1-10-2024

On a Class of James-Stein's Estimators in High-Dimensional Data

Arash Aghaei Foroushani
University of Windsor

On a Class of James-Stein's Estimators in High-Dimensional Data

by

Arash Aghaei Foroushani

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

On a Class of James-Stein's Estimators in High-Dimensional Data

by

Arash Aghaei Foroushani

APPROVED BY:

D. Li
Department of Economics

A. Hussein
Department of Mathematics and Statistics

S. Nkurunziza, Advisor
Department of Mathematics and Statistics

December 14, 2023

Declaration of Co-Authorship / Previous Publication

I. Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research, as follows: some parts of Chapters 2, 3 of the thesis were co-authored with Prof. Sévérien Nkurunziza. In all cases, the primary contributions, simulation, data analysis, interpretation, and writing were performed by the author, and the contribution of co-authors was primarily through the provision of some theoretical results.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Previous Publication

This thesis includes one original paper that has been previously published/submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication title/full citation	Publication status
Some parts of Chapter 3	Arash A. Aghaei, Sévérien Nkurunziza, 2023. Improved Gaussian Mean Matrix Estimators In High-Dimensional Data. Annals of Statistics (Submitted)	Under review

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

In this thesis, we consider the estimation problem of the mean matrix of a multivariate normal distribution in high-dimensional data. Building upon the groundwork laid by Chételat and Wells (2012), we extend their method to the cases where the parameter is the mean matrix of a matrix normal distribution. In particular, we propose a novel class of James-Stein's estimators for the mean matrix of a multivariate normal distribution with an unknown row covariance matrix and independent columns. Given a realistic assumption, we establish that our proposed estimator outperforms the classical maximum likelihood estimator (MLE) in the context of high-dimensional data. Furthermore, we investigate the conditions for which this assumption remains valid. Additionally, we identify and rectify a notable error in the proofs of a crucial result presented in Chételat and Wells (2012). Notably, the novelty of the obtained results lies in the fact that the estimator for the row covariance matrix is singular almost surely and its rank is a random variable. Finally, we present simulation results that confirm the validity of our theoretical findings.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Nkurunziza for his constant support during the career of my master. He is patient and kind all the time. Under his supervision, I feel relaxed and energized, which make me finish this thesis smoothly. I am also very grateful to all the members in my defense committee. Particularly, I would like to thank Dr. Li from department of Economics for agreeing to be my external program reader. I would also like to thank Dr. Hussein for his constant support during my graduate program.

Besides professors, I am grateful to my parents for their significant support. Without their support and care I could not finish my courses and thesis.

Finally, I would like to offer my sincere thanks to all students, faculty members and staff in the department of Mathematics and Statistics for the harmonic, friendly and positive studying and working environment.

Contents

List of Figures

Chapter 1

Introduction and Contributions

Chételat and Wells (2012) introduced a new type of estimator, based on the class of estimators proposed by Baranchik (1970). This estimator dominates the classical maximum likelihood estimator (MLE) of the mean vector in a multivariate normal distribution in high-dimensional settings. However, an error in proving one of the main results presented by Chételat and Wells (2012) motivates us to revise some of their findings. This revision not only prompts a reconsideration of their work but also encourages us to explore the problem of estimating the mean matrix in a matrix normal distribution.

In particular, we consider to estimate the mean matrix of a random matrix from a matrix normal distribution. Initially, it might seem that the classical MLE is the most suitable estimator for the mean matrix. However, in 1956, Charles Stein (refer to Stein (1956)) discovered that the classical MLE of mean vector of a p-dimensional normal random vector loses its admissibility under the quadratic loss in high-dimensional data. This finding implies the existence of alternative estimators for the mean vector that outperform the classical MLE under the aforementioned loss function. Stein (1960) introduced a novel class of biased but minimax estimators. This class of estimators dominates the classical MLE under the invariant quadratic loss.

In this thesis, our primary focus is on the generalized estimator introduced by Baranchik

(1970), particularly in the context of unknown covariance in high-dimensional data. The classical estimator in Baranchik (1970) relies on the use of traditional inverse of the covariance matrix estimator, which becomes impractical in high-dimensional settings. Indeed, in high-dimensional data, the estimator of the covariance matrix becomes singular almost surely. To overcome this problem, we utilize the Moore-Penrose inverse, instead of the traditional inverse. Because of that, classical techniques cannot be used to prove the risk dominance of the proposed class of estimators over classical MLE. Thus, the additional novelty of this thesis lies in deriving some mathematical results which are useful in establishing the risk dominance of the proposed estimators over MLE.

1.1 Organization of the thesis

This thesis is organized in 5 chapters including this chapter which gives an introduction. In Chapter 2, we begin by discussing key concepts that play a pivotal role in proving the main results and lemmas throughout this thesis. Subsequently, we present the central thesis result within the multivariate setting. Additionally, we introduce several propositions and lemmas that are essential components in demonstrating the main result outlined in Theorem 2.2. In Chapter 3, we extend the findings from Chapter 2 to the matrix normal distribution setting with an unknown row covariance and independent columns. In Chapter 4, we conduct a simulation study to validate numerically the theoretical findings presented in this thesis. In Chapter 5, we give some concluding comments. We also introduce in Chapter 5 valuable insights and ideas to serve as potential directions for future research. Finally, for the convenience of the reader, some technical proofs as well as the simulation R code are given in the Appendix A.

Chapter 2

Improved Multivariate Normal Mean Estimation

In this chapter, suppose that $Z_1, \dots, Z_N$ are independent and identically distributed random samples from $\mathcal{N}_p(\theta, \Psi)$ where Ψ represents the covariance matrix and is an unknown matrix. Then, $Z = [Z_1, \dots, Z_N]^\top$ follows $\mathcal{N}_{N\times p}(e\theta^\top, I_N \otimes \Psi)$ where $e = [1, \dots, 1]^\top$ is an N-dimensional vector. Let $X = \bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$. Therefore, $X \sim \mathcal{N}p(\theta, \Sigma)$ where $\Sigma = \frac{\Psi}{N}$. Let us consider $S = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \bar{Z})(Z_i - \bar{Z})^\top$ as an estimator of Σ and $n = N - 1$. In Appendix A.1, We show that S can be written as $S = Y^\top Y$, where Y is independent of X and follows a matrix normal distribution $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$. This implies that $S \sim Wishart_p(n, \Sigma)$.

Based on the findings from Srivastava and Khatri (1979), it is established that the matrix S is almost surely invertible when the dimensionality p is less than or equal to the sample size n, i.e., $p \leq n$. Conversely, it is almost surely singular when the dimensionality p exceeds the sample size n, i.e., $p > n$. Moreover, it has been demonstrated in Srivastava and Khatri (1979) and Srivastava (2003) that the rank of the estimator of the covariance matrix, denoted as S, is equal to the minimum of the number of observations n and the number of features

(p), almost surely.

In estimating the mean vector, denoted as θ, the unbiased maximum likelihood estimator is $\delta^0 = X$. However, according to the findings presented by Stein (1956), X becomes inadmissible under the quadratic loss function defined as $L(\theta, \delta) = (\delta - \theta)^\top(\delta - \theta)$ when $n \geq p \geq 3$.

To address the limitations of the estimator $\delta^0 = X$, especially when $n \geq p \geq 3$, Baranchik (1970) introduced a new James-Stein type of estimator, given by:

$$\delta(X, S) = \left(I - \frac{r(X^\top S^{-1} X)}{X^\top S^{-1} X}\right) X.$$

Here, the function r represents a positive, bounded, and differentiable real valued function. When the conditions $n \geq p \geq 3$ hold, this estimator is known to dominate the estimator X under the invariant quadratic loss. However, when p exceeds the sample size n, the estimator S is singular almost surely, rendering the above estimator unusable in such cases.
To overcome this issue, the Moore-Penrose inverse of S, denoted as S^+, is employed to formulate a modified Baranchik (1970) estimator:

$$\delta(X, S) = \left(I - \frac{r(X^\top S^+ X)}{X^\top S^+ X} S S^+\right) X.$$

This modification allows for a robust estimator that can handle situations where $p > n$, making it a valuable tool for estimating the mean vector θ under the given conditions.

In Section 2.2, we show that under the invariant quadratic loss, the above estimator dominates the usual estimator X. We also provide in Appendix A.2, some important concepts on the Moore-Penrose inverse and Stein's Lemma (see Stein (1981)). These concepts play a crucial role in establishing Theorem 2.2 and Theorem 3.3 which are the main results of this thesis.

To simplify the presentation of this thesis, let us introduce some notations. For $m \times n$

matrices A and B, define

$$A.B = \sum_{i,j} A_{ij} B_{ij}.$$

For special case of m-dimensional vectors A and B, we have

$$A.B = \sum_{i} A_i B_i = A^\top B.$$

Let $\text{vec}(A)$ and $\text{vec}(B)$ be the transformation of A and B to vectors of dimension mn. We have

$$A.B = \text{vec}(A).\text{vec}(B) = \text{vec}(A)^\top \text{vec}(B) = \sum_{i,j} A_{ij} B_{ij}.$$

Similarly, for $\nabla_A = \left(\frac{\partial}{\partial A_{ij}}\right)_{1 \leq i \leq m, 1 \leq j \leq n}$, define

$$\text{div}_A B = \nabla_A . B = \text{div}_{\text{vec}(A)} \text{vec}(B) = \sum_{i,j} \frac{\partial B_{ij}}{\partial A_{ij}},$$

and

$$(\nabla_A B)_{ij} = \sum_{\alpha} (\nabla_A)_{i\alpha} B_{\alpha j} = \sum_{\alpha} \frac{\partial B_{\alpha j}}{\partial A_{i\alpha}}.$$

Furthermore, let $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, be the Kronecker delta.

Before delving into the main result of this thesis in multivariate normal distribution setting, we provide the important lemmas, propositions and their corresponding proofs. These propositions lay the groundwork for the proof of Theorem 2.2.

2.1 Important Preliminary Results

In this section, we present crucial lemmas and propositions which are vital for proving some of the main results of this thesis as given in Section 2.2. To ensure the coherence of this thesis, several proofs have been moved to Appendix A.

Lemma 2.1. *Let Y be an $n \times p$ matrix and $S = Y^\top Y$. Let X be a p vector and $F = X^\top S^+ X$. Let $A \in M_{k\times p}$ and $B \in M_{p\times h}$, it then follows that*

$$(i) \quad \left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \delta_{\beta k} Y_{\alpha l} + \delta_{\beta l} Y_{\alpha k},$$

$$(ii) \quad \left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + (AY^\top)_{k\alpha} B_{\beta l},$$

$$(iii) \quad \frac{\partial F}{\partial Y_{\alpha\beta}} = -2(X^\top S^+ Y^\top)_\alpha (S^+X)_\beta + 2(X^\top S^+ S^+ Y^\top)_\alpha ((I - SS^+)X)_\beta,$$

$$\begin{aligned}(iv) \quad \left(\frac{\partial S^+XX^TSS^+}{\partial Y_{\alpha\beta}}\right)_{kl} &= -S^+_{k\beta}(YS^+XX^TSS^+)_{\alpha l} - (S^+Y^T)_{k\alpha}(S^+XX^TSS^+)_{\beta l} \\ &+ (I - SS^+)_{k\beta}(YS^+SXX^TSS^+)_{\alpha l} + (S^+S^+Y^T)_{k\alpha}((I - SS^+)XX^TSS^+)_{\beta l} \\ &+ (S^+XX^T)_{k\beta}(YS^+)_{\alpha l} + (S^+XX^TY^T)_{k\alpha}S^+_{\beta l} - (S^+XX^TSS^+)_{k\beta}(YS^+)_{\alpha l} \\ &- (S^+XX^TSS^+Y^T)_{k\alpha}S^+_{\beta l} + (S^+XX^TS^+Y^T)_{k\alpha}(I - SS^+)_{\beta l}.\end{aligned}$$

Proof. The proof of this result is given in Appendix A.3. □

Lemma 2.2. *Let Y be an $n \times p$ matrix and $S = Y^\top Y$. Let X be a p vector, $F = X^\top S^+ X$, and $G(X, S) = \frac{r^2(F)}{F^2}(S^+XX^\top SS^+)$, where r is a differentiable function. Then*

$$\begin{aligned}(i) \quad \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} =& \frac{2r(F)r'(F)}{F^2}\left(\frac{\partial F}{\partial y_{\alpha\beta}}\right)(S^+XX^\top SS^+)_{kl} - \frac{2r^2(F)}{F^3}\left(\frac{\partial F}{\partial y_{\alpha\beta}}\right)(S^+XX^\top SS^+)_{kl} \\ &+ \frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl},\end{aligned}$$

(ii) $$\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial Y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k} = -2F^2,$$

(iii) $$\sum_{\alpha,k,\beta} Y_{\alpha k}\frac{\partial}{\partial Y_{\alpha\beta}}(SS^+XX^\top S^+)_{\beta k} = F(p - 2\mathrm{tr}(SS^+) - 1),$$

(iv) $$\sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = -4r(F)r'(F) + \frac{r^2(F)}{F}\Big(p - 2\mathrm{tr}(SS^+) + 3\Big).$$

Proof. The proof of this result is given in Appendix A.4. □

Lemma 2.3. *Let Y be an $n \times p$ matrix and $S = Y^\top Y$. Let X be a p vector, $F = X^\top S^+ X$, and $g(X,S) = \frac{r(F)}{F}(SS^+X)$, where r is a differentiable function. Then*

(i) $$\frac{\partial F}{\partial X_i} = 2(S^+X)_i,$$

(ii) $$\left(\frac{\partial SS^+X}{\partial X_i}\right)_k = (SS^+)_{ki},$$

(iii) $$\frac{\partial g_k}{\partial X_i} = \frac{2(Fr'(F) - r(F))}{F^2}(S^+X)_i(SS^+X)_k + \frac{r(F)}{F}(SS^+)_{ki},$$

(iv) $$\sum_i \frac{\partial g_i}{\partial X_i} = 2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2).$$

Proof. The proof of this result is given in Appendix A.3. □

The first part of the following proposition is referenced in the proof of the main result in Chételat and Wells (2012) but it is left without proof. In this thesis, we offer a detailed proof utilizing Corollary A.2 (Stein's Lemma). Additionally, it is essential to note that the existence of the right-side expectation must hold. The conditions for the existence of this expectations will be given in Theorem 2.1.

Proposition 2.1. *Let $X \sim \mathcal{N}_p(\theta, \Sigma)$. Let $g(X, S)$ be a differentiable p vector function. Then*

$$\mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}(X-\theta)\right] = \mathrm{E}_\theta\left[\nabla_X . g(X,S)\right],$$

provided that $\mathrm{E}_\theta\left[|\nabla_X . g(X,S)|\right] < \infty$.

Proof. Let $\tilde{X} = A^{-1}(X-\theta)$ where A is a symmetric positive definite square root of Σ. Thus $\tilde{X} \sim \mathcal{N}_p(0, I_p)$. Therefore $X_i \sim \mathcal{N}(0,1)$. Let $h = A^{-1}g(X,S)$. Then, we have

$$g^\top(X,S)\Sigma^{-1}(X-\theta) = g^\top(X,S)A^{-1}A^{-1}(X-\theta).$$

Then,

$$g^\top(X,S)\Sigma^{-1}(X-\theta) = h^\top \tilde{X} = \sum_j h_{1j}^\top \tilde{X}_{j1}. \tag{2.1}$$

Therefore, by (2.1), we have

$$\mathrm{E}\left[g^\top(X,S)\Sigma^{-1}(X-\theta)\right] = \mathrm{E}\left[\sum_j h_{1j}^\top \tilde{X}_{j1}\right] = \sum_j \mathrm{E}\left[h_{1j}^\top \tilde{X}_{j1}\right] = \sum_j \mathrm{E}\left[\tilde{X}_{j1} h_{1j}^\top\right].$$

Therefore, by Corollary A.2, we get

$$\sum_j \mathrm{E}[\tilde{X}_{j1}h_{1j}^\top] = \sum_j \mathrm{E}\left[\frac{\partial}{\partial \tilde{X}_{j1}} h_{1j}^\top\right] = \sum_j \mathrm{E}\left[\frac{\partial}{\partial \tilde{X}_{j1}} h_{j1}\right] = \mathrm{E}\left[\sum_j \frac{\partial}{\partial \tilde{X}_{j1}} h_{j1}\right],$$

then,

$$\sum_j \mathrm{E}[\tilde{X}_{j1} h_{1j}^\top] = \mathrm{E}\Bigg[\sum_j \frac{\partial}{\partial \tilde{X}_{j1}} (A^{-1} g(X,S))_{j1}\Bigg] = \mathrm{E}\Bigg[\sum_j \frac{\partial}{\partial \tilde{X}_{j1}} \sum_k A_{jk}^{-1} g(X,S)_{k1}\Bigg].$$

This gives

$$\sum_j \mathrm{E}[\tilde{X}_{j1} h_{1j}^\top] = \mathrm{E}\Bigg[\sum_{j,k} A_{jk}^{-1} \frac{\partial}{\partial \tilde{X}_{j1}} g(X,S)_{k1}\Bigg]. \tag{2.2}$$

Now, by applying the chain rule in (2.2), we have

$$\begin{aligned} \mathrm{E}\Bigg[\sum_{j,k} A_{jk}^{-1} \frac{\partial}{\partial \tilde{X}_{j1}} g(X,S)_{k1}\Bigg] &= \mathrm{E}\Bigg[\sum_{j,k} A_{jk}^{-1} \sum_l \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} \frac{\partial X_{l1}}{\partial \tilde{X}_{j1}}\Bigg] \\ &= \mathrm{E}\Bigg[\sum_{j,k,l} A_{jk}^{-1} \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} \frac{\partial X_{l1}}{\partial \tilde{X}_{j1}}\Bigg]. \end{aligned} \tag{2.3}$$

Since $\tilde{X} = A^{-1}(X - \theta)$, we have

$$X_{l1} = \sum_t A_{lt} \tilde{X}_{t1} + \theta_{l1},$$

thus

$$\frac{\partial X_{l1}}{\partial \tilde{X}_{j1}} = \sum_t A_{lt} \frac{\partial \tilde{X}_{t1}}{\partial \tilde{X}_{j1}} = \sum_t A_{lt} \delta_{tj} = A_{lj}. \tag{2.4}$$

Therefore, by replacing (2.4) in (2.3) we get

$$\begin{aligned} &\mathrm{E}\Bigg[\sum_{j,k,l} A_{jk}^{-1} \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} \frac{\partial X_{l1}}{\partial \tilde{X}_{j1}}\Bigg] = \mathrm{E}\Bigg[\sum_{j,k,l} A_{jk}^{-1} \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} A_{lj}\Bigg] \\ &= \mathrm{E}\Bigg[\sum_{k,l} \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} \sum_j A_{lj} A_{jk}^{-1}\Bigg] = \mathrm{E}\Bigg[\sum_{k,l} \frac{\partial}{\partial X_{l1}} g(X,S)_{k1} (AA^{-1})_{lk}\Bigg]. \end{aligned}$$

This gives

$$\mathrm{E}\bigg[\sum_{j,k,l} A^{-1}_{jk}\frac{\partial}{\partial X_{l1}}g(X,S)_{k1}\frac{\partial X_{l1}}{\partial \tilde{X}_{j1}}\bigg] = \mathrm{E}\bigg[\sum_{k}\frac{\partial}{\partial X_{k1}}g(X,S)_{k1}\bigg] = \mathrm{E}\Big[\nabla_X.g(X,S)\Big],$$

which completes the proof. □

In the upcoming proposition, we present an enhanced version of Lemma 3 from Chételat and Wells (2012). In Part (*iii*), we utilize Corollary A.2. This outcome relies on the existence of $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$, a concept that will be thoroughly examined in Part (*i*) of Theorem 2.1.

