

Shrinkage estimation of the mean matrix of multivariate complex normal distributions

Yoshihiko KONNO

Japan Women's University

CIMPA-MICINN-UNESCO RESEARCH SCHOOL
ANALYTICAL AND ALGEBRAIC TOOLS
IN STATISTICS AND GRAPHICAL MODELS
HAMMAMET, TUNISIA

5-17 September 2011

- 1 Preliminaries : Complex normal and Wishart distributions**
 - Complex normal distribution
 - Complex Wishart distribution
- 2 Set-up of a problem**
- 3 Key ingredients**
- 4 Integration-by-parts formulae**
 - Stein identity for complex normal distribution
 - Stein-Haff identity for the complex Wishart distribution
- 5 Class of estimators and unbiased estimate of the risk**
 - Class of estimators
 - Unbiased risk estimate
 - Improved estimators
- 6 Concluding remarks**

- ★ Good reference is Andersen et al. (1995, Lecture note in Statistics vol. 101).
- ★ We say that a random variable \mathbf{Z} has a standard complex normal distributions $\mathbb{C}\mathbf{N}(\mathbf{0}, \mathbf{1})$ if

$$[\mathbf{Z}] = \begin{pmatrix} \operatorname{Re}\mathbf{Z} \\ \operatorname{Im}\mathbf{Z} \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \right].$$

- ★ Then the density (w.r.t. Lebesgue measure on \mathbb{C}) of \mathbf{Z} is given as

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{\sqrt{\pi}} \exp(-\bar{\mathbf{z}}\mathbf{z})$$

where $\bar{\mathbf{z}}$ is the conjugate of a complex \mathbf{x} ($\mathbf{x} \in \mathbb{C}$).

- ★ Assume that $\mathbf{Z} \sim \mathbb{C}\mathbf{N}(\mathbf{0}, \mathbf{1})$. For $\xi \in \mathbb{C}$, $\sigma \in \mathbb{R}_+$,

$$\mathbf{X} := \xi + \sigma\mathbf{Z} \sim \mathbb{C}\mathbf{N}(\xi, \sigma^2).$$

- ★ Let \mathbf{X} be \mathbb{C}^p -valued random vector. For $\forall \mathbf{c} \in \mathbb{C}^p$, $\xi \in \mathbb{C}^p$, and $\Sigma \in \mathbb{C}_+^{p \times p}$

$$\mathbf{c}^* \mathbf{X} \sim \mathbb{C}\mathbf{N}(\mathbf{c}^* \xi, \mathbf{c}^* \Sigma \mathbf{c}) \iff \mathbf{X} \sim \mathbb{C}\mathbf{N}_p(\xi, \Sigma),$$

where \mathbf{c}^* is the transpose complex conjugate of a column vector \mathbf{c} and $\mathbb{C}_+^{p \times p}$ is the set of all p.d. complex matrices.

- ★ Let $\mathbf{X} \sim \mathbb{C}\mathbf{N}_p(\xi, \Sigma)$. Then the density of \mathbf{X} (w.r.t. Lebesgue measure on \mathbb{C}^p) is given as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\pi^p} (\text{Det } \Sigma)^{-1} \exp\{-(\mathbf{x} - \xi)^* \Sigma^{-1} (\mathbf{x} - \xi)\}.$$

- ★ Let $\mathbf{X} \sim \mathbb{C}N_p(\boldsymbol{\xi}, \boldsymbol{\Sigma})$. Then the density of $[\mathbf{X}]$ (w.r.t Lebesgue measure on \mathbb{R}^{2p}) is given as

$$[\mathbf{X}] := \begin{pmatrix} \mathbf{Re} \mathbf{X} \\ \mathbf{Im} \mathbf{X} \end{pmatrix} \sim N_{2p} \left[\begin{pmatrix} \mathbf{Re} \boldsymbol{\xi} \\ \mathbf{Im} \boldsymbol{\xi} \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \mathbf{Re} \boldsymbol{\Sigma} & -\mathbf{Im} \boldsymbol{\Sigma} \\ \mathbf{Im} \boldsymbol{\Sigma} & \mathbf{Re} \boldsymbol{\Sigma} \end{pmatrix} \right],$$

where $\mathbf{Re} \boldsymbol{\Sigma}$ and $\mathbf{Im} \boldsymbol{\Sigma}$ are symmetric and skew-symmetric.

- ★ Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be mutually independent $\mathbb{C}\mathbf{N}_p(\mathbf{0}, \Sigma)$ random vectors. Then the distribution of

$$\mathbf{S} := \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^*$$

is called a complex Wishart distribution with parameters n, Σ (this is denoted by $\mathbb{C}\mathbf{W}_p(n, \Sigma)$)

- ★ If $n \geq p$, then $\mathbb{P}(\mathbf{S} \text{ is p.d.}) = 1$ and the density of \mathbf{S} (w.r.t. Lebesgue measure on $\mathbb{C}_+^{p \times p}$) is given as

$$f_{\mathbf{S}}(\mathbf{w}) = \frac{\text{Det}(\mathbf{w})^{n-p} \exp(-\text{Tr}(\mathbf{w}\Sigma^{-1}))}{\text{Det}(\Sigma)^n \pi^{p(p-1)/2} \prod_{j=1}^p \Gamma(n+1-j)},$$

where $\mathbf{w} \in \mathbb{C}_+^{p \times p}$, the set of all p.d. complex matrices, and $\Gamma(\cdot)$ is Euler's gamma function.

