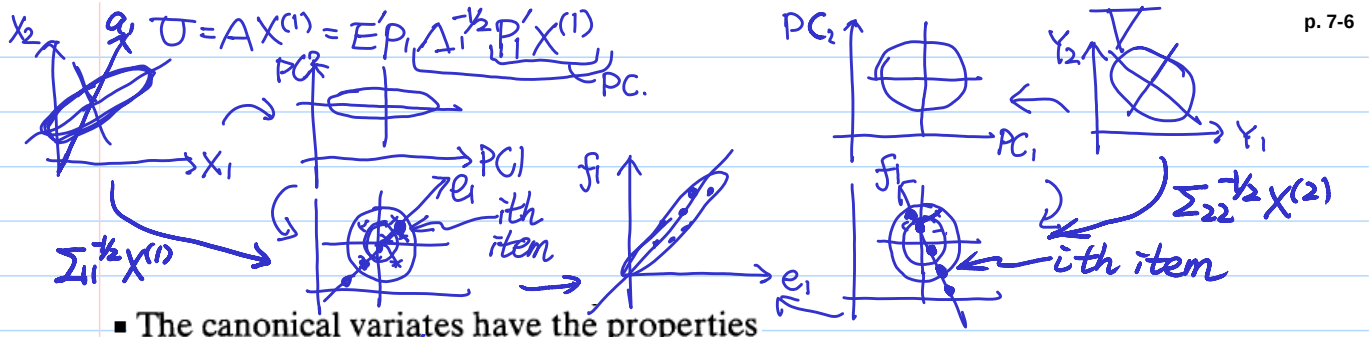

- The canonical variates have the properties

$$\begin{aligned}
\text{Var}(U_k) &= \text{Var}(V_k) = 1 \\
\text{Cov}(U_k, U_\ell) &= \text{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell \\
\text{Cov}(V_k, V_\ell) &= \text{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell \\
\text{Cov}(U_k, V_\ell) &= \text{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell
\end{aligned}$$

for $k, \ell = 1, 2, \ldots, p$.

- $\text{Cov}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) = \mathbf{A}\boldsymbol{\Sigma}_{11}$

$$\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \text{Corr}(\mathbf{U}, \mathbf{X}^{(1)}) = \text{Cov}(\mathbf{U}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}) = \text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)})$$

$$= \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2}$$

Similar calculations for the pairs $(\mathbf{U}, \mathbf{X}^{(2)})$, $(\mathbf{V}, \mathbf{X}^{(2)})$ and $(\mathbf{V}, \mathbf{X}^{(1)})$ yield

$$\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} \qquad \underset{(q\times q)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(2)}}} = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{V}_{22}^{-1/2}$$

$$\underset{(p\times q)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(2)}}} = \mathbf{A}\boldsymbol{\Sigma}_{12}\mathbf{V}_{22}^{-1/2} \qquad \underset{(q\times p)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(1)}}} = \mathbf{B}\boldsymbol{\Sigma}_{21}\mathbf{V}_{11}^{-1/2}$$

- To ease the computation burden, many people prefer to get the canonical correlations from

$$|\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \rho^{*2}\mathbf{I}| = 0$$

The coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ follow directly from the eigenvector equations

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{a} = \rho^{*2}\mathbf{a}$$

$$\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{b} = \rho^{*2}\mathbf{b}$$

➢ Interpreting the population canonical variables

- identify the canonical variables
  - the linear combinations can be interpreted much as in principal components
  - it is helpful to compute the correlation between the canonical variables and the original variables as way to determine the relative important of original variables to a particular canonical variable
- canonical correlation
  - canonical correlation generalizes the correlation between 2 variables

$$|\text{Corr}(X_i^{(1)}, X_k^{(2)})| = |\text{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)})| \le \max_{\mathbf{a},\mathbf{b}} \text{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \rho_1^*$$

  - An R^2 type statistic can be obtained to explain the total variance explained by a given set of canonical variables:

Because of its multiple correlation coefficient interpretation, the kth *squared* canonical correlation ρ_k^{*2} is the proportion of the variance of canonical variate U_k "explained" by the set $\mathbf{X}^{(2)}$. It is also the proportion of the variance of canonical variate V_k "explained" by the set $\mathbf{X}^{(1)}$