Proposition 2.2. *Let* $X \sim \mathcal{N}_p(\theta,\Sigma)$ *and* $Y \sim \mathcal{N}_{n\times p}(0, I_n\otimes\Sigma)$*. Let* $S = Y^\top Y$*. For* A *symmetric positive definite square root of* Σ *(i.e.* $A^2=\Sigma$*) define* $\tilde{Y} = YA^{-1}$, $\tilde{S} = \tilde{Y}^\top\tilde{Y}$ *and* $H = AGA^{-1}$ *where* $G(X,S)$ *is a* $p\times p$ *differentiable matrix function. Then*

(*i*) $\mathrm{tr}(\Sigma^{-1}SG) = \mathrm{tr}(\tilde{S}H)$,

(*ii*) $\mathrm{tr}(\tilde{S}H) = \mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)$,

(*iii*) $\mathrm{E}\left[\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right]$,

provided that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty$,

(*iv*)
$$\begin{aligned}\nabla_{\tilde{Y}}.(\tilde{Y}H) &= \mathrm{div}_{\mathrm{vec}(\tilde{Y})}.\mathrm{vec}(\tilde{Y}H) = n\mathrm{tr}(G) + \mathrm{tr}(Y^\top(\nabla_Y G^\top))\\ &= n\mathrm{tr}(G) + \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.\end{aligned}$$

Proof.

(*i*) $\mathrm{tr}(\tilde{S}H) = \mathrm{tr}(A^{-1}SA^{-1}AGA^{-1}) = \mathrm{tr}(A^{-1}SGA^{-1})$.

This gives

$$\mathrm{tr}(\tilde{S}H) = \mathrm{tr}(A^{-1}A^{-1}SG) = \mathrm{tr}(A^{-2}SG) = \mathrm{tr}(\Sigma^{-1}SG).$$

(ii) $\mathrm{tr}(\tilde{S}H) = \sum_{i} (\tilde{S}H)_{ii} = \sum_{i,j} \tilde{S}_{ij} H_{ji} = \sum_{i,j} (\tilde{Y}^{\top}\tilde{Y})_{ij} H_{ji} = \sum_{i,j,k} \tilde{Y}^{\top}_{ik} \tilde{Y}_{kj} H_{ji}$

$$= \sum_{i,j,k} \tilde{Y}_{ki}\tilde{Y}_{kj} H_{ji} = \sum_{i,k} \tilde{Y}_{ki} \sum_{j} \tilde{Y}_{kj} H_{ji}.$$

Hence,

$$\mathrm{tr}(\tilde{S}H) = \sum_{i,k} \tilde{Y}_{ki}(\tilde{Y}H)_{ki} = \mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H).$$

(iii) Since $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$, we get $\tilde{Y} = YA^{-1} \sim \mathcal{N}_{n\times p}(0, I_n \otimes I_p)$.

Then

$$\mathrm{vec}(\tilde{Y}) \sim \mathcal{N}_{np}(0, I_{np}),$$

Therefore,

$$\tilde{Y}_{\alpha i} \sim \mathcal{N}(0, 1).$$

Also, we have

$$\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H) = \sum_{\alpha,i} \tilde{Y}_{\alpha i}(\tilde{Y}H)_{\alpha i} = \sum_{\alpha,i} \tilde{Y}_{\alpha i} \sum_{j} \tilde{Y}_{\alpha j} H_{ji} = \sum_{\alpha,i,j} \tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j} H_{ji}.$$

Therefore, we have

$$\mathrm{E}\left[\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)\right] = \mathrm{E}\left[\sum_{\alpha,i,j} \tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j}H_{ji}\right] = \sum_{\alpha,i,j} \mathrm{E}\left[\tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j}H_{ji}\right] = \sum_{\alpha,i,j} \mathrm{E}\left[\tilde{Y}_{\alpha i}g_j(\tilde{Y}_{\alpha i})\right],$$

where $g_j(\tilde{Y}_{\alpha i}) = \tilde{Y}_{\alpha j}H_{ji}$.Therefore, by Corollary A.2, we get

$$\sum_{\alpha,i,j} \mathrm{E}\left[\tilde{Y}_{\alpha i}g_j(\tilde{Y}_{\alpha i})\right] = \sum_{\alpha,i,j} \mathrm{E}\left[\frac{\partial}{\partial \tilde{Y}_{\alpha i}} g_j(\tilde{Y}_{\alpha i})\right] = \mathrm{E}\left[\sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} g_j(\tilde{Y}_{\alpha i})\right].$$

Then,

$$\sum_{\alpha,i,j} \mathrm{E}\left[\tilde{Y}_{\alpha i}g_j(\tilde{Y}_{\alpha i})\right] = \mathrm{E}\left[\sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}H_{ji}\right] = \mathrm{E}\left[\sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{j} \tilde{Y}_{\alpha j}H_{ji}\right].$$

Therefore,

$$\sum_{\alpha,i,j} \mathrm{E}\left[\tilde{Y}_{\alpha i}g_j(\tilde{Y}_{\alpha i})\right] = \mathrm{E}\left[\sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} (\tilde{Y}H)_{\alpha i}\right] = \mathrm{E}\left[\nabla_{\tilde{Y}}.(\tilde{Y}H)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right].$$

$$(iv) \quad \nabla_{\tilde{Y}}.(\tilde{Y}H) = \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \sum_{\alpha,i} (\mathrm{div}_{\tilde{Y}})_{\alpha i}(\tilde{Y}H)_{\alpha i} = \sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{j} \tilde{Y}_{\alpha j}H_{ji}.$$

Then,

$$\nabla_{\tilde{Y}}.(\tilde{Y}H) = \sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} (\tilde{Y}_{\alpha j}H_{ji}) = \sum_{\alpha,i,j} \left((\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j})H_{ji} + \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) \right).$$

Hence,

$$\nabla_{\tilde{Y}}.(\tilde{Y}H) = \sum_{\alpha,i,j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j})H_{ji} + \sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}). \tag{2.5}$$

By applying the chain rule in the second term we get

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j} \tilde{Y}_{\alpha j} \sum_{k,\beta} \frac{\partial}{\partial Y_{k\beta}} H_{ji} \frac{\partial Y_{k\beta}}{\partial \tilde{Y}_{\alpha i}} = \sum_{\alpha,i,j,k,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} \frac{\partial (\tilde{Y} A)_{k\beta}}{\partial \tilde{Y}_{\alpha i}}$$

$$= \sum_{\alpha,i,j,k,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{l} \tilde{Y}_{kl} A_{l\beta}).$$

This gives

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,k,\beta,l} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\frac{\partial \tilde{Y}_{kl}}{\partial \tilde{Y}_{\alpha i}}) A_{l\beta} = \sum_{\alpha,i,j,k,\beta,l} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\delta_{\alpha k} \delta_{il}) A_{l\beta}.$$

Hence,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} H_{ji} A_{i\beta} = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} (AGA^{-1})_{ji} A_{i\beta}.$$

Then,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} (\sum_{k,l} A_{jk} G_{kl} A^{-1}_{li}) A_{i\beta} = \sum_{\alpha,i,j,\beta,k,l} \tilde{Y}_{\alpha j} A_{jk} \frac{\partial}{\partial Y_{\alpha\beta}} G_{kl} A^{-1}_{li} A_{i\beta}.$$

This gives,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,\beta,k,l} (\sum_{j} \tilde{Y}_{\alpha j} A_{jk}) \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} (\sum_{i} A^{-1}_{li} A_{i\beta}) = \sum_{\alpha,\beta,k,l} (\tilde{Y} A)_{\alpha k} \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} (A^{-1} A)_{l\beta}.$$

Hence,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}. \tag{2.6}$$

Also, we have

$$\sum_{\alpha,i,j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}) H_{ji} = \sum_{\alpha,i,j} \delta_{ij} H_{ji} = \sum_{\alpha,i} H_{ii} = \sum_{\alpha} \mathrm{tr}(H) = n\mathrm{tr}(H) = n\mathrm{tr}(AGA^{-1}).$$

Then

$$\sum_{\alpha,i,j}(\frac{\partial}{\partial\tilde{Y}_{\alpha i}}\tilde{Y}_{\alpha j})H_{ji} = n\mathrm{tr}(A^{-1}AG) = n\mathrm{tr}(G). \tag{2.7}$$

Therefore, by (2.6) and (2.7), we get

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = n\mathrm{tr}(G) + \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.$$

Further, we have

$$\mathrm{tr}(Y^\top(\nabla_Y G^\top)) = \sum_k (\mathrm{Y}^\top(\nabla_Y G^\top))_{kk} = \sum_{k,\alpha} Y^\top_{k\alpha}(\nabla_Y G^\top)_{\alpha k}$$

Then,

$$\mathrm{tr}(Y^\top(\nabla_Y G^\top)) = \sum_{k,\alpha} Y^\top_{k\alpha}\sum_\beta(\nabla_Y)_{\alpha\beta}G^\top_{\beta k} = \sum_{\alpha,\beta,k} Y^\top_{k\alpha}\frac{\partial G^\top_{\beta k}}{\partial Y_{\alpha\beta}} = \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}, \tag{2.8}$$

which completes the proof. □

In Propositions 2.1 and Propositions 2.2, we considered general vector g and general matrix G. Now, in the forthcoming proposition, we utilize these results for specific forms of g and G to unveil intriguing discoveries. These findings will play a pivotal role in proving Proposition 2.4. Additionally, it is worth noting that Parts (*ii*) and Parts (*iv*) of the following proposition were initially established in Lemma 1 and Lemma 2 of Chételat and Wells (2012).

Proposition 2.3. *Let Y be an $n \times p$ matrix and $S = Y^\top Y$. Let X be a p vector, $F = X^\top S^+ X$, and r be a differentiable function. Let $\tilde{Y} = YA^{-1}$, $G(X,S) = \frac{r^2(F)}{F^2}S^+XX^\top S^+S$, $g(X,S) = \frac{r(F)SS^+X}{F}$, and $H = AGA^{-1}$. Then, under the conditions of Theorem 2.2 we have*

$$(i) \quad \mathrm{tr}(G) = \frac{r^2(F)}{F},$$

$$(ii) \quad \mathrm{tr}\left(Y^\top \nabla_Y G^\top\right) = -4r(F)r'(F) + \frac{r^2(F)}{F}\left(p - 2\mathrm{tr}(SS^+) + 3\right),$$

$$(iii) \quad \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{4r(F)r'(F)}{F^2}$$

$$(iv) \quad \nabla_X . g(X,S) = 2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2),$$

$$(v) \quad g^\top(X,S)\Sigma^{-1}g(X,S) = \mathrm{tr}\left(\Sigma^{-1}SG\right),$$

$$(vi) \quad \mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right],$$

provided that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty$.

Proof.

$$(i) \quad \mathrm{tr}(G) = \mathrm{tr}\left(\frac{r^2(F)}{F^2}S^+XX^\top S^+S\right) = \frac{r^2(F)}{F^2}\mathrm{tr}(S^+XX^\top S^+S).$$

Then,

$$\mathrm{tr}(G) = \frac{r^2(F)}{F^2}\mathrm{tr}(X^\top S^+SS^+X) = \frac{r^2(F)}{F^2}\mathrm{tr}(X^\top S^+X) = \frac{r^2(F)}{F^2}F = \frac{r^2(F)}{F}.$$

(ii) From (2.8), we have

$$\mathrm{tr}\left(Y^\top(\nabla_Y G^\top)\right) = \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.$$

Therefore, by Part (*iv*) of Lemma 2.2, we get

$$\mathrm{tr}\left(Y^\top(\nabla_Y G^\top)\right) = -4r(F)r'(F) + \frac{r^2(F)}{F}\left(p - 2\mathrm{tr}(SS^+) + 3\right).$$

(*iii*) By Part (*i*) and Part (*ii*) together with Part (*iv*) of Proposition 2.2, we get

$$\begin{aligned}\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) &= n\mathrm{tr}(G) + \mathrm{tr}(Y^\top \nabla_Y G^\top) \\ &= \frac{nr^2(F)}{F} - 4r(F)r'(F) + \frac{r^2(F)}{F}\left(p - 2\mathrm{tr}(SS^+) + 3\right).\end{aligned}$$

Therefore,

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{4r(F)r'(F)}{F^2}.$$

(*iv*) By Lemma 2.3, we have

$$\nabla_X . g(X,S) = \sum_i \frac{\partial g_i}{\partial X_i} = 2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2).$$

(*v*) $g^\top(X,S)\Sigma^{-1}g(X,S) = \mathrm{tr}\Big(g^\top(X,S)\Sigma^{-1}g(X,S)\Big) = \mathrm{tr}\Big(\Sigma^{-1}g(X,S)g^\top(X,S)\Big).$

Then,

$$g^\top(X,S)\Sigma^{-1}g(X,S) = \mathrm{tr}\Big(\Sigma^{-1}\frac{r^2(F)}{F^2}SS^+XX^\top SS^+\Big) = \mathrm{tr}\Big(\Sigma^{-1}SG\Big).$$

(*vi*) From Part (*ii*) to (*v*), we have

$$\mathrm{E}\left[g^\top(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}\left[\mathrm{tr}(\Sigma^{-1}SG)\right] = \mathrm{E}\left[\mathrm{tr}(\tilde{S}H)\right] = \mathrm{E}\left[\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)\right].$$

Therefore,

$$\mathrm{E}\left[g^{\top}(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right],$$

which completes the proof. □

2.2 Main results

In this section, we introduce the primary result of this chapter, as stated in Theorem 2.2. Additionally, in Example 1, we demonstrate the improper application of the Cauchy-Schwarz inequality in the proof of Theorem 2 in Chételat and Wells (2012). Furthermore, in Example 2, we illustrate that the main result of this chapter, presented in Theorem 2.2, cannot be derived without making an assumption regarding the rank of the random matrix S.

In the following example, we show that the bound obtained in Theorem 2 of Chételat and Wells (2012) is not correct. To this end, we use the same notations as used in Chételat and Wells (2012). Let T be a symmetric matrix and A a positive definite matrix. Specifically, for a given X a column vector, in contrast with the statement in Chételat and Wells (2012)

$$X^T(T^+TA)^+(T^+TA)X \nleq X^T(T^+TA)^+(AT^+T)^+XX^T(AT^+T)(T^+TA)X.$$

Example 1. *Let* $A = I_4$, $X = \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}^T$ *and* $T = \frac{1}{48}\begin{bmatrix}7 & 7 & 1 & 1\\7 & 7 & 1 & 1\\1 & 1 & 7 & 7\\1 & 1 & 7 & 7\end{bmatrix}$. *Therefore*

$T^{+} = \frac{1}{4}\begin{bmatrix} 7 & 7 & -1 & -1 \\ 7 & 7 & -1 & -1 \\ -1 & -1 & 7 & 7 \\ -1 & -1 & 7 & 7 \end{bmatrix}$. *Then,*

$$T^{+}T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

This gives,

$$T^{+}TA = AT^{+}T = (T^{+}TA)^{+} = (AT^{+}T)^{+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Thus,

$$(T^{+}TA)^{+}(AT^{+}T)^{+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

$$(T^{+}TA)^{+}(T^{+}TA) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

$$(AT^+T)(T^+TA) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Hence,

$$X^T(T^+TA)^+(T^+TA)X = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2},$$

$$X^T(T^+TA)^+(AT^+T)^+XX^T(AT^+T)(T^+TA)X$$
$$= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{4}.$$

Hence,

$$X^T(T^+TA)^+(T^+TA)X \nleq X^T(T^+TA)^+(AT^+T)^+XX^T(AT^+T)(T^+TA)X.$$

In the following lemma, we investigate the relationship between the existence of $\mathrm{E}\left[\frac{1}{F}\right]$ and the rank of the random matrix S. In Example 2, we show that when $\mathrm{P}(R \leq 2) > 0$, $\mathrm{E}\left[\frac{1}{F}\right]$ may not exist. This observation leads us to conduct a more in-depth analysis. We aim to establish that $\mathrm{P}(R > 2) = 1$ is both necessary and sufficient for the existence of $\mathrm{E}\left[\frac{1}{F}\right]$. This lemma holds significant importance in deriving the results presented in Theorem 2.1. In Theorem 2.1, we demonstrate that the existence of $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$ can be determined by the existence of $\mathrm{E}\left[\frac{1}{F}\right]$. To be more precise, $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$ is upper-bounded by terms that involve $\mathrm{E}\left[\frac{1}{F}\right]$.

Lemma 2.4. *Let* $X \sim \mathcal{N}_p(\theta, \Sigma)$ *and* $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$. *Let* $F = X^\top S^+ X$ *where* $S = Y^\top Y$ *and* $R = \mathrm{rank}(S)$. *Then,*

$$\mathrm{E}\left[\frac{1}{F}\right] < \infty \textit{ if and only if } \mathrm{P}(R > 2) = 1.$$

Proof. Assume that $\mathrm{P}(R > 2) = 1$. Further, we have

$$X^\top S^+ X = X^\top A^{-1} A S^+ A A^{-1} X = (A^{-1}X)^\top A S^+ A (A^{-1}X) = U^\top A S^+ A U \tag{2.9}$$

where $U = A^{-1}X$.

Since $X \sim \mathcal{N}_p(\theta, \Sigma)$, we have $U = A^{-1}X \sim \mathcal{N}_p(A^{-1}\theta, I_p)$. Let C be $R \times p$-matrix of the form $C = [I_R \vdots 0_{R\times(p-R)}]$ and let $U_{(1)} = CU$. We have

$$U_{(1)}\Big|R \sim \mathcal{N}_R(CA^{-1}\theta, I_R)$$

Let λ^+_{min} and λ^+_{max} be the smallest and biggest nonzero eigenvalues of AS^+A respectively. Since AS^+A is semi-positive definite, we have

$$\lambda^+_{min} U_{(1)}^\top U_{(1)} \leq U^\top A S^+ A U \leq \lambda^+_{max} U_{(1)}^\top U_{(1)}. \tag{2.10}$$

Therefore, together with (2.9), we get

$$\frac{1}{F} \leq \frac{1}{\lambda^{+}_{min} U^{\top}_{(1)} U_{(1)}} = \frac{\lambda^{\dagger}_{max}}{U^{\top}_{(1)} U_{(1)}}, \tag{2.11}$$

where $\lambda^{\dagger}_{max}$ is the biggest nonzero eigenvalue of $(AS^{+}A)^{+} = A^{-1}SA^{-1}$. Note that $\lambda^{\dagger}_{max}$ depends on S and $U_{(1)}$ depends on R and X. Since, S and X are independent, we get

$$\mathrm{E}\left[\frac{\lambda^{\dagger}_{max}}{U^{\top}_{(1)} U_{(1)}}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{\lambda^{\dagger}_{max}}{U^{\top}_{(1)} U_{(1)}}\Big| R\right]\right] = \mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{max}\Big| R\right]\mathrm{E}\left[\frac{1}{U^{\top}_{(1)} U_{(1)}}\Big| R\right]\right].$$

Further, we have

$$\lambda^{\dagger}_{max} \leq \mathrm{tr}(A^{-1}SA^{-1}) = \mathrm{tr}(A^{-1}Y^{\top}YA^{-1}) = \mathrm{tr}((YA^{-1})^{\top}YA^{-1}) = \mathrm{vec}^{\top}(YA^{-1})\mathrm{vec}(YA^{-1})$$

where $\mathrm{vec}(YA^{-1}) \sim \mathcal{N}_{np}(0, I_p \otimes I_n)$. Therefore, we get

$$\mathrm{E}[\lambda^{\dagger}_{max}] \leq \mathrm{E}[\mathrm{vec}^{\top}(YA^{-1})\mathrm{vec}(YA^{-1})] \quad = \mathrm{tr}(I_p \otimes I_n) = \mathrm{tr}(I_{np}) = np.$$

Hence,

$$\mathrm{E}[\lambda^{\dagger}_{max}] \leq np. \tag{2.12}$$

Since $U_{(1)}\Big| R \sim \mathcal{N}_R(CA^{-1}\theta, I_R)$, we get

$$U^{\top}_{(1)} U_{(1)}\Big| R \sim \chi^2_R(\delta_R), \tag{2.13}$$

where $\delta_R = \left(CA^{-1}\theta\right)^{\top} CA^{-1}\theta$.