Observation for the problem of estimating mean matrix Ξ

- ★ Two random matrices $\mathbf{X} : m \times p$ and $\mathbf{S} : p \times p$ are independent and distributed as follows.
- ★ $\mathbf{X} : m \times p \sim \mathbb{C}N_{m \times p}(\Xi, \mathbf{I}_m \otimes \Sigma)$ and $\mathbf{S} : p \times p \sim \mathbb{C}W_p(n, \Sigma)$, where Ξ is unknown $m \times p$ matrix with complex entries and Σ is unknown Hermitian p.d. $p \times p$ matrix.
Write $\mathbf{X}' := (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$ and $\Xi' = (\xi_1, \xi_2, \dots, \xi_m)$, the transpose of the matrix Ξ . Then \mathbf{X}_i ($i = 1, 2, \dots, m$) are independently distributed as $\mathbb{C}N_p(\xi'_i, \Sigma)$.

- ★ Assume that $n > p$.

Loss function

- ★ We consider the problem of estimating unknown mean matrix Ξ based on observation \mathbf{X} and \mathbf{S} under a loss function:

$$L((\Xi, \Sigma), \hat{\Xi},) = \text{Tr} \{ \Sigma^{-1} (\hat{\Xi} - \Xi)^* (\hat{\Xi} - \Xi) \},$$

where $\hat{\Xi}$ is an estimator for Ξ .

- ★ Risk function :

$$R((\Xi, \Sigma), \hat{\Xi}) = \mathbb{E}[L((\Xi, \Sigma), \hat{\Xi})],$$

Goal

Find estimators $\hat{\Xi}$ s.t.

$$R((\Xi, \Sigma), \hat{\Xi}) \leq R((\Xi, \Sigma), \mathbf{X}) = mp \quad \text{for } \forall (\Xi, \Sigma).$$

Key ingredients

- 1 We consider a class of estimators $\hat{\Xi}$ whose modification of the estimator \mathbf{X} depends on the eigenvalues of $\mathbf{X}^* \mathbf{X} \mathbf{S}^{-1}$.
- 2 Then $R((\Xi, \Sigma), \hat{\Xi})$, the risk of estimators $\hat{\Xi}$, depends on unknown parameters (Ξ, Σ) .
- 3 To compare $R((\Xi, \Sigma), \hat{\Xi})$ with $R((\Xi, \Sigma), \mathbf{X}) = mp$, we derive the unbiased estimate of the risk $R((\Xi, \Sigma), \hat{\Xi})$.

$$R((\Xi, \Sigma), \hat{\Xi}) = mp + \mathbb{E}[\hat{\Delta}]$$

Here $\hat{\Delta}$ is a function of the eigenvalues of $\mathbf{X}^* \mathbf{X} \mathbf{S}^{-1}$.

- 4 If $\hat{\Delta} \leq 0$, then $\hat{\Xi}$ improves upon \mathbf{X} , i.e., $R((\Xi, \Sigma), \hat{\Xi}) \leq mp$ for $\forall (\Xi, \Sigma)$.
- 5 To derive $\hat{\Delta}$, we use integration-by-parts formulae (Stein identity for complex normal distribution and Stein-Haff identity for complex Wishart distribution).

Notation 1

- ★ **Rm \mathbf{C}** and **Im \mathbf{c}** : the real and imaginary parts of a complex number \mathbf{c} .
- ★ $\mathbb{R}_{\geq}^p = \{(\ell_1, \ell_2, \dots, \ell_p) \in \mathbb{R}^p : \ell_1 \geq \ell_2 \geq \dots \geq \ell_p > 0\}$
- ★ $\mathbb{C}^{m \times p}$: the sets of all $m \times p$ matrices of complex entries.
- ★ \mathbf{C}' and $\overline{\mathbf{C}}$: the transpose and the conjugate of $\mathbf{C} \in \mathbb{C}^{m \times p}$.
 $\mathbf{C}^* = \overline{\mathbf{C}'}$.