- first r canonical variables as a summary of variability
  - remember that the linear combinations are chosen to maximize correlation between the canonical variables
  - these linear combinations may be quite different from those obtained from principal component within a particular set of variables
  - the canonical may not necessarily explain the total variation within a set of original variables

➢ CCA for standardized data

If the original variables are standardized with $\mathbf{Z}^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_p^{(1)}]'$ and $\mathbf{Z}^{(2)} = [Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_q^{(2)}]'$, from first principles, the canonical variates are of the form

$$U_k = \mathbf{a}_k'\mathbf{Z}^{(1)} = \mathbf{e}_k'\boldsymbol{\rho}_{11}^{-1/2}\mathbf{Z}^{(1)}$$

$$V_k = \mathbf{b}_k'\mathbf{Z}^{(2)} = \mathbf{f}_k'\boldsymbol{\rho}_{22}^{-1/2}\mathbf{Z}^{(2)}$$

Here, $\text{Cov}(\mathbf{Z}^{(1)}) = \boldsymbol{\rho}_{11}$, $\text{Cov}(\mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{22}$, $\text{Cov}(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}) = \boldsymbol{\rho}_{12} = \boldsymbol{\rho}_{21}'$, and $\mathbf{e}_k$ and $\mathbf{f}_k$ are the eigenvectors of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ and $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$, respectively. The canonical correlations, ρ_k^*, satisfy

$$\text{Corr}(U_k, V_k) = \rho_k^*, \qquad k = 1, 2, \ldots, p$$

where $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ are the nonzero eigenvalues of the matrix $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ (or, equivalently, the largest eigenvalues of $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$).

- the canonical correlations are unchanged by the standardization (cf. PCA)

❖ **Reading**: Reference, 10.1, 10.2, 10.3

i.e. $\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}$ & $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ would have same eigenvalues.

$\mathbf{e}_k$ ↔ $\mathbf{e}_k^*$

$\mathbf{a}_k = \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_k$ ⟷ $\mathbf{a}_k^* = \boldsymbol{\rho}_{11}^{-1/2}\mathbf{e}_k^*$, $\mathbf{a}_k^* = \mathbf{V}_{11}^{1/2}\mathbf{a}_k$

- Sample Canonical Variables and Canonical Correlations

➢ Data:

$$\mathbf{X} = \left[\mathbf{X}^{(1)} \vdots \mathbf{X}^{(2)}\right] = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2p}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2q}^{(2)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{n1}^{(1)} & x_{n2}^{(1)} & \cdots & x_{np}^{(1)} & x_{n1}^{(2)} & x_{n2}^{(2)} & \cdots & x_{nq}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^{(1)\prime} & \vdots & \mathbf{x}_1^{(2)\prime} \\ \vdots & & \vdots \\ \mathbf{x}_n^{(1)\prime} & \vdots & \mathbf{x}_n^{(2)\prime} \end{bmatrix}$$

(Var(1), Var(1), Var(2), Var(2); an item)

➢ Finding canonical variables and canonical correlations

- replace population distribution by empirical distribution
- replace $\boldsymbol{\Sigma}$ by $\mathbf{S}$ (or $\mathbf{S}_n$)
- replay $\boldsymbol{\rho}$ by $\mathbf{R}$
- then, the rests the same as above

$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$

parameters: $\mathbf{A} \to \hat{\mathbf{A}}$, $\mathbf{B} \to \hat{\mathbf{B}}$, $\rho_k^* \to \hat{\rho}_k^*$; random: $\mathbf{U} \to \hat{\mathbf{U}}$, $\mathbf{V} \to \hat{\mathbf{V}}$

scores of canonical variate

➢ Scatter plots of the first $(\hat{U}_1, \hat{V}_1)$ pair may reveal atypical observations $\mathbf{x}_j$ requiring further study. If the canonical correlations $\hat{\rho}_2^*, \hat{\rho}_3^*, \ldots$ are also moderately large, scatter plots of the pairs $(\hat{U}_2, \hat{V}_2), (\hat{U}_3, \hat{V}_3), \ldots$ may also be helpful in this respect.