Let Z be a random variable such that $Z\Big| R \sim Poisson(\delta_R/2)$. By (2.13), We have

$$\mathrm{E}\left[\left(U^{\top}_{(1)} U_{(1)}\right)^{-1}\Big| R\right] = \mathrm{E}\left[\mathrm{E}\left[\left(U^{\top}_{(1)} U_{(1)}\right)^{-1}\Big| R, Z\right]\Big| R\right] = \mathrm{E}\left[\mathrm{E}\left[\left(\chi^2_{R+2Z}\right)^{-1}\Big| R, Z\right]\Big| R\right].$$

This gives

$$\mathrm{E}\left[\left(U_{(1)}^{\top}U_{(1)}\right)^{-1}\Big|R\right]=\mathrm{E}\left[\frac{2^{-1}\Gamma(\frac{R+2Z}{2}-1)}{\Gamma(\frac{R+2Z}{2})}\Big|R\right].$$

We have, $\frac{2^{-1}\Gamma(\frac{R+2Z}{2}-1)}{\Gamma(\frac{R+2Z}{2})}=\frac{1}{R+2Z-2}$. Therefore, since $\mathrm{P}(R>2)=\mathrm{P}(R\geq 3)=1$ and $\mathrm{P}\left(Z\geq 0\right)=1$, we have, $R+2Z-2\geq 1$ with probability one and then,

$$\mathrm{E}\left[\left(U_{(1)}^{\top}U_{(1)}\right)^{-1}\Big|R\right]=\mathrm{E}\left[\frac{1}{q+2Z-2}\Big|R\right]\leq 1\quad\text{almost surely.}$$

Therefore, together with (2.11) and (2.12), we get

$$\mathrm{E}\left[\frac{1}{F}\right]\leq\mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{max}\Big|R\right]\mathrm{E}\left[\frac{1}{U_{(1)}^{\top}U_{(1)}}\Big|R\right]\right]\leq\mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{max}\Big|R\right]\right]=\mathrm{E}\left[\lambda^{\dagger}_{max}\right]\leq np.$$

Hence,

$$\mathrm{E}\left[\frac{1}{F}\right]\leq np<\infty.$$

Now, assume that $\mathrm{E}\left[\frac{1}{F}\right]<\infty$. Further, from (2.10), we have

$$\frac{1}{\lambda^{+}_{max}U_{(1)}^{\top}U_{(1)}}=\frac{\lambda^{\dagger}_{min}}{U_{(1)}^{\top}U_{(1)}}\leq\frac{1}{F},$$

where $\lambda^{\dagger}_{min}$ is the smallest nonzero eigenvalue of $A^{-1}SA^{-1}$. Again, note that $\lambda^{\dagger}_{min}$ depends on S and $U_{(1)}$ depends on R and X. Since, S and X are independent, we get

$$\begin{aligned}\mathrm{E}\left[\frac{\lambda^{\dagger}_{min}}{U_{(1)}^{\top}U_{(1)}}\right]&=\mathrm{E}\left[\mathrm{E}\left[\frac{\lambda^{\dagger}_{min}}{U_{(1)}^{\top}U_{(1)}}\Big|R\right]\right]\\&=\mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{min}\Big|R\right]\mathrm{E}\left[\frac{1}{U_{(1)}^{\top}U_{(1)}}\Big|R\right]\right]\leq\mathrm{E}\left[\frac{1}{F}\right]<\infty.\end{aligned}$$

Then,

$$\mathrm{P}\left(\mathrm{E}\left[\lambda_{min}^{\dagger}\middle|R\right]\mathrm{E}\left[\frac{1}{U_{(1)}^{\top}U_{(1)}}\middle|R\right]<\infty\right)=1. \tag{2.14}$$

Since $0<\mathrm{E}\left[\lambda_{min}^{\dagger}\middle|R\right]<\infty$, we get

$$\mathrm{P}\left(\mathrm{E}\left[\frac{1}{U_{(1)}^{\top}U_{(1)}}\middle|R\right]<\infty\right)=1.$$

Therefore, by (2.13), we get

$$\mathrm{P}\left(\mathrm{E}\left[(\chi_R^2(\delta_R))^{-1}\middle|R\right]<\infty\right)=1.$$

Hence,

$$\mathrm{P}\left(R>2\right)=1,$$

which completes the proof. □

Part (i) of Proposition 2.1, and Part (vi) Proposition 2.3, are valid under the assumption that $\mathrm{E}\left[|\nabla_X.g(X,S)|\right]$ and $\mathrm{E}\left[\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right|\right]$, respectively, exist. This motivates us to explore the conditions under which these expectations are well-defined.

In the subsequent theorem, we establish that the condition $\mathrm{P}(R>2)=1$ ensures the existence of these expectations.

Theorem 2.1. *Let $X\sim N_p(\theta,\Sigma)$,$Y\sim N_{n\times p}(0,I_n\otimes\Sigma)$ and for A the symmetric positive definite square root of Σ, let $\tilde{Y}=YA^{-1}$. Let r be any bounded differentiable non-negative function $r:\mathbb{R}\longrightarrow[0,C_1]$ with bounded derivative $|r'|\leq C_2$. Define $G=\frac{r^2(F)}{F^2}S^+XX^\top S^+S$, and $g(X,S)=\frac{r(F)SS^+X}{F}$, where $F=X^\top S^+X$ and $H=AGA^{-1}$. Let $R=\mathrm{rank}(S)$ and* *suppose that* $\mathrm{P}(R > 2) = 1$. *Then*

(i) $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty$,

(ii) $\mathrm{E}\left[|\nabla_X . g(X, S)|\right] < \infty$.

Proof. (i) By Proposition 2.3 and triangle inequality we get

$$\begin{aligned}\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| &= \left|\frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - 4r(F)r'(F)\right| \\ &\leq \frac{r^2(F)}{F}\left|n + p - 2\mathrm{tr}(SS^+) + 3\right| + 4r(F)r'(F).\end{aligned}$$

This gives

$$\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| \leq \frac{C_1^2}{F}\left|n + p - 2\mathrm{tr}(SS^+) + 3\right| + 4C_1C_2.$$

Therefore, since $\mathrm{tr}(SS^+) = \min(n, p)$ almost surely, we have

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] \leq C_1^2\left|n + p - 2\min(n, p) - 1\right|\mathrm{E}\left[\frac{1}{F}\right] + 4C_1C_2. \tag{2.15}$$

Further, since $\mathrm{P}(R > 2) = 1$ then by Lemma 2.4, we get $\mathrm{E}\left[\frac{1}{F}\right] < \infty$. Then,

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty.$$

(ii) Similarly to Part (i), by Part (iv) of Proposition 2.2, we get

$$\mathrm{E}\left[|\nabla_X \cdot g(X, S)|\right] = \mathrm{E}\left[\left|2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2)\right|\right] \leq 2C_2 + C_1|\min(n, p) - 2|\mathrm{E}\left[\frac{1}{F}\right] < \infty,$$

which completes the proof. □

In the previous theorem, we demonstrated that $\mathrm{P}(R>2)=1$ is a sufficient condition for the existence of $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$. Now, in the following corollary, we explore the conditions under which $\mathrm{P}(R>2)=1$ is both necessary and sufficient for the existence of $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$.

Corollary 2.1. *Let $X\sim\mathcal{N}_p(\theta,\Sigma)$,$Y\sim\mathcal{N}_{n\times p}(0,I_n\otimes\Sigma)$ and for A the symmetric positive definite square root of Σ, let $\tilde{Y}=YA^{-1}$. Let r be any bounded differentiable positive function $r:\mathbb{R}\longrightarrow[C^*,C_1]$ with bounded derivative $|r'|\leq C_2$. Suppose that $|p-n|>1$. Define $G=\frac{r^2(F)}{F^2}S^+XX^\top S^+S$, where $F=X^\top S^+X$ and $H=AGA^{-1}$. Then, $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]<\infty$, if and only if $P(R>2)=1$.*

Proof. If $\mathrm{P}(R>2)=1$ then, by Theorem 2.1, we get

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]<\infty.$$

Now, assume that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]<\infty$. We have

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)=\frac{r^2(F)}{F}\left(n+p-2\min(p,n)+3\right)-4r(F)r'(F).$$

Therefore since $n+p-2\min(p,n)=|p-n|$, we get

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\geq\frac{r^2(F)}{F}\left(|p-n|-1\right)-4C_1C_2.$$

Therefore since $|p-n|>1$, we get

$$\begin{aligned}\frac{r^2(F)}{F}&\leq\frac{1}{|p-n|-1}\left(\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)+4C_1C_2\right)\\&=\frac{1}{|p-n|-1}\left(\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)+4C_1C_2\right|\right).\end{aligned}$$

Then,

$$\frac{r^2(F)}{F} \leq \frac{1}{|p-n|-1}\left(\left|\operatorname{div}_{\operatorname{vec}(\tilde{Y})}\operatorname{vec}(\tilde{Y}H)\right| + 4C_1C_2\right).$$

Therefore,

$$\mathrm{E}\left[\frac{r^2(F)}{F}\right] \leq \frac{1}{|p-n|-1}\left(\mathrm{E}\left[\left|\operatorname{div}_{\operatorname{vec}(\tilde{Y})}\operatorname{vec}(\tilde{Y}H)\right|\right] + 4C_1C_2\right) < \infty.$$

Further, we have

$$\frac{(C^*)^2}{F} \leq \frac{r^2(F)}{F}.$$

Therefore,

$$(C^*)^2\mathrm{E}\left[\frac{1}{F}\right] \leq \mathrm{E}\left[\frac{r^2(F)}{F}\right] < \infty,$$

which implies that

$$\mathrm{E}\left[\frac{1}{F}\right] < \infty.$$

Therefore, by Lemma 2.4, we get,

$$\mathrm{P}(R > 2) = 1,$$

which completes the proof. □

In the following example, we consider a positive function r such that $\mathrm{E}\left[|\operatorname{div}_{\operatorname{vec}(\tilde{Y})}\operatorname{vec}(\tilde{Y}H)|\right] = \infty$. This emphasizes the significance of the assumption regarding $R > 2$ with probability one, where $R = \operatorname{rank}(S)$. In particular, we demonstrate that when $\mathrm{P}(R \leq 2) > 0$, it is possible to have $\mathrm{E}\left[\frac{1}{F}\right] = \infty$, rendering obsolete the Theorem 2 of Chételat and Wells (2012).

Example 2. *Let* $X \sim \mathcal{N}_2\left(\begin{bmatrix}1\\1\end{bmatrix}, I_2\right)$ *and* $Y = \begin{bmatrix}U \vdots V\end{bmatrix}$ *where* U *and* V *are independent random variable distributed as* $\mathcal{N}(0,\,1)$ *i.e.* $Y \sim \mathcal{N}_{1\times 2}(0, 1\otimes I_2)$*. let* $r(x) = \frac{1}{1+e^{-x}}$*. Let* $S^+ = PD^+P^\top$ *be the spectral decomposition of* S^+ *where* $D^+ = diag(d_1, 0)$*. Since*

$$F = X^\top S^+ X = X^\top P D^+ P^\top X = (P^\top X)^\top D^+ P^\top X.$$

Therefore,

$$\frac{F}{d_1} = (P^\top X)^\top \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} P^\top X.$$

Note that d_1 *and* P *are functions of* (U, V) *and note that* $P^\top X \Big| U, V \sim \mathcal{N}_2\left(P^\top \begin{bmatrix}1\\1\end{bmatrix}, I_2\right)$*.*

Then,

$$\frac{X^\top S^+ X}{d_1}\Big| U, V \sim \chi_1^2(\delta_0)$$

where $\delta_0 = \begin{bmatrix}1 & 1\end{bmatrix} P \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} P^\top \begin{bmatrix}1\\1\end{bmatrix}$*. Therefore,*

$$\mathrm{E}\left[\frac{d_1}{F}\Big|U, V\right] = \mathrm{E}\left[\frac{d_1}{X^\top S^+ X}\Big|U, V\right] = \mathrm{E}\left[(\chi_1^2(\delta_0))^{-1}\Big|U, V\right] = \infty,$$

almost surely. Then,

$$\mathrm{E}\left[\frac{1}{F}\Big|U, V\right] = \frac{1}{d_1}\mathrm{E}\left[\frac{d_1}{F}\Big|U, V\right] = \infty,$$

with probability one. Hence,

$$\mathrm{E}\left[\frac{1}{F}\right] = \infty.$$

Further, we have

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}(n + p - 2\min(n,p) + 3) - 4r(F)r'(F),$$

where $\min(n,p) = 1$*. Therefore, we get*

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}(1 + 2 - 2 + 3) - 4r(F)r'(F) = \frac{4r^2(F)}{F} - 4r(F)r'(F).$$

Then,

$$\frac{4r^2(F)}{F} = \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) + 4r(F)r'(F) = \left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) + 4r(F)r'(F)\right|.$$

Hence,

$$\frac{4r^2(F)}{F} \leq \left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| + 4r(F)r'(F).$$

Since $r(F)$ *and* $r'(F)$ *are bounded by 1 we get*

$$\frac{4r^2(F)}{F} \leq \left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| + 4. \tag{2.16}$$

We also have,

$$\frac{4r^2(F)}{F} = \frac{4}{F(1 + e^{-F})^2}.$$

Since $1 < 1 + e^{-F} \leq 2$ *thus* $1 < (1 + e^{-F})^2 \leq 4$. *Therefore,*

$$\frac{1}{F} \leq \frac{4}{F(1+e^{-F})^2}.$$

Then, we get

$$\mathrm{E}\left[\frac{1}{F}\right] \leq \mathrm{E}\left[\frac{4}{F(1+e^{-F})^2}\right].$$

But, since $\mathrm{E}\left[\frac{1}{F}\right] = \infty$, *we get*

$$\mathrm{E}\left[\frac{4}{F(1+e^{-F})^2}\right] = \infty$$

Then from (2.16), *we get*

$$\mathrm{E}\left[\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right|\right] = \infty.$$

Finally, we are now ready to present and substantiate the primary proposition that plays a crucial role in establishing the main result of this chapter in Theorem 2.2.

Proposition 2.4. *Let* $X \sim \mathcal{N}_p(\theta, \Sigma)$, $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$ *and* $F = X^\top S^+ X$ *where* $S = Y^\top Y$. *Let* $g(X,S) = \frac{r(F)SS^+X}{F}$, *where* r *is a differentiable function. Let* $R = \mathrm{rank}(S)$ *and suppose* $\mathrm{P}(R > 2) = 1$, *then*

$$(i) \quad \mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}(X-\theta)\right] = \mathrm{E}_\theta\left[2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2)\right],$$

$$(ii) \quad \mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}_\theta\left[\frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{4r(F)r'(F)}{F^2}\right].$$

Proof. (i) By, Part (i) of Proposition 2.1, we have

$$\mathrm{E}_\theta[g^\top(X,S)\Sigma^{-1}(X-\theta)] = \mathrm{E}_\theta[\nabla_X.g(X,S)],$$

and from, Part (iv) of Proposition 2.3, we have

$$\nabla_X.g(X,S) = 2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2).$$

This gives

$$\mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}(X-\theta)\right] = \mathrm{E}_\theta\left[2r'(F) + \frac{r(F)}{F}(\mathrm{tr}(SS^+) - 2)\right].$$

(ii) Since $\mathrm{P}(R > 2) = 1$, by Theorem 2.1, we get $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty$. Therefore from Part (vi) of Proposition 2.3, we have

$$\mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}_\theta\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right].$$

Further from Part (iii) of Proposition 2.3 we have

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{4r(F)r'(F)}{F^2}.$$

Hence

$$\mathrm{E}_\theta\left[g^\top(X,S)\Sigma^{-1}g(X,S)\right] = \mathrm{E}_\theta\left[\frac{r^2(F)}{F}\left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{4r(F)r'(F)}{F^2}\right],$$

which completes the proof. □

The upcoming theorem serves as the central finding in this chapter. Utilizing Proposition 2.4, we are ready to provide a high-dimensional Baranchik (1970) type estimator, for the mean vector of a p-dimensional multivariate normal distribution. This result was initially introduced by Chételat and Wells (2012) in their Theorem 1. However, in Example 2, we demonstrated that this result might not hold without an additional assumption on the rank of the random matrix S. Hence, it is essential to integrate this assumption into the statement of the theorem.

Theorem 2.2. *Let* $X \sim \mathcal{N}_p(\theta, \Sigma)$,$Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$ *and* $S = Y^\top Y$. *Let* $F = X^\top S^+ X$, $\delta_r(X, S) = \left(I - \frac{r(F)SS^+}{F}\right) X$, *where* r *is a differentiable function, and* $\delta^0(X) = X$. *Suppose that* $\mathrm{P}(R > 2) = 1$, *where* $R = \mathrm{rank}(S)$. *Suppose that*

(i) r *satisfies* $0 \le r \le \dfrac{2(\min(n,p) - 2)}{n + p - 2\min(n,p) + 3}$

(ii) r *is non-decreasing*

(iii) r' *is bounded*

Then, under invariant quadratic loss, $L(\theta, \delta) = (\delta - \theta)^\top \Sigma^{-1} (\delta - \theta)$, δ_r *dominates* δ^0.