Notation 2

- ★ For $\mathbf{z} = \mathbf{x} + \sqrt{-1} \mathbf{y}$ ($\mathbf{x}, \mathbf{y} \in \mathbb{R}$) and differentiable function $\mathbf{g}(\mathbf{z}) = \mathbf{u}(\mathbf{z}) + \sqrt{-1} \mathbf{v}(\mathbf{z})$ ($\mathbf{u}(\mathbf{z})$ and $\mathbf{v}(\mathbf{z})$ are real-valued differential functions),

$$\begin{aligned} \frac{\partial}{\partial \mathbf{z}} \mathbf{g} &= \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{x}} - \sqrt{-1} \frac{\partial}{\partial \mathbf{y}} \right) \mathbf{g} \\ &= \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \right) + \frac{\sqrt{-1}}{2} \left(\frac{\partial \mathbf{v}}{\partial \mathbf{x}} - \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \right). \end{aligned}$$

Therefore

$$\operatorname{Re} \left(\frac{\partial \mathbf{g}}{\partial \mathbf{z}} \right) = \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \right) = \frac{1}{2} \left(\frac{\partial(\operatorname{Re} \mathbf{g})}{\partial(\operatorname{Re} \mathbf{z})} + \frac{\partial(\operatorname{Im} \mathbf{g})}{\partial(\operatorname{Im} \mathbf{z})} \right)$$

Notation 3

- ★ $\mathbf{G} = (\mathbf{g}_{ij})_{i=1,2,\dots,m,j=1,2,\dots,p}$ be an $m \times p$ matrix (\mathbf{g}_{ij} 's are complex-valued differentiable functions on $\mathbb{C}^{m \times p}$).
- ★ For $\mathbf{x} = (\mathbf{x}_{ij})_{i=1,2,\dots,m,j=1,2,\dots,p} \in \mathbb{C}^{m \times p}$, set

$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial \mathbf{x}_{ij}} \right)_{i=1,2,\dots,m,j=1,2,\dots,p},$$

and define

$$\begin{aligned} \operatorname{Re}(\operatorname{Tr}(\nabla'_{\mathbf{x}} \mathbf{G})) &= \operatorname{Tr}(\operatorname{Re}(\nabla'_{\mathbf{x}} \mathbf{G})) \\ &= \frac{1}{2} \sum_{j=1}^p \sum_{i=1}^m \left\{ \frac{\partial(\operatorname{Re} \mathbf{g}_{ij})}{\partial(\operatorname{Re} \mathbf{x}_{ij})} + \frac{\partial(\operatorname{Im} \mathbf{g}_{ij})}{\partial(\operatorname{Im} \mathbf{x}_{ij})} \right\}. \end{aligned}$$

Lemma (Stein's identity for complex normal distribution)

Let \mathbf{X} be a $m \times p$ matrix having $\mathbb{C}N_{m \times p}(\Xi, \mathbf{I}_m \otimes \Sigma)$ and let $\mathbf{G} = (\mathbf{g}_{ij})_{i=1,2,\dots,m,j=1,2,\dots,p}$ with

$$\mathbb{E} \left| \frac{\partial (\operatorname{Re} \mathbf{g}_{ij})}{\partial (\operatorname{Re} \mathbf{x}_{ij})} \right|_{\mathbf{x}=\mathbf{X}} < \infty, \quad \mathbb{E} \left| \frac{\partial (\operatorname{Im} \mathbf{g}_{ij})}{\partial (\operatorname{Im} \mathbf{x}_{ij})} \right|_{\mathbf{x}=\mathbf{X}} < \infty,$$

and $\mathbf{x} = (\mathbf{x}_{ij})_{i=1,2,\dots,m,j=1,2,\dots,p}$ in $\mathbb{C}^{m \times p}$. Then we have

$$\begin{aligned} & \mathbb{E}[\operatorname{Tr} \{ \Sigma^{-1} (\mathbf{X} - \Xi)^* \mathbf{G}(\mathbf{X}) + \operatorname{Tr} \{ \Sigma^{-1} \mathbf{G}^*(\mathbf{X})(\mathbf{X} - \Xi) \}] \\ & = 2 \mathbb{E}[\operatorname{Tr} (\operatorname{Re}(\nabla'_{\mathbf{x}} \mathbf{G}))|_{\mathbf{x}=\mathbf{X}}] \end{aligned}$$

Sketch of proof

Recall that, for $\mathbf{X} \sim \mathbb{C}N_p(\boldsymbol{\xi}, \boldsymbol{\Sigma})$, we have

$$[\mathbf{X}] := \begin{pmatrix} \operatorname{Re} \mathbf{X} \\ \operatorname{Im} \mathbf{X} \end{pmatrix} \sim N_{2p} \left[\underbrace{\begin{pmatrix} \operatorname{Re} \boldsymbol{\xi} \\ \operatorname{Im} \boldsymbol{\xi} \end{pmatrix}}_{[\boldsymbol{\xi}]: 2p \times 1}, \frac{1}{2} \underbrace{\begin{pmatrix} \operatorname{Re} \boldsymbol{\Sigma} & -\operatorname{Im} \boldsymbol{\Sigma} \\ \operatorname{Im} \boldsymbol{\Sigma} & \operatorname{Re} \boldsymbol{\Sigma} \end{pmatrix}}_{\{\boldsymbol{\Sigma}\}: 2p \times 2p} \right],$$

and note that

$$(\mathbf{X} - \boldsymbol{\xi})^* \boldsymbol{\Sigma}^{-1} \mathbf{g}(\mathbf{X}) + \mathbf{g}^*(\mathbf{X}) \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\xi}) = 2[\mathbf{X} - \boldsymbol{\xi}]' \{\boldsymbol{\Sigma}\}^{-1} [\mathbf{g}].$$