Proof. Let $g(X, S) = \frac{r(F)SS^+X}{F}$. Thus $\delta_r = X - g(X, S)$. The risk difference under the quadratic loss between δ_r and δ^0 is

$$\begin{aligned}\Delta_\theta =& \mathrm{E}_\theta \left[\left(X - g(X,S) - \theta)^\top\right) \Sigma^{-1} \left(X - g(X,S) - \theta\right) \right] \\ & - \mathrm{E}_\theta \left[(X - \theta)^\top \Sigma^{-1} (X - \theta) \right] \\ & = -2\mathrm{E}_\theta \left[g^\top(X,S)\Sigma^{-1}(X - \theta) \right] + \mathrm{E}_\theta \left[g^\top(X,S)\Sigma^{-1} g(X,S) \right]. \end{aligned} \tag{2.17}$$

From Proposition 2.4, we have

$$\Delta_\theta = \mathrm{E}_\theta \left[\frac{r^2(F)}{F} \left(n + p - 2\mathrm{tr}(SS^+) + 3\right) - \frac{2r(F)}{F} (\mathrm{tr}(SS^+) - 2) - 4r'(F)(1 + \frac{r(F)}{F^2}) \right].$$

Since r is non-negative and non-decreasing, therefore $-4r'(F)(1 + \frac{r(F)}{F^2}) \le 0$. Under the condition (i) on r, we have

$$r(F) \leq \frac{2(\min(n,p) - 2)}{n + p - 2\min(n,p) + 3}.$$

Then,

$$\frac{r^2(F)}{F}(n + p - 2\min(n,p) + 3) \leq \frac{2r(F)}{F}(\min(n,p) - 2).$$

Therefore, since $\mathrm{tr}(SS^+) = \min(n,p)$ almost surely, we get

$$\mathrm{E}\left[\frac{r^2(F)}{F}(n + p - 2\mathrm{tr}(SS^+) + 3) - \frac{2r(F)}{F}(\mathrm{tr}(SS^+) - 2)\right] \leq 0.$$

Therefore,

$\Delta_\theta \leq 0$, which completes the proof. □

Chapter 3

The Case of Matrix Normal Mean Estimation

In this chapter, we suppose that $Z_1, \ldots, Z_N$ are independent and identically distributed random samples from $\mathcal{N}_{p\times q}(\theta, \Psi \otimes I_q)$ where Ψ represents the row covariance matrix and is an unknown matrix. Then, $Z = [Z_1, \ldots, Z_N]^\top$ follows $\mathcal{N}_{Nq\times p}(\gamma\theta^\top, I_{Nq} \otimes \Psi)$ where $\gamma = e \otimes I_q$ and $e = [1, \ldots, 1]^\top$ is an N-dimensional vector. Let $X = \bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$. Therefore, $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$ where $\Sigma = \frac{\Psi}{N}$. Let us consider $S = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \bar{Z})(Z_i - \bar{Z})^\top$ as an estimator of Σ and $n = N-1$. In Appendix A.1, We show that S can be written as $S = Y^\top Y$, where Y is independent of X and follows a matrix normal distribution $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$. This implies that $S \sim Wishart_p(nq, \Sigma)$.

This chapter is divided into two main sections. In Section 3.1, we introduce crucial Lemmas and propositions that play pivotal roles in proving the results outlined in Section 3.2. Section 3.2 focuses on essential results that form the basis for the main result of this chapter, as presented in Theorem 3.3.

In Theorem 3.3, we extend the findings of Theorem 2.2, as discussed in Chapter 2. The outcomes detailed in this chapter also serve as generalizations of the key results established in Chételat and Wells (2012). In particular, in Theorem 3.3, we establish that the Baranchik (1970) type estimator

$$\delta_r(X, S) = \left(I - \frac{r(\operatorname{tr}(X^\top S^+ X))}{\operatorname{tr}(X^\top S^+ X)} SS^+ \right) X,$$

outperforms the usual estimator $\delta^0 = X$ under the invariant quadratic loss,

$$L(\theta, \delta) = \operatorname{tr}\left((\delta - \theta)^\top \Sigma^{-1} (\delta - \theta) \right),$$

when $\mathrm{P}(qR > 2) = 1$, where $R = \operatorname{rank}(S)$. Once again, it is worth noting that the function r in the above estimator represents a positive, bounded, and differentiable real-valued function.

3.1 Important Preliminary Results

In Section 3.1, we introduce several technical lemmas and propositions that play a pivotal role in the development of results presented in Section 3.2. For the sake of maintaining the simplicity and clarity of this thesis, most proofs have been relocated to the Appendix A.

Lemma 3.1. *Let Y be an $nq \times p$ matrix and $S = Y^\top Y$. Let X be a $p \times q$ matrix and $F = \operatorname{tr}(X^\top S^+ X)$. Let $A \in M_{k\times p}$ and $B \in M_{p\times h}$, it then follows that*

(i) $$\left(\frac{\partial S}{\partial Y_{\alpha\beta}} \right)_{kl} = \delta_{\beta k} Y_{\alpha l} + \delta_{\beta l} Y_{\alpha k},$$

(ii) $$\left(A \frac{\partial S}{\partial Y_{\alpha\beta}} B \right)_{kl} = A_{k\beta} (YB)_{\alpha l} + (AY^\top)_{k\alpha} B_{\beta l},$$

(iii) $$\left(\frac{\partial X^\top S^+ X}{\partial Y_{\alpha\beta}} \right)_{kk} = -\, 2(X^\top S^+ Y^\top)_{k\alpha} (S^+ X)_{\beta k}$$

$$+ 2(X^\top S^+ S^+ Y^\top)_{k\alpha}((I - SS^+)X)_{\beta k},$$

$$(iv) \quad \frac{\partial F}{\partial Y_{\alpha\beta}} = -2(S^+ X X^\top S^+ Y^\top)_{\beta\alpha} + 2((I - SS^+)XX^\top S^+ S^+ Y^\top)_{\beta\alpha},$$

$$(v) \quad \left(\frac{\partial S^+ X X^T S S^+}{\partial Y_{\alpha\beta}}\right)_{kl} = -S^+_{k\beta}(Y S^+ X X^T S S^+)_{\alpha l} - (S^+ Y^T)_{k\alpha}(S^+ X X^T S S^+)_{\beta l}$$
$$+ (I - SS^+)_{k\beta}(Y S^+ S X X^T S S^+)_{\alpha l} + (S^+ S^+ Y^T)_{k\alpha}((I - SS^+)X X^T S S^+)_{\beta l}$$
$$+ (S^+ X X^T)_{k\beta}(Y S^+)_{\alpha l} + (S^+ X X^T Y^T)_{k\alpha} S^+_{\beta l} - (S^+ X X^T S S^+)_{k\beta}(Y S^+)_{\alpha l}$$
$$- (S^+ X X^T S S^+ Y^T)_{k\alpha} S^+_{\beta l} + (S^+ X X^T S^+ Y^T)_{k\alpha}(I - SS^+)_{\beta l}.$$

Proof. The proof of this result is given in the Appendix A.4. □

Lemma 3.2. *Let Y be an $nq \times p$ matrix and $S = Y^\top Y$. Let X be a $p \times q$ matrix, $F = \mathrm{tr}(X^\top S^+ X)$, and $G(X,S) = \frac{r^2(F)}{F^2}(S^+ X X^\top S S^+)$, where r is a differentiable function. Then*

$$(i) \quad \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} = \frac{2r(F)r'(F)}{F^2}\left(\frac{\partial F}{\partial y_{\alpha\beta}}\right)(S^+ X X^\top S S^+)_{kl} - \frac{2r^2(F)}{F^3}\left(\frac{\partial F}{\partial y_{\alpha\beta}}\right)(S^+ X X^\top S S^+)_{kl}$$
$$+ \frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+ X X^\top S S^+)_{kl},$$

$$(ii) \quad \sum_{\alpha,k,\beta} Y_{\alpha k}\left(\frac{\partial F}{\partial Y_{\alpha\beta}}\right)(SS^+ X X^\top S^+)_{\beta k} = -2\mathrm{tr}((X^\top S^+ X)^2),$$

$$(iii) \quad \sum_{\alpha,k,\beta} Y_{\alpha k}\frac{\partial}{\partial Y_{\alpha\beta}}(SS^+ X X^\top S^+)_{\beta k} = F(p - 2\mathrm{tr}(SS^+) - 1),$$

$$(iv) \quad \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = -\frac{4r(F)r'(F)}{F^2} \mathrm{tr}\left((X^\top S^+ X)^2\right) + \frac{r^2(F)}{F}\left(\frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2} + p - 2\mathrm{tr}(SS^+) - 1\right).$$

Proof. The proof of this result is given in the Appendix A.4. □

Lemma 3.3. *Let Y be an $nq \times p$ matrix and $S = Y^\top Y$. Let X be a $p \times q$ matrix and $F = \mathrm{tr}(X^\top S^+ X)$, and $g(X,S) = \frac{r(F)}{F}(SS^+X)$, where r is a differentiable function. Then*

$$(i) \quad \frac{\partial F}{\partial X_{ij}} = 2(S^+X)_{ij},$$

$$(ii) \quad \left(\frac{\partial SS^+X}{\partial X_{ij}}\right)_{kl} = (SS^+)_{ki}\delta_{lj},$$

$$(iii) \quad \frac{\partial g_{kl}}{\partial X_{ij}} = \frac{2(Fr'(F) - r(F))}{F^2}(S^+X)_{ij}(SS^+X)_{kl} + \frac{r(F)}{F}(SS^+)_{ki}\delta_{lj},$$

$$(iv) \quad \sum_{i,j} \frac{\partial g_{ij}}{\partial X_{ij}} = 2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2).$$

Proof. The proof of this result is given in the Appendix A.4. □

The forthcoming proposition can be seen as an expansion of proposition 2.1. The proof for this Proposition can be established by applying Corollary A.2 once again. In this chapter, we investigate the existence of the expectation on the right-hand side of part (i) in Theorem 3.1 for a specific form of $g(X,S)$.

Proposition 3.1. *Let $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$. Let $g(X,S)$ be a differentiable $p \times q$ matrix function. Then*

$$(i) \quad \mathrm{E}_\theta\Big[\mathrm{tr}(g^\top(X,S)\Sigma^{-1}(X-\theta))\Big] = \mathrm{E}_\theta\Big[\mathrm{tr}(\nabla_X g^\top(X,S))\Big],$$

provided that $\mathrm{E}_\theta\Big[|\mathrm{tr}(\nabla_X g^\top(X,S))|\Big] < \infty,$

$$(ii) \quad \mathrm{tr}\Big(\nabla_X g^\top(X,S)\Big) = \sum_{i,j} \frac{\partial g_{ij}}{\partial X_{ij}}.$$

Proof. (i) Let $\tilde{X} = A^{-1}(X-\theta)$ where A is a symmetric positive definite square root of Σ. Thus $\tilde{X} \sim \mathcal{N}_{p\times q}(0, I_p \otimes I_q)$. Therefore, $X_{ij} \sim \mathcal{N}(0,1)$. Let $h = A^{-1}g(X,S)$. Then, we have

$$\mathrm{tr}(g^\top(X,S)\Sigma^{-1}(X-\theta)) = \mathrm{tr}(g^\top(X,S)A^{-1}A^{-1}(X-\theta)) = \mathrm{tr}(h^\top\tilde{X}) = \sum_i (h^\top\tilde{X})_{ii}.$$

Then,

$$\mathrm{tr}(g^\top(X,S)\Sigma^{-1}(X-\theta)) = \sum_{i,j} h^\top_{ij}\tilde{X}_{ji}. \tag{3.1}$$

Therefore, by (3.1), we have

$$\mathrm{E}\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}(X-\theta)\right)\right] = \mathrm{E}\left[\sum_{i,j} h^\top_{ij}\tilde{X}_{ji}\right] = \sum_{i,j}\mathrm{E}\left[h^\top_{ij}\tilde{X}_{ji}\right] = \sum_{i,j}\mathrm{E}\left[\tilde{X}_{ji}h^\top_{ij}\right].$$

Therefore, by Corollary A.2, we get

$$\sum_{i,j}\mathrm{E}[\tilde{X}_{ji}h^\top_{ij}] = \sum_{i,j}\mathrm{E}\left[\frac{\partial}{\partial\tilde{X}_{ji}}h^\top_{ij}\right] = \sum_{i,j}\mathrm{E}\left[\frac{\partial}{\partial\tilde{X}_{ji}}h_{ji}\right] = \mathrm{E}\left[\sum_{i,j}\frac{\partial}{\partial\tilde{X}_{ji}}h_{ji}\right].$$

Then,

$$\sum_{i,j}\mathrm{E}[\tilde{X}_{ji}h^\top_{ij}] = \mathrm{E}\left[\sum_{i,j}\frac{\partial}{\partial\tilde{X}_{ji}}(A^{-1}g(X,S))_{ji}\right] = \mathrm{E}\left[\sum_{i,j}\frac{\partial}{\partial\tilde{X}_{ji}}\sum_k A^{-1}_{jk}g(X,S)_{ki}\right].$$

Hence,

$$\sum_{i,j} \mathrm{E}[\tilde{X}_{ji} h_{ij}^{\top}] = \mathrm{E}\bigg[\sum_{i,j,k} A_{jk}^{-1} \frac{\partial}{\partial \tilde{X}_{ji}} g(X,S)_{ki}\bigg]. \tag{3.2}$$

Now, by applying the chain rule in (3.2), we have

$$\begin{aligned} \mathrm{E}\bigg[\sum_{i,j,k} A_{jk}^{-1} \frac{\partial}{\partial \tilde{X}_{ji}} g(X,S)_{ki}\bigg] &= \mathrm{E}\bigg[\sum_{i,j,k} A_{jk}^{-1} \sum_{l,\alpha} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\bigg] \\ &= \mathrm{E}\bigg[\sum_{i,j,k,l,\alpha} A_{jk}^{-1} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\bigg] \end{aligned} \tag{3.3}$$

Since $\tilde{X} = A^{-1}(X - \theta)$, we have

$$X_{l\alpha} = \sum_t A_{lt} \tilde{X}_{t\alpha} + \theta_{l\alpha},$$

then,

$$\frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}} = \sum_t A_{lt} \frac{\partial \tilde{X}_{t\alpha}}{\partial \tilde{X}_{ji}} = \sum_t A_{lt} \delta_{tj} \delta_{\alpha i} = A_{lj} \delta_{\alpha i}. \tag{3.4}$$

Therefore, by replacing (3.4) in (3.3), we get

$$\mathrm{E}\bigg[\sum_{i,j,k,l,\alpha} A_{jk}^{-1} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\bigg] = \mathrm{E}\bigg[\sum_{i,j,k,l,\alpha} A_{jk}^{-1} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} A_{lj} \delta_{\alpha i}\bigg],$$

and then,

$$\begin{aligned} \mathrm{E}\bigg[\sum_{i,j,k,l,\alpha} A_{jk}^{-1} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\bigg] &= \mathrm{E}\bigg[\sum_{i,k,l} \frac{\partial}{\partial X_{li}} g(X,S)_{ki} \sum_j A_{lj} A_{jk}^{-1}\bigg] \\ &= \mathrm{E}\bigg[\sum_{i,k,l} \frac{\partial}{\partial X_{li}} g(X,S)_{ki} (AA^{-1})_{lk}\bigg]. \end{aligned}$$

This gives

$$\mathrm{E}\left[\sum_{i,j,k,l,\alpha} A^{-1}_{jk} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\right] = \mathrm{E}\left[\sum_{i,k} \frac{\partial}{\partial X_{ki}} g(X,S)_{ki}\right] = \mathrm{E}\left[\sum_{i,k} (\nabla_X)_{ki} g^\top(X,S)_{ik}\right],$$

and then,

$$\mathrm{E}\left[\sum_{i,j,k,l,\alpha} A^{-1}_{jk} \frac{\partial}{\partial X_{l\alpha}} g(X,S)_{ki} \frac{\partial X_{l\alpha}}{\partial \tilde{X}_{ji}}\right] = \mathrm{E}\left[\sum_{k} (\nabla_X g^\top(X,S))_{kk}\right] = \mathrm{E}\Big[\mathrm{tr}(\nabla_X g^\top(X,S))\Big].$$

$$(ii) \quad \mathrm{tr}\Bigg(\nabla_X g^\top(X,S)\Bigg) = \sum_i (\nabla_X g^\top)_{ii} = \sum_{i,j} (\nabla_X)_{ij} g^\top_{ji} = \sum_{i,j} (\nabla_X)_{ij} g_{ij} = \sum_{i,j} \frac{\partial g_{ij}}{\partial X_{ij}},$$

which completes the proof. □

In the following proposition, we build upon the ideas from Lemma 3 in the work by Chételat and Wells (2012). Our approach refines and organizes their lemma, providing a detailed breakdown of each step in the proof that leads to the end result stated in Lemma 3 of Chételat and Wells (2012).

Proposition 3.2. *Let $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$ and $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$. Let $S = Y^\top Y$. For A symmetric positive definite square root of Σ (i.e. $A^2 = \Sigma$) define $\tilde{Y} = YA^{-1}$, $\tilde{S} = \tilde{Y}^\top \tilde{Y}$ and $H = AGA^{-1}$ where $G(X,S)$ is a differentiable $p \times p$ matrix function. Then,*

$$(i) \quad \mathrm{tr}(\Sigma^{-1} S G) = \mathrm{tr}(\tilde{S} H),$$

$$(ii) \quad \mathrm{tr}(\tilde{S} H) = \mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y} H),$$

(iii) $\mathrm{E}\left[\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right],$

provided that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{YH})|\right] < \infty.$

(iv)
$$\begin{aligned}\nabla_{\tilde{Y}}.(\tilde{Y}H) &= \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = nq\mathrm{tr}(G) + \mathrm{tr}(Y^\top(\nabla_Y G^\top)) \\ &= nq\mathrm{tr}(G) + \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.\end{aligned}$$

Proof.

(i) $\mathrm{tr}(\tilde{S}H) = \mathrm{tr}(A^{-1}SA^{-1}AGA^{-1}) = \mathrm{tr}(A^{-1}SGA^{-1}) = \mathrm{tr}(A^{-1}A^{-1}SG) = \mathrm{tr}(A^{-2}SG).$

Then,

$$\mathrm{tr}(\tilde{S}H) = \mathrm{tr}(\Sigma^{-1}SG).$$

(ii) $\mathrm{tr}(\tilde{S}H) = \sum_i (\tilde{S}H)_{ii} = \sum_{i,j} \tilde{S}_{ij}H_{ji} = \sum_{i,j} (\tilde{Y}^\top\tilde{Y})_{ij}H_{ji} = \sum_{i,j,k} \tilde{Y}_{ik}^\top\tilde{Y}_{kj}H_{ji}.$

Then,

$$\mathrm{tr}(\tilde{S}H) = \sum_{i,j,k} \tilde{Y}_{ki}\tilde{Y}_{kj}H_{ji} = \sum_{i,k} \tilde{Y}_{ki}\sum_j \tilde{Y}_{kj}H_{ji} = \sum_{i,k} \tilde{Y}_{ki}(\tilde{Y}H)_{ki} = \mathrm{vec}(\tilde{Y}) \cdot \mathrm{vec}(\tilde{Y}H).$$

(iii) Since $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq}\otimes\Sigma)$, we have $\tilde{Y} = YA^{-1} \sim \mathcal{N}_{nq\times p}(0, I_{nq}\otimes I_p)$.

Then,

$$\text{vec}(\tilde{Y}) \sim \mathcal{N}_{npq}(0, I_{npq}).$$

Therefore,

$$\tilde{Y}_{\alpha i} \sim \mathcal{N}(0, 1).$$

Also, we have

$$\text{vec}(\tilde{Y}).\text{vec}(\tilde{Y}H) = \sum_{\alpha,i} \tilde{Y}_{\alpha i}(\tilde{Y}H)_{\alpha i} = \sum_{\alpha,i} \tilde{Y}_{\alpha i} \sum_{j} \tilde{Y}_{\alpha j} H_{ji} = \sum_{\alpha,i,j} \tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j} H_{ji}.$$

Therefore, we have

$$\text{E}\Bigg[\text{vec}(\tilde{Y}).\text{vec}(\tilde{Y}H)\Bigg] = \text{E}\Bigg[\sum_{\alpha,i,j} \tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j} H_{ji})\Bigg] = \sum_{\alpha,i,j} \text{E}\Bigg[\tilde{Y}_{\alpha i}\tilde{Y}_{\alpha j} H_{ji}\Bigg] = \sum_{\alpha,i,j} \text{E}\Bigg[\tilde{Y}_{\alpha i} g_j(\tilde{Y}_{\alpha i})\Bigg],$$

where $g_j(\tilde{Y}_{\alpha i}) = \tilde{Y}_{\alpha j} H_{ji}$.Therefore, by Corollary A.2, we get

$$\sum_{\alpha,i,j} \text{E}\Bigg[\tilde{Y}_{\alpha i} g_j(\tilde{Y}_{\alpha i})\Bigg] = \sum_{\alpha,i,j} \text{E}\Bigg[\frac{\partial}{\partial \tilde{Y}_{\alpha i}} g_j(\tilde{Y}_{\alpha i})\Bigg] = \text{E}\Bigg[\sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} g_j(\tilde{Y}_{\alpha i})\Bigg] = \text{E}\Bigg[\sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j} H_{ji}\Bigg].$$

Then,

$$\sum_{\alpha,i,j} \text{E}\Bigg[\tilde{Y}_{\alpha i} g_j(\tilde{Y}_{\alpha i})\Bigg] = \text{E}\Bigg[\sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{j} \tilde{Y}_{\alpha j} H_{ji}\Bigg] = \text{E}\Bigg[\sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} (\tilde{Y}H)_{\alpha i}\Bigg] = \text{E}\Bigg[\nabla_{\tilde{Y}} \cdot (\tilde{Y}H)\Bigg].$$

Hence,

$$\sum_{\alpha,i,j} \text{E}\Bigg[\tilde{Y}_{\alpha i} g_j(\tilde{Y}_{\alpha i})\Bigg] = \text{E}\Bigg[\text{div}_{\text{vec}(\tilde{Y})} \text{vec}(\tilde{Y}H)\Bigg].$$

$$(iv) \quad \nabla_{\tilde{Y}} \cdot (\tilde{Y}H) = \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}}(\tilde{Y}H)_{\alpha i} = \sum_{\alpha,i} \frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{j} \tilde{Y}_{\alpha j} H_{ji}.$$

Then,

$$\nabla_{\tilde{Y}}.(\tilde{Y}H) = \sum_{\alpha,i,j} \frac{\partial}{\partial \tilde{Y}_{\alpha i}}(\tilde{Y}_{\alpha j} H_{ji}) = \sum_{\alpha,i,j} \bigg((\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}) H_{ji} + \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) \bigg).$$

Hence,

$$\nabla_{\tilde{Y}}.(\tilde{Y}H) = \sum_{\alpha,i,j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}) H_{ji} + \sum_{\alpha,i,j} \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}). \tag{3.5}$$

By applying the chain rule in the second term of (3.5), we get

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j} \tilde{Y}_{\alpha j} \sum_{k,\beta} \frac{\partial}{\partial Y_{k\beta}} H_{ji} \frac{\partial Y_{k\beta}}{\partial \tilde{Y}_{\alpha i}} = \sum_{\alpha,i,j,k,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} \frac{\partial (\tilde{Y}A)_{k\beta}}{\partial \tilde{Y}_{\alpha i}}.$$

Then,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,k,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \sum_{l} \tilde{Y}_{kl} A_{l\beta}) = \sum_{\alpha,i,j,k,\beta,l} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\frac{\partial \tilde{Y}_{kl}}{\partial \tilde{Y}_{\alpha i}}) A_{l\beta}.$$

Hence,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,k,\beta,l} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{k\beta}} H_{ji} (\delta_{\alpha k} \delta_{il}) A_{l\beta} = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} H_{ji} A_{i\beta}.$$

This gives

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} (AGA^{-1})_{ji} A_{i\beta} = \sum_{\alpha,i,j,\beta} \tilde{Y}_{\alpha j} \frac{\partial}{\partial Y_{\alpha\beta}} (\sum_{k,l} A_{jk} G_{kl} A^{-1}_{li}) A_{i\beta},$$

and then,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,i,j,\beta,k,l} \tilde{Y}_{\alpha j} A_{jk} \frac{\partial}{\partial Y_{\alpha\beta}} G_{kl} A_{li}^{-1} A_{i\beta} = \sum_{\alpha,\beta,k,l} (\sum_j \tilde{Y}_{\alpha j} A_{jk}) \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} (\sum_i A_{li}^{-1} A_{i\beta}).$$

Hence,

$$\sum_{\alpha,i,j} \tilde{Y}_{\alpha j}(\frac{\partial}{\partial \tilde{Y}_{\alpha i}} H_{ji}) = \sum_{\alpha,\beta,k,l} (\tilde{Y} A)_{\alpha k} \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} (A^{-1}A)_{l\beta} = \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}. \tag{3.6}$$

Also, we have

$$\sum_{\alpha,i,j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}) H_{ji} = \sum_{\alpha,i,j} \delta_{ij} H_{ji} = \sum_{\alpha,i} H_{ii} = \sum_{\alpha} \mathrm{tr}(H) = nq\mathrm{tr}(H) = nq\mathrm{tr}(AGA^{-1})$$

Then,

$$\sum_{\alpha,i,j} (\frac{\partial}{\partial \tilde{Y}_{\alpha i}} \tilde{Y}_{\alpha j}) H_{ji} = nq\mathrm{tr}(A^{-1}AG) = nq\mathrm{tr}(G). \tag{3.7}$$

Therefore, by (3.6) and (3.7), we get

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y} H) = nq\mathrm{tr}(G) + \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.$$

Further, we have

$$\mathrm{tr}(Y^\top (\nabla_Y G^\top)) = \sum_k (Y^\top (\nabla_Y G^\top))_{kk} = \sum_{k,\alpha} Y^\top_{k\alpha} (\nabla_Y G^\top)_{\alpha k} = \sum_{k,\alpha} Y^\top_{k\alpha} \sum_\beta (\nabla_Y)_{\alpha\beta} G^\top_{\beta k}.$$

Then,

$$\mathrm{tr}(Y^\top (\nabla_Y G^\top)) = \sum_{\alpha,\beta,k} Y^\top_{k\alpha} \frac{\partial G^\top_{\beta k}}{\partial Y_{\alpha\beta}} = \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}, \tag{3.8}$$

which completes the proof.

□

In Propositions 3.1 and Propositions 3.2, we considered matrices g and G in their general forms. In the following proposition, we apply these general results to specific forms of matrices g and G. This proposition combines and extends the ideas presented in Lemma 1 and Lemma 2 of Chételat and Wells (2012). Specifically, Parts (*ii*) and (*iv*) present generalized versions of the results found in Lemma 1 and Lemma 2 of Chételat and Wells (2012).

Proposition 3.3. *Let Y be an $nq \times p$ matrix and $S = Y^\top Y$. Let X be a $p \times q$ matrix, $F = \mathrm{tr}(X^\top S^+ X)$ and r be a differentiable function. Let $\tilde{Y} = YA^{-1}$, $G(X,S) = \frac{r^2(F)}{F^2} S^+ X X^\top S^+ S$, $g(X,S) = \frac{r(F) S S^+ X}{F}$ and $H = AGA^{-1}$. Then, under the conditions of Theorem 3.3, we have*

$$(i) \quad \mathrm{tr}(G) = \frac{r^2(F)}{F},$$

$$(ii) \quad \mathrm{tr}\left(Y^\top \nabla_Y G^\top\right)$$
$$= -\frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+ X)^2\right) + \frac{r^2(F)}{F}\left(\frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2} + p - 2\mathrm{tr}(SS^+) - 1\right),$$

$$(iii) \quad \mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}\left(nq + p - 2\mathrm{tr}(SS^+) - 1 + \frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\right) - \frac{4r(F)r'(F)}{F^2},$$

$$(iv) \quad \mathrm{tr}(\nabla_X g(X,S)^\top) = 2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2),$$

$$(v) \quad \mathrm{tr}\left(g^\top(X,S)\Sigma^{-1} g(X,S)\right) = \mathrm{tr}\left(\Sigma^{-1} S G\right),$$

$$(vi) \quad \mathrm{E}_\theta \left[\mathrm{tr} \left(g^\top(X,S)\Sigma^{-1} g(X,S) \right) \right] = \mathrm{E} \left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y}H) \right],$$

provided that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{YH})| \right] < \infty$.

Proof.

$$(i) \quad \mathrm{tr}(G) = \mathrm{tr}\left(\frac{r^2(F)}{F^2} S^+ X X^\top S^+ S \right) = \frac{r^2(F)}{F^2} \mathrm{tr}(S^+ X X^\top S^+ S) = \frac{r^2(F)}{F^2} \mathrm{tr}(X^\top S^+ S S^+ X).$$

Then,

$$\mathrm{tr}(G) = \frac{r^2(F)}{F^2} \mathrm{tr}(X^\top S^+ X) = \frac{r^2(F)}{F^2} F = \frac{r^2(F)}{F}.$$

(ii) From (3.8), we have

$$\mathrm{tr}\left(Y^\top (\nabla_Y G^\top) \right) = \sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}}.$$

Therefore, by Part (iv) of Lemma 3.2, we get

$$\begin{aligned} &\mathrm{tr}\left(Y^\top (\nabla_Y G^\top) \right) \\ &\quad = -\frac{4r(F)r'(F)}{F^2} \mathrm{tr}\left((X^\top S^+ X)^2 \right) + \frac{r^2(F)}{F} \left(\frac{4\mathrm{tr}\left((X^\top S^+ X)^2 \right)}{F^2} + p - 2\mathrm{tr}(SS^+) - 1 \right). \end{aligned}$$

(iii) By Part (i) and Part (ii) together with Part (iv) of Proposition 3.2, we get

$$\begin{aligned} \mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y}H) = nq\mathrm{tr}(G) + \mathrm{tr}(Y^\top \nabla_Y G^\top) &= \frac{nqr^2(F)}{F} - \frac{4r(F)r'(F)}{F^2} \mathrm{tr}\left((X^\top S^+ X)^2 \right) \\ &+ \frac{r^2(F)}{F} \left(\frac{4\mathrm{tr}\left((X^\top S^+ X)^2 \right)}{F^2} + p - 2\mathrm{tr}(SS^+) - 1 \right). \end{aligned}$$

Then,

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) = \frac{r^2(F)}{F}\left(nq + p - 2\mathrm{tr}(SS^+) - 1 + \frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\right) - \frac{4r(F)r'(F)}{F^2}.$$

(iv) From Part (ii) of Proposition 3.1, we get

$$\mathrm{tr}\left(\nabla_X g^\top(X,S)\right) = \sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}},$$

and by Part (iv) of Lemma 3.3, we have

$$\sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}} = 2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2).$$

(v) $\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right) = \mathrm{tr}\left(\Sigma^{-1}g(X,S)g^\top(X,S)\right).$

Then,

$$\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right) = \mathrm{tr}\left(\Sigma^{-1}\frac{r^2(F)}{F^2}SS^+XX^\top SS^+\right) = \mathrm{tr}\left(\Sigma^{-1}SG\right).$$

(vi) From Part (ii) to (v) we have

$$\mathrm{E}\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right] = \mathrm{E}\left[\mathrm{tr}(\Sigma^{-1}SG)\right] = \mathrm{E}\left[\mathrm{tr}(\tilde{S}H)\right].$$

Therefore,

$$\mathrm{E}\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right] = \mathrm{E}\left[\mathrm{vec}(\tilde{Y}).\mathrm{vec}(\tilde{Y}H)\right] = \mathrm{E}\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right],$$

which completes the proof. □

As previously mentioned, Parts (ii) and Part (iv) of Proposition 3.3 provide generalizations of Chételat and Wells (2012)'s Lemma 1 and Lemma 2. Indeed in special case where $q = 1$, we have $F = \mathrm{tr}(X^\top S^+ X) = X^\top S^+ X$, and this yields the results established in Chételat and Wells (2012) (Lemma 1 and Lemma 2).

3.2 Main results

In this section, we show the main theorem of this thesis, as stated in Theorem 3.3. We demonstrate that the proposed Baranchik (1970) type estimator, outperforms the classical maximum likelihood estimator (MLE) for the mean matrix in the context of matrix normal distribution.

The following Lemma establishes an intriguing connection between the existence of $\mathrm{E}\left[\frac{1}{F}\right]$ and the rank of the matrix S, denoted as R. It highlights the significance of having $\mathrm{P}(qR > 2) = 1$. Without this condition, the result of Theorem 3.1 does not hold.

Lemma 3.4. *Let $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$ and $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$. Let $F = \mathrm{tr}(X^\top S^+ X)$ where $S = Y^\top Y$ and $R = \mathrm{rank}(S)$. Then, $\mathrm{E}\left[\frac{1}{F}\right] < \infty$ if and only if $\mathrm{P}(qR > 2) = 1$.*

Proof. Assume that $P(qR > 2) = 1$. Let $U = A^{-1}X$. Then, we have

$$X^\top S^+ X = X^\top A^{-1} A S^+ A A^{-1} X = (A^{-1}X)^\top A S^+ A (A^{-1}X) = U^\top A S^+ A U. \tag{3.9}$$

Since $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$, thus $U = A^{-1}X \sim \mathcal{N}_{p\times q}(A^{-1}\theta, I_p \otimes I_q)$. Therefore,

$$\mathrm{vec}(U) \sim \mathcal{N}_{pq}(\mathrm{vec}(A^{-1}\theta), I_{pq}). \tag{3.10}$$

Let C be $R \times p$-matrix of the form $C = [I_R \vdots 0_{R\times(p-R)}]$ and let $U_{(1)} = CU$. We have

$$\mathrm{vec}(U_{(1)}) = \mathrm{vec}(CU) = (I_q \otimes C)\mathrm{vec}(U).$$

Then,

$$\text{vec}(U_{(1)})\Big|R \sim \mathcal{N}_{qR}((I_q \otimes C)\text{vec}(A^{-1}\theta), (I_q \otimes C)(I_q \otimes C)^\top).$$

Therefore,

$$\text{vec}(U_{(1)})\Big|R \sim \mathcal{N}_{qR}((I_q \otimes C)\text{vec}(A^{-1}\theta), I_{qR}). \tag{3.11}$$

Let λ^+_{min} and λ^+_{max} be the smallest and biggest nonzero eigenvalues of AS^+A respectively. Since AS^+A is semi-positive definite, we have

$$\lambda^+_{min} U_{(1)}^\top U_{(1)} \le U^\top AS^+AU \le \lambda^+_{max} U_{(1)}^\top U_{(1)}. \tag{3.12}$$

Therefore, together with (3.9), we get

$$\frac{1}{F} \le \frac{1}{\lambda^+_{min}\text{tr}(U_{(1)}^\top U_{(1)})} = \frac{\lambda^\dagger_{max}}{\text{tr}(U_{(1)}^\top U_{(1)})} = \frac{\lambda^\dagger_{max}}{\text{vec}^\top(U_{(1)})\text{vec}(U_{(1)})}, \tag{3.13}$$

where $\lambda^\dagger_{max}$ is the biggest nonzero eigenvalue of $(AS^+A)^+ = A^{-1}SA^{-1}$. Note that $\lambda^\dagger_{max}$ depends on S and $U_{(1)}$ depends on R and X. Since, S and X are independent, we get

$$\begin{aligned}\text{E}\left[\frac{\lambda^\dagger_{max}}{\text{vec}^\top(U_{(1)})\text{vec}(U_{(1)})}\right] &= \text{E}\left[\text{E}\left[\frac{\lambda^\dagger_{max}}{\text{vec}^\top(U_{(1)})\text{vec}(U_{(1)})}\Big|R\right]\right] \\ &= \text{E}\left[\text{E}\left[\lambda^\dagger_{max}\Big|R\right]\text{E}\left[\frac{1}{\text{vec}^\top(U_{(1)})\text{vec}(U_{(1)})}\Big|R\right]\right].\end{aligned}$$

Further, we have

$$\lambda^\dagger_{max} \le \text{tr}(A^{-1}SA^{-1}) = \text{tr}(A^{-1}Y^\top YA^{-1}) = \text{tr}((YA^{-1})^\top YA^{-1}) = \text{vec}^\top(YA^{-1})\text{vec}(YA^{-1}),$$

where $\mathrm{vec}(YA^{-1}) \sim \mathcal{N}_{npq}(0, I_p \otimes I_{nq})$. Therefore, we get

$$\mathrm{E}[\lambda^{\dagger}_{max}] \leq \mathrm{E}[\mathrm{vec}^{\top}(YA^{-1})\mathrm{vec}(YA^{-1})] \quad = \mathrm{tr}(I_p \otimes I_{nq}) = \mathrm{tr}(I_{npq}) = npq.$$

Hence

$$\mathrm{E}[\lambda^{\dagger}_{max}] \leq npq. \tag{3.14}$$

Since $\mathrm{vec}(U_{(1)})\Big|R \sim \mathcal{N}_{qR}((I_q \otimes C)\mathrm{vec}(A^{-1}\theta), I_{qR})$, we get

$$\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})\Big|R \sim \chi^2_{qR}(\delta_R), \tag{3.15}$$

where $\delta_R = \Big((I_q \otimes C)\mathrm{vec}(A^{-1}\theta)\Big)^{\top}\Big((I_q \otimes C)\mathrm{vec}(A^{-1}\theta)\Big)$.