From Stein's identity for the multivariate real normal distribution, we have

$$\mathbb{E}[[\mathbf{X} - \boldsymbol{\xi}]' (\{\boldsymbol{\Sigma}\}/2)^{-1} [\mathbf{g}]] = \mathbb{E} \left[\sum_{i=1}^p \left(\frac{\partial(\operatorname{Re} g_i)}{\partial(\operatorname{Re} x_i)} + \frac{\partial(\operatorname{Im} g_i)}{\partial(\operatorname{Im} x_i)} \right) \Big|_{\mathbf{x}=\mathbf{X}} \right],$$

where $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ and $\mathbf{g}' = (g_1, g_2, \dots, g_p)$.

We consider a real normal case for simplicity.

For $Z \sim N(\xi, 1)$ and a function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that

$\mathbb{E}[g'(Z)] < \infty$, we have

$$\begin{aligned} & \int_{-\infty}^{\infty} (z - \xi)g(z) \exp(-(z - \xi)^2/2) dz \\ &= \underbrace{\left[g(z) \exp(-(z - \xi)^2/2) \right]_{-\infty}^{\infty}}_0 + \int_{-\infty}^{\infty} g'(z) \exp(-(z - \xi)^2/2) dz \end{aligned}$$

because $\int_{-\infty}^{\infty} g(z) \exp(-(z - \xi)^2/2) dz < \infty$. This shows that

$$\mathbb{E}[(Z - \xi)g(Z)] = \mathbb{E}[g'(Z)].$$

Note that $\mathbb{E}[g'(Z)] < \infty$ implies $\mathbb{E}[Zg(Z)] < \infty$ and $\mathbb{E}[g(Z)] < \infty$.

See Sourav Chatterjee's lecture note(Stein's method and applications).

Notation

★ Let $\mathbf{s} : \mathbf{p} \times \mathbf{p} = (\mathbf{s}_{jk})$ and $\mathbf{G}(\mathbf{s}) = (\mathbf{g}_{ij}(\mathbf{s}))$.

★

$$\frac{\partial}{\partial \mathbf{s}_{jk}} = \frac{1}{2}(1 + \delta_{jk}) \left\{ \frac{\partial}{\partial(\operatorname{Re} \mathbf{s}_{jk})} + (1 - \delta_{jk}) \sqrt{-1} \frac{\partial}{\partial(\operatorname{Im} \mathbf{s}_{jk})} \right\},$$

for $j, k = 1, 2, \dots, \mathbf{p}$ and δ_{jk} is the Kronecker delta.

★ Let $\mathbf{D}_S = (\partial/\partial \mathbf{s}_{jk})$ be a $\mathbf{p} \times \mathbf{p}$ operator matrix given by

$$\begin{aligned} \{\mathbf{D}_S \mathbf{G}(\mathbf{S})\}_{jk} &= \sum_{l=1}^{\mathbf{p}} \frac{\partial \mathbf{g}_{lk}}{\partial \mathbf{s}_{jl}}(\mathbf{S}) \\ &= \frac{1 + \delta_{jl}}{2} \sum_{l=1}^{\mathbf{p}} \left\{ \frac{\partial \mathbf{g}_{lk}}{\partial(\operatorname{Re} \mathbf{s}_{jl})} + (1 - \delta_{jl}) \sqrt{-1} \frac{\partial \mathbf{g}_{lk}}{\partial(\operatorname{Im} \mathbf{s}_{jl})} \right\} \Big|_{\mathbf{s}=\mathbf{S}}. \end{aligned}$$

Lemma (Stein-Haff identity for complex Wishart distribution)

Assume that each entry of $\mathbf{G}(\mathbf{S})$ is a partially differentiable function with respect to $\mathbf{Re} \mathbf{s}_{jk}$ and $\mathbf{Im} \mathbf{s}_{jk}$, $j, k = 1, 2, \dots, p$. Under some regularity conditions on $\mathbf{G}(\mathbf{S})$, the following identity holds:

$$\mathbb{E}[\mathrm{Tr}(\mathbf{G}(\mathbf{S})\boldsymbol{\Sigma}^{-1})] = \mathbb{E}[(n - p)\mathrm{Tr}(\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}) + \mathrm{Tr}(\mathbf{D}_{\mathbf{S}}\mathbf{G}(\mathbf{S}))]. \quad (1)$$

Sketch of proof

For $\mathbf{x} = (\mathbf{x}_{ij})_{i=1, \dots, n; j=1, \dots, p}$, note that

$$\frac{\partial}{\partial \mathbf{x}_{ij}} \exp(-\text{Tr } \mathbf{x}^* \mathbf{x}) = -\bar{\mathbf{x}}_{ij} \exp(-\text{Tr } \mathbf{x}^* \mathbf{x}),$$

Assume that $\mathbf{X} \sim \mathbb{C}\mathbf{N}_{n \times p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_p)$ and put

$\mathbf{S} = (\mathbf{s}_{ij}) = \mathbf{X}^* \mathbf{X} \sim \mathbb{C}\mathbf{W}(n, \mathbf{I}_p)$. Then we have

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^p \mathbf{s}_{ij} \mathbf{g}_{jl} \right] &= \mathbb{E} \left[\sum_{j=1}^p \sum_{k=1}^n \bar{\mathbf{x}}_{ki} \mathbf{x}_{kj} \mathbf{g}_{jl} \right] = \mathbb{E} \left[\sum_{j=1}^p \sum_{k=1}^n \frac{\partial}{\partial \mathbf{x}_{ki}} (\mathbf{x}_{kj} \mathbf{g}_{jl}) \right] \\ &= \mathbb{E} \left[n \mathbf{g}_{il} + \sum_{j=1}^p \sum_{k=1}^n \mathbf{x}_{kj} \frac{\partial \mathbf{g}_{jl}}{\partial \mathbf{x}_{ki}} \right]. \end{aligned}$$

In matrix form

$$\mathbb{E}[\mathbf{S}\mathbf{G}] = \mathbb{E}[\mathbf{n}\mathbf{G} + (\mathbf{X}'\nabla_{\mathbf{X}})'\mathbf{G}].$$

Transform $\mathbf{X} \rightarrow \mathbf{X}(\mathbf{P}^*)^{-1}$ ($\mathbf{S} \rightarrow \mathbf{P}^{-1}\mathbf{S}(\mathbf{P}^*)^{-1}$) and $\mathbf{G} \rightarrow \mathbf{P}^*\mathbf{G}(\mathbf{P}^*)^{-1}$, where $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}^*$ and \mathbf{P} is a $p \times p$ nonsingular matrix. Then

$$\mathbb{E}[\boldsymbol{\Sigma}^{-1}\mathbf{S}\mathbf{G}] = \mathbb{E}[\mathbf{n}\mathbf{G} + (\mathbf{X}'\nabla_{\mathbf{X}})'\mathbf{G}].$$

Replacing \mathbf{S} with $\mathbf{S}^{-1}\mathbf{G}$, taking the trace, and using the chain-rule of differentiation, we have

$$\mathbb{E}[\text{Tr } \boldsymbol{\Sigma}^{-1}\mathbf{G}] = \mathbb{E}[(n - p)\text{Tr } (\mathbf{G}\mathbf{S}^{-1}) + \text{Tr } (D_{\mathbf{S}}\mathbf{G})].$$

□

Class of estimators

- ★ Let $\mathbf{F} = \text{Diag}(f_1, f_2, \dots, f_{\min(m,p)})$ be the eigenvalues of $\mathbf{X}^* \mathbf{X} \mathbf{S}^{-1}$.
- ★ For $p > m$ decompose $\mathbf{X} \mathbf{S}^{-1} \mathbf{X}^* = \mathbf{U} \mathbf{F} \mathbf{U}^*$, where \mathbf{U} is an $m \times m$ unitary matrix.
- ★ For $m > p$ we decompose $\mathbf{S} = (\mathbf{A}^*)^{-1} \mathbf{A}^{-1}$ and $\mathbf{X}^* \mathbf{X} = (\mathbf{A}^*)^{-1} \mathbf{F} \mathbf{A}^{-1}$, where \mathbf{A} is a $p \times p$ non-singular matrix.
- ★ Consider a class of estimators of the form

$$\widehat{\Xi}_H := \widehat{\Xi}_H(\mathbf{X}, \mathbf{S}) = \begin{cases} \mathbf{X} \{ \mathbf{I}_p + \mathbf{A} \mathbf{H}(\mathbf{F}) \mathbf{A}^{-1} \} & \text{if } m > p \\ \{ \mathbf{I}_m + \mathbf{U} \mathbf{H}(\mathbf{F}) \mathbf{U}^* \} \mathbf{X} & \text{if } m < p \end{cases}, \quad (2)$$

where $\mathbf{H} := \mathbf{H}(\mathbf{F}) = \text{Diag}(h_1(\mathbf{F}), h_2(\mathbf{F}), \dots, h_{\min(m,p)}(\mathbf{F}))$
 whose i -th element $h_i := h_i(\mathbf{F})$, $i = 1, 2, \dots, \min(m, p)$, is
 a real-valued function on $\mathbb{R}_{\geq}^{\min(m,p)}$.