Let Z be a random variable such that $Z\Big|R \sim Poisson(\delta_R/2)$. By (3.15), we have

$$\begin{aligned}
\mathrm{E}\left[\Big(\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})\Big)^{-1}\Big|R\right] &= \mathrm{E}\left[\mathrm{E}\left[\Big(\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})\Big)^{-1}\Big|R, Z\right]\Big|R\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\Big(\chi^2_{qR+2Z}\Big)^{-1}\Big|R, Z\right]\Big|R\right] \\
&= \mathrm{E}\left[\frac{2^{-1}\Gamma(\frac{qR+2Z}{2} - 1)}{\Gamma(\frac{qR+2Z}{2})}\Big|R\right].
\end{aligned}$$

We have $\frac{2^{-1}\Gamma(\frac{qR+2Z}{2}-1)}{\Gamma(\frac{qR+2Z}{2})} = \frac{1}{qR+2Z-2}$. Therefore, since $\mathrm{P}(qR > 2) = \mathrm{P}(qR \geq 3) = 1$ and $\mathrm{P}(Z \geq 0) = 1$, we have $qR + 2Z - 2 \geq 1$ with probability one, and then,

$$\mathrm{E}\left[\Big(\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})\Big)^{-1}\Big|R\right] = \mathrm{E}\left[\frac{1}{Rq + 2Z - 2}\Big|R\right] \leq 1 \quad \text{almost surely.}$$

Therefore, together with (3.13) and (3.14), we get

$$\mathrm{E}\left[\frac{1}{F}\right] \leq \mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{max}\Big|R\right]\mathrm{E}\left[\frac{1}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\Big|R\right]\right] \leq \mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{max}\Big|R\right]\right] = \mathrm{E}\left[\lambda^{\dagger}_{max}\right] \leq npq.$$

Hence,

$$\mathrm{E}\left[\frac{1}{F}\right] \leq npq < \infty.$$

Now, assume that $\mathrm{E}[\frac{1}{F}] < \infty$. Further, from (3.12), we have

$$\frac{1}{\lambda^{+}_{max}\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})} = \frac{\lambda^{\dagger}_{min}}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})} \leq \frac{1}{F},$$

where $\lambda^{\dagger}_{min}$ is the smallest nonzero eigenvalue of $A^{-1}SA^{-1}$. Again note that $\lambda^{\dagger}_{min}$ depends on S and $U_{(1)}$ depends on R and X. Since, S and X are independent, we get

$$\mathrm{E}\left[\frac{\lambda^{\dagger}_{min}}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{\lambda^{\dagger}_{min}}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\Big| R\right]\right].$$

This gives

$$\mathrm{E}\left[\frac{\lambda^{\dagger}_{min}}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\right] = \mathrm{E}\left[\mathrm{E}\left[\lambda^{\dagger}_{min}\Big| R\right]\mathrm{E}\left[\frac{1}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\Big| R\right]\right] \leq \mathrm{E}[\frac{1}{F}] < \infty.$$

Then,

$$\mathrm{P}\left(\mathrm{E}\left[\lambda^{\dagger}_{min}\Big| R\right]\mathrm{E}\left[\frac{1}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\Big| R\right] < \infty\right) = 1. \tag{3.16}$$

Since $0 < \mathrm{E}\left[\lambda^{\dagger}_{min}\Big| R\right] < \infty$, we get

$$\mathrm{P}\left(\mathrm{E}\left[\frac{1}{\mathrm{vec}^{\top}(U_{(1)})\mathrm{vec}(U_{(1)})}\Big| R\right] < \infty\right) = 1.$$

Therefore, by (3.15), we get

$$\mathrm{P}\left(\mathrm{E}\left[(\chi^2_{qR}(\delta_R))^{-1}\Big| R\right] < \infty\right) = 1,$$

This implies that

$$\mathrm{P}\left(qR > 2\right) = 1,$$

which completes the proof. □

In Part (i) of Proposition 3.1, and Part (vi) of Proposition 3.3, we suppose that the quantities $\mathrm{E}\left[|\mathrm{tr}(\nabla_X g(X,S)^\top)|\right]$ and $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]$ exist. Now, in the following theorem, we give the conditions under which these expectations are well-defined.

Theorem 3.1. *Let $X \sim N_{p\times q}(\theta, \Sigma \otimes I_q)$,$Y \sim N_{nq\times p}(0, I_{nq} \otimes \Sigma)$ and for A the symmetric positive definite square root of Σ, let $\tilde{Y} = YA^{-1}$. Let r be any bounded differentiable non-negative function $r : \mathbb{R} \longrightarrow [0, C_1]$ with bounded derivative $|r'| \leq C_2$. Define*

$$G(X,S) = \frac{r^2(F)}{F^2} S^+ X X^\top S^+ S \text{ and } g(X,S) = \frac{r(F) S S^+ X}{F},$$

where $F = \mathrm{tr}(X^\top S^+ X)$ and $H = AGA^{-1}$. Let $R = rank(S)$ and suppose that $\mathrm{P}(qR > 2) = 1$. Then

(i) $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty,$

(ii) $\mathrm{E}\left[|\mathrm{tr}(\nabla_X g(X,S)^\top)|\right] < \infty.$

Proof. (i) By Proposition 3.3, we get

$$\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| = \left|\frac{r^2(F)}{F}\left(nq + p - 2\mathrm{tr}(SS^+) - 1 + \frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\right) - \frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+ X)^2\right)\right|.$$

Then, by triangle inequality, we have

$$\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| \leq \frac{r^2(F)}{F}\left|nq+p-2\mathrm{tr}(SS^+)-1+\frac{4\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^2}\right| + \frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+X)^2\right).$$

Hence,

$$\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| \leq \frac{C_1^2}{F}\left|nq+p-2\mathrm{tr}(SS^+)-1+\frac{4\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^2}\right| + \frac{4C_1C_2}{F^2}\mathrm{tr}\left((X^\top S^+X)^2\right),$$

and then,

$$\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| \leq \frac{C_1^2}{F}\left|nq+p-2\mathrm{tr}(SS^+)-1\right| + \frac{4C_1^2\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^3} + \frac{4C_1C_2}{F^2}\mathrm{tr}\left((X^\top S^+X)^2\right).$$

Therefore, since $\mathrm{tr}(SS^+) = \min(nq, p)$ almost surely, we have

$$\mathrm{E}\left[\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right|\right] \leq C_1^2|nq+p-2\min(nq,p)-1|\mathrm{E}\left[\frac{1}{F}\right] + 4C_1^2\mathrm{E}\left[\frac{\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^3}\right] + 4C_1C_2\mathrm{E}\left[\frac{\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^2}\right]. \tag{3.17}$$

Let d_i's be the eigenvalues of $X^\top S^+X$. Since $X^\top S^+X$ is semi-positive definite, we have

$$\mathrm{tr}\left((X^\top S^+X)^2\right) = \sum_i d_i^2 \leq (\sum_i d_i)^2 = \mathrm{tr}^2\left(X^\top S^+X\right) = F^2.$$

Then,

$$\frac{\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2} \le 1. \tag{3.18}$$

Therefore,

$$\mathrm{E}\left[\frac{\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\right] \le 1, \tag{3.19}$$

and then

$$\frac{\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^3} = \frac{\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\frac{1}{F} \le \frac{1}{F}.$$

Therefore,

$$\mathrm{E}\left[\frac{\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^3}\right] \le \mathrm{E}\left[\frac{1}{F}\right]. \tag{3.20}$$

Then, by (3.19) and (3.20) together with (3.17), we get

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] \le C_1^2\left|nq+p-2\mathrm{min}(nq,p)-1\right|\mathrm{E}\left[\frac{1}{F}\right] + 4C_1^2\mathrm{E}\left[\frac{1}{F}\right] + 4C_1C_2. \tag{3.21}$$

Further, since $\mathrm{P}(qR > 2) = 1$, by Lemma 3.4, we get $\mathrm{E}\left[\frac{1}{F}\right] < \infty$.

(ii) Similarly to Part (i) , by Part (iv) of Proposition 3.3, we get

$$\begin{aligned}\mathrm{E}\left[|\mathrm{tr}(\nabla_X g(X,S)^\top)|\right] &= \mathrm{E}\left[\left|2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2)\right|\right] \\ &\le 2C_2 + C_1|q\mathrm{min}(nq,p) - 2|\mathrm{E}\left[\frac{1}{F}\right] < \infty,\end{aligned}$$

which completes the proof. □

The following corollary demonstrates that if the number of columns of X, denoted as q, is greater than or equal to 3, then the result stated in Theorem 3.1 can be automatically derived.

Corollary 3.1. *Let* $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$,$Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$ *and for* A *the symmetric positive definite square root of* Σ*, let* $\tilde{Y} = YA^{-1}$*. Let* r *be any bounded differentiable non-negative function* $r : \mathbb{R} \longrightarrow [0, C_1]$ *with bounded derivative* $|r'| \leq C_2$*. Define*

$$G(X, S) = \frac{r^2(F)}{F^2} S^+ X X^\top S^+ S, \text{ where } F = \operatorname{tr}(X^\top S^+ X), \text{ and } H = AGA^{-1}.$$

If $q \geq 3$*, then, for all* n *and* p

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y} H)|\right] < \infty.$$

Proof. Since $q \geq 3$ and $R \geq 1$, we have $qR > 2$. Therefore, $\mathrm{P}(qR > 2) = 1$. Then, by Theorem 3.1, we get

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y} H)|\right] < \infty,$$

which completes the proof. □

The subsequent corollary shows that under some conditions, $\mathrm{P}(qR > 2) = 1$ becomes both necessary and sufficient for the existence of $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y} H)|\right]$. In essence, this corollary generalizes the findings of Corollary 2.1, where similar results were derived in the context of p-dimensional normal distribution.

Corollary 3.2. *Let* $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$,$Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$ *and for* A *the symmetric positive definite square root of* Σ*, let* $\tilde{Y} = YA^{-1}$*. Let* r *be any bounded differentiable positive function* $r : \mathbb{R} \longrightarrow [C^*, C_1]$ *with bounded derivative* $|r'| \leq C_2$*. Suppose that* $|p - nq| > 1$*. Define* $G(X, S) = \frac{r^2(F)}{F^2} S^+ X X^\top S^+ S$*, where* $F = \operatorname{tr}(X^\top S^+ X)$ *and* $H = AGA^{-1}$*. Then*

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})} \mathrm{vec}(\tilde{Y} H)|\right] < \infty$$

if and only if $\mathrm{P}(qR > 2) = 1$.

Proof. If $\mathrm{P}(qR > 2) = 1$, then by Theorem 3.1, we get

$$\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty.$$

Now, assume that $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right] < \infty$. We have

$$\begin{aligned}\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) =& \frac{r^2(F)}{F}\left(nq + p - 2\mathrm{min}(p, nq) - 1 + \frac{4\mathrm{tr}\left((X^\top S^+ X)^2\right)}{F^2}\right) \\ &- \frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+ X)^2\right).\end{aligned}$$

Therefore, since $0 \leq \frac{\mathrm{tr}((X^\top S^+ X)^2)}{F^2} \leq 1$ and $nq + p - 2\mathrm{min}(p, nq) = |p - nq|$, we get

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) \geq \frac{r^2(F)}{F}\left(|p - nq| - 1\right) - 4C_1C_2.$$

Therefore, since $|p - nq| > 1$, we get

$$\begin{aligned}\frac{r^2(F)}{F} \leq& \frac{1}{|p - nq| - 1}\left(\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) + 4C_1C_2\right) \\ =& \frac{1}{|p - nq| - 1}\left(\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H) + 4C_1C_2\right|\right) \\ \leq& \frac{1}{|p - nq| - 1}\left(\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right| + 4C_1C_2\right).\end{aligned}$$

Therefore,

$$\mathrm{E}\left[\frac{r^2(F)}{F}\right] \leq \frac{1}{|p - nq| - 1}\left(\mathrm{E}\left[\left|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right|\right] + 4C_1C_2\right) < \infty.$$

Further, we have

$$\frac{(C^*)^2}{F} \leq \frac{r^2(F)}{F} \Rightarrow (C^*)^2\mathrm{E}\left[\frac{1}{F}\right] \leq \mathrm{E}\left[\frac{r^2(F)}{F}\right] < \infty,$$

which implies that

$$\mathrm{E}\left[\frac{1}{F}\right] < \infty,$$

which completes the proof. □

Now, we introduce the main results of this chapter, crucial in proving Theorem 3.3, where we establish that the proposed estimator, δ_r, outperforms the MLE, X. These results give a generalized version of Proposition 2.4. Specifically, by setting $q = 1$, $F = \mathrm{tr}(X^\top S^+ X) = X^\top S^+ X$. This leads to the results presented in Proposition 2.4.

Theorem 3.2. *Let $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$, $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$ and $F = \mathrm{tr}(X^\top S^+ X)$ where $S = Y^\top Y$. Let $g(X,S) = \frac{r(F)SS^+X}{F}$, where r is a differentiable function. Let $R = rank(S)$ and suppose that $\mathrm{P}(qR > 2) = 1$, then*

$$(i) \quad \mathrm{E}_\theta\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}(X-\theta)\right)\right] = \mathrm{E}_\theta\left[2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2)\right],$$

$$(ii) \quad \mathrm{E}_\theta\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right]$$
$$= \mathrm{E}_\theta\left[\frac{r^2(F)}{F}\left(nq + p - 2\mathrm{tr}(SS^+) - 1 + \frac{4\mathrm{tr}((X^\top S^+ X)^2)}{F^2}\right) - \frac{4r(F)r'(F)}{F^2}\right].$$

Proof. (i) From Part (i) of Proposition 3.1, we have

$$\mathrm{E}_\theta[\mathrm{tr}(g^\top(X,S)\Sigma^{-1}(X-\theta))] = \mathrm{E}_\theta[\mathrm{tr}(\nabla_X g^\top(X,S))],$$

and from Part (iv) of Proposition 3.3, we have

$$\mathrm{tr}(\nabla_X g^\top(X,S)) = 2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2),$$

then

$$\mathrm{E}_\theta\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}(X-\theta)\right)\right]=\mathrm{E}_\theta\left[2r'(F)+\frac{r(F)}{F}(q\mathrm{tr}(SS^+)-2)\right].$$

(ii) Since $\mathrm{P}(qR>2)=1$, by Theorem 3.1, we get $\mathrm{E}\left[|\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)|\right]<\infty$. Therefore, from Part (vi) of Proposition 3.3, we have

$$\mathrm{E}_\theta\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right]=\mathrm{E}_\theta\left[\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)\right].$$

Further from Part (iii) of Proposition 3.3, we have

$$\mathrm{div}_{\mathrm{vec}(\tilde{Y})}\mathrm{vec}(\tilde{Y}H)=\frac{r^2(F)}{F}\left(nq+p-2\mathrm{tr}(SS^+)-1+\frac{4\mathrm{tr}((X^\top S^+X)^2)}{F^2}\right)-\frac{4r(F)r'(F)}{F^2}.$$

Hence

$$\begin{aligned}&\mathrm{E}_\theta\left[\mathrm{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right]\\&\quad=\mathrm{E}_\theta\left[\frac{r^2(F)}{F}\left(nq+p-2\mathrm{tr}(SS^+)-1+\frac{4\mathrm{tr}((X^\top S^+X)^2)}{F^2}\right)-\frac{4r(F)r'(F)}{F^2}\right],\end{aligned}$$

which completes the proof. □

We are now prepared to present the central finding of this thesis. In Theorem 3.3, we establish that under the invariant quadratic loss, the proposed Baranchik (1970) type estimator for the mean matrix of a matrix normal distribution with independent columns and unknown row covariance outperforms the maximum likelihood estimator. The proof of this theorem relies heavily on Theorem 3.2. Notably, this theorem extends the primary result of Chételat and Wells (2012) and Theorem 2.2 in Chapter 2 to the case of matrix normal distribution.

Theorem 3.3. *Let* $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$*,* $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$ *and* $S = Y^\top Y$*. Let* $F = \operatorname{tr}(X^\top S^+ X)$*,* $\delta_r(X,S) = \left(I - \frac{r(F)SS^+}{F}\right) X$*, where* r *is a differentiable function, and* $\delta^0(X) = X$*. Let* $R = \operatorname{rank}(S)$*. Suppose that* $\mathrm{P}(qR > 2) = 1$*, and suppose that*

(i) r *satisfies* $0 \le r \le \dfrac{2(q.\min(nq,p) - 2)}{nq + p - 2\min(nq,p) + 3}$

(ii) r *is non-decreasing*

(iii) r' *is bounded*

Then, under invariant quadratic loss $L(\theta,\delta) = \operatorname{tr}\left((\delta-\theta)^\top \Sigma^{-1}(\delta-\theta)\right)$*,* δ_r *dominates* δ^0*.*

Proof. Let $g(X,S) = \frac{r(F)SS^+X}{F}$. Therefore, $\delta_r = X - g(X,S)$. The risk difference under the quadratic loss between δ_r and δ^0 is

$$\begin{aligned}\Delta_\theta =& \mathrm{E}_\theta\left[\operatorname{tr}\left(\left(X - g(X,S) - \theta)^\top\right)\Sigma^{-1}\left(X - g(X,S) - \theta\right)\right)\right] \\ &- \mathrm{E}_\theta\left[\operatorname{tr}\left(\left(X-\theta\right)^\top \Sigma^{-1}\left(X-\theta\right)\right)\right] \\ &= -2\mathrm{E}_\theta\left[\operatorname{tr}\left(g^\top(X,S)\Sigma^{-1}(X-\theta)\right)\right] + \mathrm{E}_\theta\left[\operatorname{tr}\left(g^\top(X,S)\Sigma^{-1}g(X,S)\right)\right]. \end{aligned} \tag{3.22}$$

From Theorem 3.2 we have

$$\begin{aligned}\Delta_\theta = \mathrm{E}_\theta\Bigg[&\frac{r^2(F)}{F}\left(nq + p - 2\operatorname{tr}(SS^+) - 1 + \frac{4\operatorname{tr}((X^\top S^+X)^2)}{F^2}\right) - \frac{2r(F)}{F}(q\operatorname{tr}(SS^+) - 2) \\ &- 4r'(F)(1 + \frac{r(F)}{F^2})\Bigg].\end{aligned}$$

Since r is non-negative and non-decreasing, $-4r'(F)(1 + \frac{r(F)}{F^2}) \le 0$. Further, since $\frac{\operatorname{tr}((X^\top S^+X)^2)}{F^2} \le 1$ we have

$$\begin{aligned}&\frac{r^2(F)}{F}(nq + p - 2\operatorname{tr}(SS^+) - 1 + \frac{4\operatorname{tr}((X^\top S^+X)^2)}{F^2}) \\ &\qquad \le \frac{r^2(F)}{F}(nq + p - 2\operatorname{tr}(SS^+) + 3).\end{aligned} \tag{3.23}$$

Under the condition (i) on r, we have

$$r(F) \leq \frac{2(q\min(nq,p)-2)}{nq+p-2\min(nq,p)+3},$$

then,

$$\frac{r^2(F)}{F}(nq+p-2\min(nq,p)+3) \leq \frac{2r(F)}{F}(q\min(nq,p)-2).$$

Therefore, by (3.23) and since $\mathrm{tr}(SS^+) = \min(nq,p)$ almost surely, we get

$$\mathrm{E}\left[\frac{r^2(F)}{F}(nq+p-2\mathrm{tr}(SS^+)-1+\frac{4\mathrm{tr}((X^TS^+X)^2)}{F^2}) - \frac{2r(F)}{F}(q\mathrm{tr}(SS^+)-2)\right] \leq 0.$$

Hence, $\Delta_\theta \leq 0$, which completes the proof. □

Chapter 4

Numerical study

In Chapter 2 and Chapter 3, we illustrated that under certain conditions outlined in Theorem 2.2 and Theorem 3.3, the proposed δ_r estimator outperforms the Maximum Likelihood Estimator (MLE) under the invariant quadratic loss function. This significant finding motivates us to carry out some simulations in order to conduct a comparative analysis of the two estimators.

Namely, in this chapter, we conduct a comprehensive simulation study to highlight the risk dominance of the proposed estimator over the maximum likelihood estimator (MLE). We illustrate that, according to the conditions outlined in Theorem 3.3, the proposed estimator outperforms the MLE in high-dimensional settings for specific functions of r. The R code for this simulation is given in Appendix B. In this simulation, we consider $F = \mathrm{tr}(X^\top S^+ X)$, $r = \frac{1}{1+e^{-F}}$, and the proposed estimator is $\delta_r = (I - \frac{r(F)}{F} SS^+)X$. For the sake of simplicity, we assume that $\Sigma = I_p$. We generate samples for various values of $p(24, 32, 56$ and $104)$ along with 11 different matrix θ configurations for $q = 3$ fixed. For each value of p, we explore four distinct sample sizes: $n = \frac{p}{8}, \frac{p}{4}, p - 1$ and $2p$. This comprehensive approach allows us to investigate the impact of different p, n and $||\theta||$ combinations on the results of the simulation.

The Figure 4.1 gives the simulation results. One can see that the simulation study supports the theoretical findings. As expected, the risk difference between the suggested estimator δ_r and the classical MLE $\delta_0 = X$ is not positive. This leads to the dominance of δ_r over δ_0.

Furthermore, as presented in Figure 4.1, a consistent pattern becomes evident across all four cases. The risk difference between the two estimators diminishes as the norm of the mean matrix, $||\theta||$, increases. This intriguing observation serves as a compelling incentive for potential future research. Further, exploring how the mean matrix θ is related to the difference in risk between these estimators under the invariant quadratic loss opens up an interesting path for further investigation.

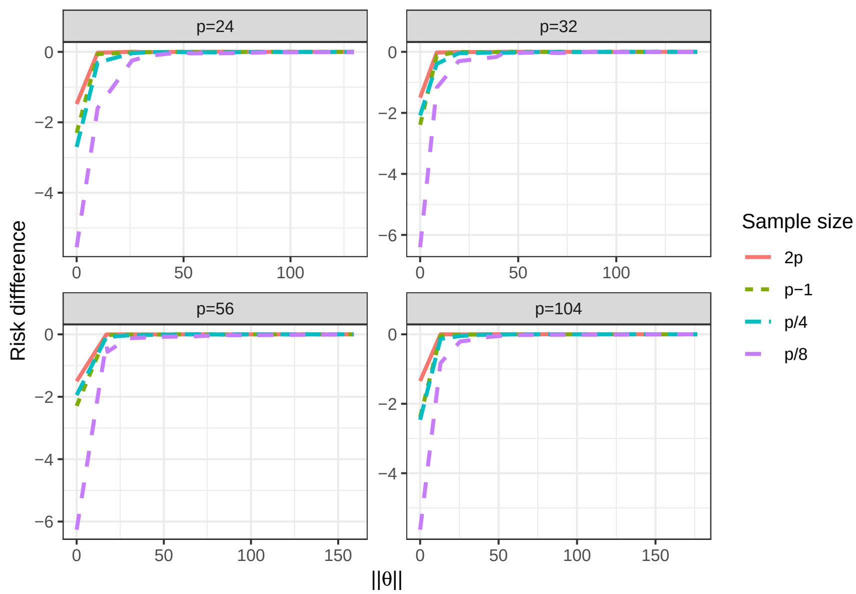

Figure 4.1: The risk difference between the proposed estimator δ_r and the MLE

Chapter 5

Conclusion

In this thesis, we demonstrated the risk dominance of our Baranchik (1970) type estimator over the classical MLE in high-dimensional data, where the number of features surpasses the number of observations, under the invariant quadratic loss. Additionally, thanks to some explorations of the estimator's rank of the unknown row covariance matrix, we established a new methodology highlighting specific conditions crucial for this dominance. Moreover, this innovative approach allowed us to revise Theorem 2 of Chételat and Wells (2012). As a direction for future research, we could aim to discover a function r that establishes dominance of the proposed estimator δ_r over the high-dimensional James-Stein estimator, $\delta^{JS}(X, S) = (I - \frac{c}{F} SS^+)X$, for any constant c. Moreover, it would be interesting to explore whether it is possible to relax the bounds discussed in Theorem 3.3. This gives us greater flexibility to select the function r, while maintaining the dominance of δ_r over δ_0.

Appendices

Appendix A

Some Technical Proofs

A.1 Distribution of Sample Covariance

Theorem A.1. *Let* $Z = [Z_1, \dots, Z_N]^\top$ *follows* $\mathcal{N}_{N\times p}(e\theta^\top, I_N \otimes \Psi)$ *where* $Z_1, \dots, Z_N$ *are independent and identically distributed random samples from* $\mathcal{N}_p(\theta, \Psi)$ *and* $e = [1, \dots, 1]^\top$ *is an N-dimensional vector. Let* $X = \bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$ *and*

$$S = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \bar{Z})(Z_i - \bar{Z})^\top.$$

Let $n = N - 1$ *and* $\Sigma = \frac{\Psi}{N}$. *Then,*

(i) S *is independent of* X *and can be rewritten as* $S = Y^\top Y$ *where* $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$,

(ii) $X \sim \mathcal{N}_p(\theta, \Sigma)$ *and* $S \sim Wishart_p(n, \Sigma)$.

Proof. *(i)* Let $Q = \sum_{i=1}^{N}(Z_i - X)(Z_i - X)^\top$. Let $U = \Gamma Z$ where Γ is an $N \times N$ orthogonal matrix with a last row $N^{-\frac{1}{2}}e^\top$. Since Γ is orthogonal, the Jacobian of transformation is

$J(Z \to U) = |\det(\Gamma)^p| = 1$. Furthermore, U can be partitioned as $\begin{bmatrix} V \\ \hline W^\top \end{bmatrix}$ where V is an $n \times p$ matrix and $W = N^{\frac{1}{2}}\bar{Z}$ is a p-dimensional vector. Then,

$$Z^\top Z = (\Gamma^\top U)^\top \Gamma^\top U = U^\top \Gamma\Gamma^\top U = U^\top U = V^\top V + WW^\top. \tag{A.1}$$

Therefore, by (A.1), we get

$$Q = \sum_{i=1}^{N} (Z_i - \bar{Z})(Z_i - \bar{Z})^\top = Z^\top Z - N\bar{Z}\bar{Z}^\top = V^\top V + WW^\top - N\bar{Z}\bar{Z}^\top.$$

Since $WW^\top = N\bar{Z}\bar{Z}^\top$, we get

$$Q = V^\top V + N\bar{Z}\bar{Z}^\top - N\bar{Z}\bar{Z}^\top = V^\top V. \tag{A.2}$$

We also have,

$$\begin{aligned}(Z - e\theta^\top)^\top (Z - e\theta^\top) &= Z^\top Z - Z^\top e\theta^\top - \theta e^\top Z + N\theta\theta^\top \\ &= V^\top V + WW^\top - Z^\top e\theta^\top - (Z^\top e\theta^\top)^\top + N\theta\theta^\top. \end{aligned} \tag{A.3}$$

Since the first n rows of Γ are orthogonal to the N-dimensional vector e, we get

$$Z^\top e\theta^\top = U^\top \Gamma e\theta^\top = [V^\top \vdots W][0 \ldots 0 N^{\frac{1}{2}}]^\top \theta^\top = N^{\frac{1}{2}} W\theta^\top. \tag{A.4}$$

Therefore, By using A.4 in A.3, we get

$$(Z - e\theta^\top)^\top (Z - e\theta^\top) = v^\top v + (W - N^{\frac{1}{2}}\theta)(W - N^{\frac{1}{2}}\theta)^\top. \tag{A.5}$$

The probability density function (pdf) of $Z \sim \mathcal{N}_{N\times p}(e\theta^\top, I_N \otimes \Psi)$ is given by

$$f_Z(z) = (2\pi)^{-\frac{Np}{2}} (\det(\Psi))^{-\frac{N}{2}} \operatorname{etr}\left[-\frac{1}{2}\Psi^{-1}(z - e\theta^\top)^\top (z - e\theta^\top)\right]. \tag{A.6}$$

Therefore, by A.5, the joint pdf of (V, W) can be written as

$$f_{(V,W)}(v,w) = (2\pi)^{-\frac{(n+1)p}{2}}(\det(\Psi))^{-\frac{n+1}{2}}\text{etr}\left[-\frac{1}{2}\Psi^{-1}\left(v^\top v + (w - N^{\frac{1}{2}}\theta)(w - N^{\frac{1}{2}}\theta)^\top\right)\right].$$

Since

$$\begin{aligned}\text{tr}\Bigg(\Psi^{-1}(w - N^{\frac{1}{2}}\theta)(w - N^{\frac{1}{2}}\theta)^\top\Bigg) &= \text{tr}\Bigg((w - N^{\frac{1}{2}}\theta)^\top\Psi^{-1}(w - N^{\frac{1}{2}}\theta)\Bigg)\\ &= (w - N^{\frac{1}{2}}\theta)^\top\Psi^{-1}(w - N^{\frac{1}{2}}\theta),\end{aligned}$$

then,

$$\begin{aligned}f_{(V,W)}(v,w) = (2\pi)^{-\frac{np}{2}}(\det(\Psi))^{-\frac{n}{2}}\text{etr}\Bigg[-\frac{1}{2}\Psi^{-1}v^\top v\Bigg](2\pi)^{-\frac{p}{2}}(\det(\Psi))^{-\frac{1}{2}}\\ \exp\Bigg(-\frac{1}{2}(w - N^{\frac{1}{2}}\theta)^\top\Psi^{-1}(w - N^{\frac{1}{2}}\theta)\Bigg).\end{aligned} \tag{A.7}$$

Therefore, by A.7, we get $V \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Psi)$ independent of $W \sim \mathcal{N}_p(N^{\frac{1}{2}}\theta, \Psi)$. Let $Y = N^{-\frac{1}{2}}V$. Then, $S = N^{-1}Q = N^{-1}V^\top V = Y^\top Y$ and $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$.

(ii) Since $Y \sim \mathcal{N}_{n\times p}(0, I_n \otimes \Sigma)$, by the definition of wishart distribution we get,

$$S = Y^\top Y \sim \text{Wishart}_p(n, \Sigma).$$

We also have $X = \bar{Z} = N^{-\frac{1}{2}}W \sim \mathcal{N}_p(\theta, \Sigma)$, which completes the proof. □

Theorem A.2. *Let* $Z = [Z_1, \dots, Z_N]^\top$ *follows* $\mathcal{N}_{Nq\times p}(\gamma\theta^\top, I_{Nq} \otimes \Psi)$ *where* $Z_1, \dots, Z_N$ *are independent and identically distributed random samples from* $\mathcal{N}_{p\times q}(\theta, \Psi \otimes I_q)$*,* $\gamma = e \otimes I_q$ *and* $e = [1, \dots, 1]^\top$ *is an N-dimensional vector. Let* $X = \bar{Z} = \frac{1}{N}\sum_{i=1}^N Z_i$ *and*

$$S = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \bar{Z})(Z_i - \bar{Z})^\top.$$

Let $n = N - 1$ *and* $\Sigma = \frac{\Psi}{N}$*. Then,*

(i) *S is independent of X and can be rewritten as* $S = Y^\top Y$ *where* $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$*,*

(ii) $X \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$ *and* $S \sim Wishart_p(nq, \Sigma)$*.*

Proof. *(i)* Let $U = (\Gamma \otimes I_q)Z$ where Γ is an $N \times N$ orthogonal matrix with a last row $N^{-\frac{1}{2}}e^\top$. The Jacobian of transformation is

$$J(Z \to U) = |\det{(\Gamma \otimes I_q)^p}| = \left|\left(\det^N(\Gamma)\det^q(I_q)\right)^p\right| = 1.$$

U can be partitioned as $\left[\begin{array}{c} V \\ \hline W^\top \end{array}\right]$ where V is an $nq \times p$ matrix and W is a $p \times q$ matrix. Then,

$$\begin{aligned} Z^\top Z = \left((\Gamma \otimes I_q)^{-1}U\right)^\top (\Gamma \otimes I_q)^{-1}U &= U^\top(\Gamma^{-1} \otimes I_q)^\top(\Gamma^{-1} \otimes I_q)U \\ &= U^\top(\Gamma \otimes I_q)(\Gamma^{-1} \otimes I_q)U. \end{aligned}$$

Therefore,

$$Z^\top Z = U^\top\left((\Gamma\Gamma^{-1}) \otimes (I_qI_q)\right)U = U^\top\left(I_N \otimes I_q\right)U = U^\top I_{Nq}U = U^\top U. \tag{A.8}$$

We also have,

$$U^\top U = V^\top V + WW^\top. \tag{A.9}$$

Hence, by A.8 and A.9, we get

$$Q = \sum_{i=1}^{N} (Z_i - \bar{Z})(Z_i - \bar{Z})^\top = Z^\top Z - N\bar{Z}\bar{Z}^\top = V^\top V + WW^\top - N\bar{Z}\bar{Z}^\top. \tag{A.10}$$

We also have,

$$W^\top = (N^{\frac{1}{2}} e^\top \otimes I_q)Z = N^{-\frac{1}{2}}[I_q \dots I_q]Z = N^{-\frac{1}{2}} \sum_{i=1}^{N} Z_i^\top = N^{\frac{1}{2}} \bar{Z}^\top. \tag{A.11}$$

Therefore, by A.10 and A.11, we get

$$Q = V^\top V. \tag{A.12}$$

The probability density function (pdf) of $Z \sim \mathcal{N}_{Nq\times p}(\gamma\theta^\top, I_{Nq} \otimes \Psi)$ is given by

$$f_Z(z) = (2\pi)^{-\frac{Nqp}{2}} (\det(\Psi))^{-\frac{Nq}{2}} \operatorname{etr}\left[-\frac{1}{2}\Psi^{-1}(z - \gamma\theta^\top)^\top(z - \gamma\theta^\top)\right]. \tag{A.13}$$

We also have,

$$Z^\top \gamma\theta^\top = U^\top(\Gamma \otimes I_q)(e \otimes I_q)\theta^\top = U^\top \left((\Gamma e) \otimes (I_q I_q)\right)\theta^\top.$$

Since the first n rows of Γ is orthogonal to e, we get

$$Z^\top \gamma\theta^\top = U^\top([0 \dots 0 N^{\frac{1}{2}}]^\top \otimes I_q)\theta^\top = [V^\top \vdots W][0_q \dots 0_q N^{\frac{1}{2}} I_q]^\top \theta^\top = N^{\frac{1}{2}} W\theta^\top, \tag{A.14}$$

where 0_q is $q \times q$ square matrix of zeros.

We also have,

$$\gamma^\top \gamma = (e \otimes I_q)^\top (e \otimes I_q) = (e^\top \otimes I_q)(e \otimes I_q) = (e^\top e) \otimes (I_q I_q) = N \otimes I_q = N I_q.$$

Therefore,

$$(Z - \gamma\theta^\top)^\top (Z - \gamma\theta^\top) = Z^\top Z - Z^\top \gamma\theta^\top - (Z^\top \gamma\theta^\top)^\top + N\theta\theta^\top.$$

By using A.8, A.9 and A.14, we get

$$\begin{aligned}(Z - \gamma\theta^\top)^\top (Z - \gamma\theta^\top) &= Z^\top Z - N^{\frac{1}{2}} W\theta^\top - N^{\frac{1}{2}}\theta W^\top + N\theta\theta^\top \\ &= V^\top V + (W - N^{\frac{1}{2}}\theta)(W - N^{\frac{1}{2}}\theta)^\top. \end{aligned} \tag{A.15}$$

Therefore the joint pdf of (V, W) can be shown as

$$f_{(V,W)}(v, w) = (2\pi)^{-\frac{nqp}{2}} (\det(\Psi))^{-\frac{nq}{2}} \operatorname{etr}\left[-\frac{1}{2}\Psi^{-1} v^\top v \right] (2\pi)^{-\frac{qp}{2}} (\det(\Psi))^{-\frac{q}{2}} \operatorname{etr}\left(-\frac{1}{2}(w - N^{\frac{1}{2}}\theta)^\top \Psi^{-1} (w - N^{\frac{1}{2}}\theta) \right). \tag{A.16}$$

Therefore, by A.16, we get $V \sim \mathcal{N}_{nq\times p}(0, I_n \otimes \Psi)$ independent of $W \sim \mathcal{N}_{p\times q}(N^{\frac{1}{2}}\theta, \Psi \otimes I_q)$. Let $Y = N^{-\frac{1}{2}} V$. Then, $S = N^{-1} Q = N^{-1} V^\top V = Y^\top Y$ and $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$.

(ii) Since $Y \sim \mathcal{N}_{nq\times p}(0, I_{nq} \otimes \Sigma)$, by the definition of wishart distribution we get,

$$S = Y^\top Y \sim \text{Wishart}_p(nq, \Sigma).$$

We also have $X = \bar{Z} = N^{-\frac{1}{2}} W \sim \mathcal{N}_{p\times q}(\theta, \Sigma \otimes I_q)$, which completes the proof. □

A.2 On the Moore-Penrose inverse and Stein's Lemma

Proposition A.1. *Let A be a $m \times n$ matrix. Then*

(i) $A^+ = A^+(A^+)^\top A^\top$

(ii) $A^\top = A^\top A A^+$

Proof. *(i)* By properties of Moore-Penrose inverse we have

$$(AA^+)^\top = AA^+.$$

Since $A^+ = A^+AA^+$, we get

$$A^+ = A^+AA^+ = A^+(AA^+)^\top = A^+(A^+)^\top A^\top.$$

(ii) Similar to Part *(i)*, we get

$$A = AA^+A = (AA^+)^\top A = (A^+)^\top A^\top A.$$

Hence,

$$A^\top = ((A^+)^\top A^\top A)^\top = A^\top AA^+.$$

□

Corollary A.1. *Let $S = Y^\top Y$. Then, we have*

$$SS^+Y^\top = Y^\top.$$

Proof. The proof follows from Proposition A.1. □

Proposition A.2. *For $A(t)$ a differentiable matrix function of constant rank, we have*

$$(i) \quad \frac{\partial A^{+}}{\partial t} = -A^{+}\frac{\partial A}{\partial t}A^{+} + (I - AA^{+})\frac{\partial A^{\top}}{\partial t}(A^{+})^{\top}A^{+} + A^{+}(A^{+})^{\top}\frac{\partial A^{\top}}{\partial t}(I - AA^{+}).$$

(ii) *For the symmetric matrix, S, we have*

$$\frac{\partial S^{+}}{\partial t} = -S^{+}\frac{\partial S}{\partial t}S^{+} + (I - SS^{+})\frac{\partial S}{\partial t}S^{+}S^{+} + S^{+}S^{+}\frac{\partial S}{\partial t}(I - SS^{+}).$$

Proof. The proof of this proposition is given in Theorem 4.3 of Golub and Pereyra (1973). □

The following proposition generalizes Lemma 1 from Stein (1981) to distributions with probability density functions that exhibit the property:

$$\lim_{y \to -\infty} f_Y(y) = \lim_{y \to \infty} f_Y(y) = 0.$$

This proposition can be employed as a foundational step in deriving Lemma 1 from Stein (1981).