Theorem

Under the suitable conditions, we have

$$R((\Xi, \Sigma), \widehat{\Xi}_H) = \begin{cases} mp + \mathbb{E}[\widehat{\Delta}(n, m, p; \mathbf{H})] & \text{if } m > p \\ mp + \mathbb{E}[\widehat{\Delta}(n + m - p, p, m; \mathbf{H})] & \text{if } m < p \end{cases}$$

where $\mathbf{h}_{kk}(\mathbf{F}) = (\partial \mathbf{h}_k / \partial \mathbf{f}_k)(\mathbf{F})$ and

$$\begin{aligned} \widehat{\Delta}(n, m, p; \mathbf{H}) = & \sum_{k=1}^p \left\{ 2(m - p + 1)h_k(\mathbf{F}) + 2f_k h_{kk}(\mathbf{F}) \right. \\ & + 4 \sum_{b>k} \frac{f_k h_k(\mathbf{F}) - f_b h_b(\mathbf{F})}{f_k - f_b} + (n + p - 2)f_k h_k^2(\mathbf{F}) \\ & \left. - 2f_k^2 h_{kk}(\mathbf{F}) h_k(\mathbf{F}) - 2 \sum_{b>k} \frac{f_k^2 h_k^2(\mathbf{F}) - f_b^2 h_b^2(\mathbf{F})}{f_k - f_b} \right\} \mathbb{I}\{f_1 > \dots > f_p > 0\} \end{aligned}$$

Evaluation of risk

★ We have

$$\begin{aligned}
 R((\Xi, \Sigma), \widehat{\Xi}_H) &= \mathbb{E} \left[\text{Tr} \{ \Sigma^{-1} (\mathbf{X} + \mathbf{G} - \Xi)^* (\mathbf{X} + \mathbf{G} - \Xi) \} \right] \\
 &= \underbrace{\mathbb{E} \left[\text{Tr} \{ \Sigma^{-1} (\mathbf{X} - \Xi)^* (\mathbf{X} - \Xi) \} \right]}_{mp} \\
 &\quad + \underbrace{\mathbb{E} \left[\text{Tr} \{ \Sigma^{-1} \mathbf{G}^* (\mathbf{X} - \Xi) \} + \text{Tr} \{ \Sigma^{-1} (\mathbf{X} - \Xi)^* \mathbf{G} \} \right]}_{\text{Stein's identity}} \\
 &\quad + \underbrace{\mathbb{E} \left[\text{Tr} \{ \Sigma^{-1} \mathbf{G}^* \mathbf{G} \} \right]}_{\text{Stein and Haff identity}} \\
 &= mp + \mathbb{E} \left[2 \text{Tr} \{ \text{Re} (\nabla'_X \mathbf{G}) \} \right] \\
 &\quad + \mathbb{E} \left[\text{Tr} \{ \mathbf{D}_S \mathbf{G}^* \mathbf{G} \} + (n - p) \text{Tr} \{ \mathbf{S}^{-1} \mathbf{G}^* \mathbf{G} \} \right].
 \end{aligned}$$

- ★ When $m > p$, recall that $\widehat{\Xi}_H = \mathbf{X} + \mathbf{G} = \mathbf{X} + \mathbf{X}\mathbf{A}\mathbf{H}(\mathbf{F})\mathbf{A}^{-1}$, where \mathbf{A} is a $p \times p$ nonsingular matrix s.t. $\mathbf{S} = (\mathbf{A}^*)^{-1}\mathbf{A}^{-1}$ and $\mathbf{A}^*\mathbf{X}^*\mathbf{X}\mathbf{A} = \mathbf{F} = \text{Diag}(f_1, \dots, f_p)$, and that the modification \mathbf{H} depends only on the eigenvalues of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$, .

★

$$\begin{aligned}
 R((\Xi, \Sigma), \widehat{\Xi}_H) &= mp + \mathbb{E}\left[2 \text{Tr}\{\text{Re}(\nabla'_X \mathbf{G})\}\right] \\
 &\quad + \mathbb{E}\left[\text{Tr}\{\mathbf{D}_S \mathbf{G}^* \mathbf{G}\} + (n-p)\text{Tr}\{\mathbf{S}^{-1} \mathbf{G}^* \mathbf{G}\}\right]. \\
 &= mp + \mathbb{E}\left[2 \text{Tr}\{\text{Re}(\nabla'_X \mathbf{X}\mathbf{A}\mathbf{H}\mathbf{A}^{-1})\}\right] \\
 &\quad + \mathbb{E}\left[\text{Tr}\{\mathbf{D}_S \underbrace{(\mathbf{X}\mathbf{A}\mathbf{H}\mathbf{A}^{-1})^* (\mathbf{X}\mathbf{A}\mathbf{H}\mathbf{A}^{-1})}_{(\mathbf{A}^*)^{-1} \mathbf{H} \mathbf{F} \mathbf{H} (\mathbf{F}) \mathbf{A}^{-1}}}\right] \\
 &\quad + (n-p) \underbrace{\text{Tr}\{\mathbf{S}^{-1} (\mathbf{X}\mathbf{A}\mathbf{H}\mathbf{A}^{-1})^* (\mathbf{X}\mathbf{A}\mathbf{H}\mathbf{A}^{-1})\}}_{\text{Tr}\{\mathbf{F}\mathbf{H}^2\}}.
 \end{aligned}$$

- ★ The terms inside the first and second expectations are written as

$$\begin{aligned} & \text{Tr} \{ \text{Re} (\nabla'_X \mathbf{X} \mathbf{A} \mathbf{H} \mathbf{A}^{-1}) \} \\ &= \sum_{k=1}^p \left\{ f_k h_{kk} + (m - p + 1) h_{kk} + 2 \sum_{b>k} \frac{f_k h_k - f_b h_b}{f_k - f_b} \right\} \\ & \text{Tr} \{ \mathbf{D}_S ((\mathbf{A}^*)^{-1} \mathbf{H} \mathbf{F} \mathbf{H} \mathbf{A}^{-1}) \} \\ &= \sum_{k=1}^p \left\{ 2(p-1) f_k h_k^2 - 2 f_k h_{kk} h_k - 2 \sum_{b>k} \frac{f_k^2 h_k^2 - f_b^2 h_b^2}{f_k - f_b} \right\}, \end{aligned}$$

where $\mathbf{H} = \text{Diag}(h_1, \dots, h_p)$ and $h_{kk} = \partial h_k / \partial f_k$, and \mathbf{A} is a $p \times p$ nonsingular matrix such $\mathbf{S} = (\mathbf{A}^*)^{-1} \mathbf{A}^{-1}$, and that the modification \mathbf{H} depends only on the eigenvalues of $\mathbf{X}^* \mathbf{X} \mathbf{S}^{-1}$, $\mathbf{F} = \text{Diag}(f_1, f_2, \dots, f_p)$.

- ★ When $m < p$, recall that $\widehat{\Xi}_H := \mathbf{X} + \mathbf{G}$ where $\mathbf{G} = \mathbf{U}\mathbf{H}\mathbf{U}^*\mathbf{X}$, where \mathbf{U} is an $m \times m$ unitary matrix such that $\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^* = \mathbf{U}\mathbf{F}\mathbf{U}^*$, and recall that an $m \times m$ diagonal matrix \mathbf{H} depends only on $\mathbf{F} = \text{Diag}(f_1, f_2, \dots, f_m)$.
- ★ For $m < p$, we have

$$\begin{aligned}
 R((\Xi, \Sigma), \widehat{\Xi}_H) &= mp + \mathbb{E}\left[2 \text{Tr}\{\text{Re}(\nabla'_X \mathbf{G})\}\right] \\
 &\quad + \mathbb{E}\left[\text{Tr}\{\mathbf{D}_S \mathbf{G}^* \mathbf{G}\} + (n-p) \text{Tr}\{\mathbf{S}^{-1} \mathbf{G}^* \mathbf{G}\}\right]. \\
 &= mp + \mathbb{E}\left[2 \text{Tr}\{\text{Re}(\mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\mathbf{X})\}\right] \\
 &\quad + \mathbb{E}\left[\text{Tr}\{\mathbf{D}_S \underbrace{(\mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\mathbf{X})^* (\mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\mathbf{X})}_{\mathbf{X}^* \mathbf{U} \mathbf{H}^2 \mathbf{U}^* \mathbf{X}}\}\right] \\
 &\quad + (n-p) \underbrace{\text{Tr}\{\mathbf{S}^{-1} (\mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\mathbf{X})^* (\mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\mathbf{X})\}}_{\text{Tr}\{\mathbf{F}\mathbf{H}^2\}}.
 \end{aligned}$$

- ★ The terms inside the first and second expectations are written as

$$\begin{aligned} & \text{Tr} \{ \text{Re} (\nabla'_X \mathbf{U} \mathbf{H} \mathbf{U}^* \mathbf{X}) \} \\ &= \sum_{k=1}^m \left\{ f_k h_{kk} + (p - m + 1) h_{kk} + 2 \sum_{b>k} \frac{f_k h_k - f_b h_b}{f_k - f_b} \right\} \\ & \text{Tr} \{ \mathbf{D}_S (\mathbf{X}^* \mathbf{U} \mathbf{H}^2 \mathbf{U} \mathbf{X}) \} \\ &= \sum_{k=1}^m \left\{ 2(m-1) f_k h_k^2 - 2 f_k^2 h_{kk} h_k - 2 \sum_{b>k} \frac{f_k^2 h_k - f_b^2 h_b}{f_k - f_b} \right\}, \end{aligned}$$

where $\mathbf{H} = \text{Diag}(h_1, \dots, h_m)$ and $h_{kk} = \partial h_k / \partial f_k$, and \mathbf{U} is an $m \times m$ unitary matrix s.t.

$$\mathbf{X} \mathbf{S}^{-1} \mathbf{X}^* = \mathbf{U} \text{Diag}(f_1, f_2, \dots, f_m) \mathbf{U}^*.$$

Theorem (Efron-Morris type estimator)

Let

$$\mathbf{c} = \begin{cases} \frac{m-p}{n+p} & \text{if } m > p \\ \frac{p-m}{n+2m-p} & \text{if } m < p \end{cases}$$

and put

$$\mathbf{H}(\mathbf{F}) = -\text{Diag} \left(\frac{\mathbf{c}}{f_1}, \frac{\mathbf{c}}{f_2}, \dots, \frac{\mathbf{c}}{f_{\min(m,p)}} \right)$$

Then the estimator

$$\hat{\Xi}_H = \begin{cases} \mathbf{X} \{ \mathbf{I}_p - \mathbf{c}(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{S} \} & \text{if } m > p \\ \{ \mathbf{I}_m - \mathbf{c}(\mathbf{X} \mathbf{S}^{-1} \mathbf{X}^*)^{-1} \} \mathbf{X} & \text{if } m < p \end{cases}$$

improves upon \mathbf{X} , i.e., $R((\Xi, \Sigma), \hat{\Xi}_H) \leq R((\Xi, \Sigma), \mathbf{X})$ for $\forall (\Xi, \Sigma)$.

Theorem

Let $\mathbf{c}_k = \frac{m+p-2k}{n-p+2k}$ ($k = 1, 2, \dots, \min(m, p)$) and put

$$\mathbf{H}(\mathbf{F}) = -\text{Diag} \left(\frac{\mathbf{c}_1}{f_1}, \frac{\mathbf{c}_2}{f_2}, \dots, \frac{\mathbf{c}_{\min(m,p)}}{f_{\min(m,p)}} \right).$$

Then the estimator

$$\widehat{\Xi} = \begin{cases} \mathbf{X}\{\mathbf{I}_p + \mathbf{A}\mathbf{H}(\mathbf{F})\mathbf{A}^{-1}\} & \text{if } m > p \\ \{\mathbf{I}_m + \mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\}\mathbf{X} & \text{if } m > p \end{cases}$$

improves upon \mathbf{X} , i.e., $R((\Xi, \Sigma), \widehat{\Xi}_H) \leq R((\Xi, \Sigma), \mathbf{X})$ for $\forall(\Xi, \Sigma)$.

Recall that \mathbf{U} is an $m \times m$ unitary matrix s.t. $\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^* = \mathbf{U}\mathbf{F}\mathbf{U}^*$ and that \mathbf{A} is a $p \times p$ nonsingular matrix s.t. $\mathbf{S} = (\mathbf{A}^*)^{-1}\mathbf{A}^{-1}$.

- 1 The results are complex version of those for estimation problem for the mean matrix of the real multivariate normal distribution (see Konno (JMVA, 1991) and Kariya et al. (JMVA, 1999)).
- 2 However the results are parallel to real case.
- 3 Note that the substitution rule to get the second part from the first part is $(\boldsymbol{\rho}, \boldsymbol{m}, \boldsymbol{n}) \rightarrow (\boldsymbol{m}, \boldsymbol{\rho}, \boldsymbol{n} + \boldsymbol{m} - \boldsymbol{\rho})$, i.e.,

$$R((\Xi, \Sigma), \hat{\Xi}_H) = \begin{cases} \boldsymbol{m}\boldsymbol{\rho} + \mathbb{E}\left[\widehat{\Delta}(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{\rho}; \mathbf{H})\right] & \text{if } \boldsymbol{m} > \boldsymbol{\rho} \\ \boldsymbol{m}\boldsymbol{\rho} + \mathbb{E}\left[\widehat{\Delta}(\boldsymbol{n} + \boldsymbol{m} - \boldsymbol{\rho}, \boldsymbol{\rho}, \boldsymbol{m}; \mathbf{H})\right] & \text{if } \boldsymbol{m} < \boldsymbol{\rho} \end{cases}$$

This is the same as the rule substitution for the joint distribution of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.

When $\Xi = \mathbf{0}$, the joint distribution of $\mathbf{X}^* \mathbf{X} \mathbf{S}^{-1}$ is, apart from normalizing constants,

$$\prod_{k=1}^p \frac{f_k^{m-p}}{(1+f_k)^{n+m}} \prod_{k=1}^{p-1} \prod_{j=k+1}^p (f_k - f_j)^2 \prod_{k=1}^p df_k$$

if $m > p$ while it is

$$\prod_{k=1}^m \frac{f_k^{p-m}}{(1+f_k)^{n+m}} \prod_{k=1}^{m-1} \prod_{j=k+1}^m (f_k - f_j)^2 \prod_{k=1}^m df_k,$$

if $p > m$. The second one is obtained by the substitution rule $(p, m, n) \rightarrow (m, p, n + m - p)$. Good reference is Khatri(1965, Ann. ,Math. Statist).