Proposition A.3. *Let Y be a random variable with pdf $f_Y(y)$ and $\lim_{y \to -\infty} f_Y(y) = \lim_{y \to \infty} f_Y(y) = 0$. Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be an indefinite integral of the Lebesgue measurable function g' the derivative of g. Suppose $\mathrm{E}\Big[|g'(Y)|\Big] < \infty$. Then*

$$\mathrm{E}\Big[g'(Y)\Big] = -\mathrm{E}\left[g(Y)\frac{f'_Y(Y)}{f_Y(Y)}\right]$$

Proof. Since $\lim_{y \to -\infty} f_Y(y) = \lim_{y \to \infty} f_Y(y) = 0$, we get

$$\begin{aligned} \int_{-\infty}^{y} f'_Y(z)\,dz &= (f_Y(y) - f_Y(-\infty)) = f_Y(y) \\ -\int_{y}^{\infty} f'_Y(z)\,dz &= -(f_Y(\infty) - f_Y(y)) = f_Y(y) \end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathrm{E}\Big[g'(Y)\Big] &= \int_{-\infty}^{\infty} g'(y) f_Y(Y)\, dy \; = \int_{-\infty}^{0} g'(y) f_Y(Y)\, dy \; + \int_{0}^{\infty} g'(y) f_Y(Y)\, dy \\
&= \int_{-\infty}^{0} g'(y) \left(\int_{-\infty}^{y} f'_Y(z)\, dz \right) dy \; + \int_{0}^{\infty} g'(y) \left(- \int_{y}^{\infty} f'_Y(z)\, dz \right) dy \\
&= \int_{-\infty}^{0} \int_{-\infty}^{y} g'(y) f'_Y(z)\, dz \;\, dy \; - \int_{0}^{\infty} \int_{y}^{\infty} g'(y) f'_Y(z)\, dz \;\, dy \; .
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\Big[g'(Y)\Big] &= \int_{-\infty}^{0} \int_{z}^{0} g'(y) f'_Y(z)\, dy \;\, dz \; - \int_{0}^{\infty} \int_{0}^{z} g'(y) f'_Y(z)\, dy \;\, dz \\
&= \int_{-\infty}^{0} f'_Y(z) \left(\int_{z}^{0} g'(y)\, dy \right) dz \; - \int_{0}^{\infty} f'_Y(z) \left(\int_{0}^{z} g'(y)\, dy \right) dz \\
&= \int_{-\infty}^{0} f'_Y(z)(g(0) - g(z))\, dz \; - \int_{0}^{\infty} f'_Y(z)(g(z) - g(0))\, dz \\
&= \int_{-\infty}^{0} f'_Y(z)(g(0) - g(z))\, dz \; + \int_{0}^{\infty} f'_Y(z)(g(0) - g(z))\, dz \; .
\end{aligned}$$

This gives

$$\begin{aligned}
\mathrm{E}\Big[g'(Y)\Big] &= \int_{-\infty}^{\infty} f'_Y(z)(g(0) - g(z))\, dz \; = \int_{-\infty}^{\infty} f'_Y(z) g(0)\, dz \; - \int_{-\infty}^{\infty} f'_Y(z) g(z)\, dz \\
&= g(0) \int_{-\infty}^{\infty} f'_Y(z)\, dz \; - \int_{-\infty}^{\infty} g(z) f'_Y(z)\, dz \\
&= g(0)(f_Y(\infty) - f_Y(-\infty)) - \int_{-\infty}^{\infty} g(z) \frac{f'_Y(z)}{f_Y(z)} f_Y(z)\, dz \; .
\end{aligned}$$

Hence,

$$\mathrm{E}\Big[g'(Y)\Big] = - \int_{-\infty}^{\infty} g(z) \frac{f'_Y(z)}{f_Y(z)} f_Y(z)\, dz \; = -\mathrm{E}\left[g(Y) \frac{f'_Y(Y)}{f_Y(Y)} \right],$$

which completes the proof. □

With the assistance of the aforementioned results, we can now directly derive Lemma 1 from Stein (1981).

Corollary A.2 (Stein (1981) Lemma 1)**.** *Let $Y \sim N(0,1)$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ be an indefinite integral of the Lebesgue measurable function g' the derivative of g. Suppose $E[|g'(Y)|] < \infty$. Then*

$$\mathrm{E}\Big[g'(Y)\Big] = \mathrm{E}\Big[Yg(Y)\Big].$$

Proof.

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \Rightarrow f'_Y(y) = -y\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = -yf_Y(y)$$

Therefore, by Proposition A.3, we get

$$\mathrm{E}\Big[g'(Y)\Big] = -\mathrm{E}\left[g(Y)\frac{f'_Y(Y)}{f_Y(Y)}\right] = -\mathrm{E}\left[g(Y)\frac{-Yf_Y(Y)}{f_Y(Y)}\right] = \mathrm{E}\Big[Yg(Y)\Big],$$

which completes the proof. □

A.3 Some Technical Proofs of Chapter 2

Proof of Lemma 2.1. (*i*) Let δ_{ij} be the Kronecker delta. We have

$$\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \frac{\partial}{\partial Y_{\alpha\beta}} \sum_q Y^{\top}_{kq} Y_{ql} = \sum_q \frac{\partial}{\partial Y_{\alpha\beta}} (Y_{qk}Y_{ql}) = \sum_q \left[(\frac{\partial Y_{qk}}{\partial Y_{\alpha\beta}}) Y_{ql} + Y_{qk} (\frac{\partial Y_{ql}}{\partial Y_{\alpha\beta}}) \right].$$

Then,

$$\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \sum_q (\frac{\partial Y_{qk}}{\partial Y_{\alpha\beta}}) Y_{ql} + \sum_q Y_{qk} (\frac{\partial Y_{ql}}{\partial Y_{\alpha\beta}}) = \sum_q \delta_{\beta k} Y_{ql} + \sum_q Y_{qk} \delta_{\beta l}.$$

Hence,

$$\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \delta_{\beta k} Y_{\alpha l} + \delta_{\beta l} Y_{\alpha k}.$$

(ii) By part (i) we get

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = \sum_j (A\frac{\partial S}{\partial Y_{\alpha\beta}})_{kj}B_{jl} = \sum_j \left(\sum_i A_{ki}\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{ij}\right)B_{jl}$$
$$= \sum_j \left(\sum_i A_{ki}\Big\{\delta_{\beta i}Y_{\alpha j} + \delta_{\beta j}Y_{\alpha i}\Big\}\right)B_{jl}.$$

Then,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = \sum_j \left(\sum_i A_{ki}\delta_{\beta i}Y_{\alpha j} + \sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right)B_{jl} = \sum_j \left(A_{k\beta}Y_{\alpha j} + \sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right)B_{jl}.$$

Then,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = \sum_j A_{k\beta}Y_{\alpha j}B_{jl} + \sum_j \left(\sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right)B_{jl} = A_{k\beta}\sum_j Y_{\alpha j}B_{jl}$$
$$+ \sum_i A_{ki}Y_{\alpha i}\sum_j \delta_{\beta j}B_{jl}.$$

This gives,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + \sum_i A_{ki}Y_{\alpha i}B_{\beta l} = A_{k\beta}(YB)_{\alpha l} + B_{\beta l}\sum_i^p A_{ki}Y^\top_{i\alpha}.$$

Therefore,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + B_{\beta l}(AY^\top)_{k\alpha} = A_{k\beta}(YB)_{\alpha l} + (AY^\top)_{k\alpha}B_{\beta l}.$$

(iii) We have $\dfrac{\partial F}{\partial Y_{\alpha\beta}} = \dfrac{\partial}{\partial Y_{\alpha\beta}}(X^\top S^+ X) = X^\top(\dfrac{\partial S^+}{\partial Y_{\alpha\beta}})X.$

From, Proposition A.2, we get

$$X^\top(\frac{\partial S^+}{\partial Y_{\alpha\beta}})X = X^\top\left(-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+ + (I - SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+ + S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I - SS^+)\right)X.$$

This gives

$$X^\top(\frac{\partial S^+}{\partial Y_{\alpha\beta}})X = -X^\top S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+X + X^\top(I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+X + X^\top S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)X$$
$$= -X^\top S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+X + X^\top(I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+X + X^\top S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)X.$$

Now, by Part (ii), for $k=1$ and $l=1$ we get

$$\begin{aligned}\frac{\partial X^\top S^+X}{\partial Y_{\alpha\beta}} =& -(X^\top S^+)_{1\beta}(YS^+X)_{\alpha 1} - (X^\top S^+Y^\top)_{1\alpha}(S^+X)_{\beta 1}\\ &+ (X^\top(I-SS^+))_{1\beta}(YSS^+X)_{\alpha 1} + (X^\top(I-SS^+)Y^\top)_{1\alpha}(SS^+X)_{\beta 1}\\ &+ (X^\top S^+S^+)_{1\beta}(Y(I-SS^+)X)_{\alpha 1} + (X^\top S^+S^+Y^\top)_{1\alpha}((I-SS^+)X)_{\beta 1}.\end{aligned}$$

Since

$(X^\top S^+)_{1\beta}(YS^+X)_{\alpha 1} = (X^\top S^+Y^\top)_{1\alpha}(S^+X)_{\beta 1}$,

$(X^\top(I-SS^+))_{1\beta}(YSS^+X)_{\alpha 1} = (X^\top S^+S^+Y^\top)_{1\alpha}((I-SS^+)X)_{\beta 1}$ and

$Y(I-SS^+) = (I-SS^+)Y^\top = 0$,

we get

$$\frac{\partial X^\top S^+X}{\partial Y_{\alpha\beta}} = -2(X^\top S^+Y^\top)_{1\alpha}(S^+X)_{\beta 1} + 2(X^\top S^+S^+Y^\top)_{1\alpha}((I-SS^+)X)_{\beta 1}.$$

(iv)
$$\begin{aligned}&\left(\frac{\partial S^+XX^\top SS^+}{\partial Y_{\alpha\beta}}\right)_{kl}\\ &= \left(\frac{\partial S^+}{\partial Y_{\alpha\beta}}XX^\top SS^+\right)_{kl} + \left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} + \left(S^+XX^\top S\frac{\partial S^+}{\partial Y_{\alpha\beta}}\right)_{kl}\\ &= \left((-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+ + (I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+ + S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+))XX^\top SS^+\right)_{kl}\\ &+ \left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl}\end{aligned}$$

$$+\left(S^{+}XX^{\top}S(-S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}+(I-SS^{+})\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}S^{+}+S^{+}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^{+}))\right)_{kl}.$$

Then,

$$\begin{aligned}\left(\frac{\partial S^{+}XX^{\top}SS^{+}}{\partial Y_{\alpha\beta}}\right)_{kl} &= \left(-S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}XX^{\top}SS^{+}\right)_{kl}+\left((I-SS^{+})\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}S^{+}XX^{\top}SS^{+}\right)_{kl}\\ &+\left(S^{+}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^{+})XX^{\top}SS^{+}\right)_{kl}+\left(S^{+}XX^{\top}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}\right)_{kl}\\ &-\left(S^{+}XX^{\top}SS^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}\right)_{kl}+\left(S^{+}XX^{\top}S(I-SS^{+})\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}S^{+}\right)_{kl}\\ &+\left(S^{+}XX^{\top}SS^{+}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^{+})\right)_{kl}.\end{aligned}$$

Now, by using Part (ii), we get

$$\left(-S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}XX^{\top}SS^{+}\right)_{kl}=-S^{+}_{k\beta}(YS^{+}XX^{\top}SS^{+})_{\alpha l}-(S^{+}Y^{\top})_{k\alpha}(S^{+}XX^{\top}SS^{+})_{\beta l}. \tag{A.17}$$

Further,

$$\begin{aligned}&\left((I-SS^{+})\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}S^{+}XX^{\top}SS^{+}\right)_{kl}\\ &=(I-SS^{+})_{k\beta}(YS^{+}S^{+}XX^{\top}SS^{+})_{\alpha l}+((I-SS^{+})Y^{\top})_{k\alpha}(S^{+}S^{+}XX^{\top}SS^{+})_{\beta l}\\ &=(I-SS^{+})_{k\beta}(YS^{+}S^{+}XX^{\top}SS^{+})_{\alpha l}.\end{aligned} \tag{A.18}$$

Since, $Y(I-SS^{+})=0$ (see Corollary A.1), we get

$$\begin{aligned}&\left(S^{+}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^{+})XX^{\top}SS^{+}\right)_{kl}\\ &=(S^{+}S^{+})_{k\beta}(Y(I-SS^{+})XX^{\top}SS^{+})_{\alpha l}+(S^{+}S^{+}Y^{\top})_{k\alpha}((I-SS^{+})XX^{\top}SS^{+})_{\beta l}\\ &=(S^{+}S^{+}Y^{\top})_{k\alpha}((I-SS^{+})XX^{\top}SS^{+})_{\beta l}.\end{aligned} \tag{A.19}$$

We also have

$$\left(S^{+}XX^{\top}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}\right)_{kl} = (S^{+}XX^{\top})_{k\beta}(YS^{+})_{\alpha l} + (S^{+}XX^{\top}Y^{\top})_{k\alpha}S^{+}_{\beta l}. \tag{A.20}$$

Further, we have

$$\left(-S^{+}XX^{\top}SS^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}\right)_{kl} = -(S^{+}XX^{\top}SS^{+})_{k\beta}(YS^{+})_{\alpha l}(S^{+}XX^{\top}SS^{+}Y^{\top})_{k\alpha}S^{+}_{\beta l}.$$

Since $SS^{+}Y^{\top} = Y^{\top}$ (see Proposition A.1), we get

$$\left(-S^{+}XX^{\top}SS^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}\right)_{kl} = -(S^{+}XX^{\top}SS^{+})_{k\beta}(YS^{+})_{\alpha l}(S^{+}XX^{\top}Y^{\top})_{k\alpha}S^{+}_{\beta l}. \tag{A.21}$$

Since $S(I - SS^{+}) = S^{+}(I - SS^{+}) = 0$, we get

$$\left(S^{+}XX^{\top}S(I - SS^{+})\frac{\partial S}{\partial Y_{\alpha\beta}}S^{+}S^{+}\right)_{kl} = 0, \tag{A.22}$$

then,

$$\begin{aligned}
&\left(S^{+}XX^{\top}SS^{+}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I - SS^{+})\right)_{kl} = \left(S^{+}XX^{\top}S^{+}\frac{\partial S}{\partial Y_{\alpha\beta}}(I - SS^{+})\right)_{kl} \\
&= \left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + (AY^{\top})_{k\alpha}B_{\beta l} \\
&= (S^{+}XX^{\top}S^{+})_{k\beta}(Y(I - SS^{+}))_{\alpha l} + (S^{+}XX^{\top}S^{+}Y^{\top})_{k\alpha}(I - SS^{+})_{\beta l} \\
&= (S^{+}XX^{\top}S^{+}Y^{\top})_{k\alpha}(I - SS^{+})_{\beta l}.
\end{aligned} \tag{A.23}$$

Therefore, by (A.17), (A.18), $\cdots$, (A.23), we get

$$\left(\frac{\partial S^{+}XX^{\top}SS^{+}}{\partial Y_{\alpha\beta}}\right)_{kl} = -S^{+}_{k\beta}(YS^{+}XX^{\top}SS^{+})_{\alpha l} - (S^{+}Y^{T})_{k\alpha}(S^{+}XX^{\top}SS^{+})_{\beta l}$$

$$+ (I - SS^+)_{k\beta}(YS^+SXX^\top SS^+)_{\alpha l} + (S^+S^+Y^\top)_{k\alpha}((I - SS^+)XX^\top SS^+)_{\beta l}$$
$$+ (S^+XX^\top)_{k\beta}(YS^+)_{\alpha l} + (S^+XX^\top Y^\top)_{k\alpha}S^+_{\beta l} - (S^+XX^\top SS^+)_{k\beta}(YS^+)_{\alpha l}$$
$$- (S^+XX^\top SS^+Y^\top)_{k\alpha}S^+_{\beta l} + (S^+XX^\top S^+Y^\top)_{k\alpha}(I - SS^+)_{\beta l},$$

which completes the proof. □

Proof of Lemma 2.2.

(i)
$$\frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} = \frac{\partial}{\partial Y_{\alpha\beta}}\left(\frac{r^2(F)}{F^2}(S^+XX^\top SS^+)_{kl}\right)$$
$$= \frac{\partial}{\partial Y_{\alpha\beta}}(\frac{r^2(F)}{F^2})(S^+XX^\top SS^+)_{kl} + \frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl}.$$

Then,

$$\frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} = \frac{2r(F)r'(F)(\frac{\partial F}{\partial y_{\alpha\beta}})F^2 - 2F(\frac{\partial F}{\partial y_{\alpha\beta}})r^2(F)}{F^4}(S^+XX^\top SS^+)_{kl}$$
$$+ \frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl}.$$

Hence,

$$\frac{\partial G_{kl}}{\partial Y_{\alpha\beta}} = \frac{2r(F)r'(F)}{F^2}(\frac{\partial F}{\partial y_{\alpha\beta}})(S^+XX^\top SS^+)_{kl} - \frac{2r^2(F)}{F^3}(\frac{\partial F}{\partial y_{\alpha\beta}})(S^+XX^\top SS^+)_{kl}$$
$$+ \frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl}.$$

(ii) By Part (iii) of Lemma 2.1, we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k}$$
$$= -2\sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{\beta\alpha}(SS^+XX^\top S^+)_{\beta k}$$

$$
\begin{aligned}
&+ 2 \sum_{\alpha,k,\beta} Y_{\alpha k}((I - SS^{+})XX^{\top}S^{+}S^{+}Y^{\top})_{\beta\alpha}(SS^{+}XX^{\top}S^{+})_{\beta k} \\
&= -2 \sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (YS^{+}XX^{\top}S^{+})_{\alpha\beta}(SS^{+}XX^{T}S^{+})_{\beta k} \\
&+ 2 \sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (YS^{+}S^{+}XX^{\top}(I - SS^{+}))_{\alpha\beta}(SS^{+}XX^{\top}S^{+})_{\beta k}.
\end{aligned}
$$

This gives

$$
\begin{aligned}
\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^{+}XX^{\top}S^{+})_{\beta k} =& -2 \sum_{\alpha,k} Y_{\alpha k}(YS^{+}XX^{\top}S^{+}SS^{+}XX^{\top}S^{+})_{\alpha k} \\
&+ 2 \sum_{\alpha,k} Y_{\alpha k}(YS^{+}S^{+}XX^{\top}(I - SS^{+})SS^{+}XX^{\top}S^{+})_{\alpha k}.
\end{aligned}
$$

Since $(I - SS^{+})SS^{+} = 0$, we get

$$
\begin{aligned}
\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^{+}XX^{\top}S^{+})_{\beta k} &= -2 \sum_{\alpha,k} Y^{\top}_{k\alpha}(YS^{+}XX^{\top}S^{+}XX^{\top}S^{+})_{\alpha k} \\
&= -2 \sum_{k} (Y^{\top}YS^{+}XX^{\top}S^{+}XX^{\top}S^{+})_{kk} \\
&= -2 \sum_{k} (SS^{+}XX^{\top}S^{+}XX^{\top}S^{+})_{kk}.
\end{aligned}
$$

This gives

$$
\begin{aligned}
\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^{+}XX^{\top}S^{+})_{\beta k} &= -2\mathrm{tr}(SS^{+}XX^{\top}S^{+}XX^{\top}S^{+}) \\
&= -2\mathrm{tr}(X^{\top}S^{+}XX^{\top}S^{+}SS^{+}X).
\end{aligned}
$$

Hence,

$$
\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^{+}XX^{\top}S^{+})_{\beta k} = -2\mathrm{tr}(X^{\top}S^{+}XX^{\top}S^{+}X) = -2\mathrm{tr}((X^{\top}S^{+}X)^{2}) = -2F^{2}.
$$

$$(iii) \quad \sum_{\alpha,k,\beta} Y_{\alpha k} \frac{\partial}{\partial y_{\alpha\beta}} (SS^{+}XX^{\top}S^{+})_{\beta k} = \sum_{\alpha,k,\beta} Y_{\alpha k} \frac{\partial}{\partial y_{\alpha\beta}} (S^{+}XX^{\top}SS^{+})_{k\beta}$$

By Part (iv) of Lemma 2.1, for appropriate $A_1^{\alpha,k,\beta}, A_2^{\alpha,k,\beta}, \cdots, A_9^{\alpha,k,\beta}$ and for $l = \beta$, we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k} \frac{\partial}{\partial y_{\alpha\beta}} (S^{+}XX^{\top}SS^{+})_{k\beta} = \sum_{\alpha,k,\beta} Y_{\alpha k} (A_1^{\alpha,k,\beta} + A_2^{\alpha,k,\beta} + \cdots + A_9^{\alpha,k,\beta}). \tag{A.24}$$

Further, we have

$$\begin{aligned}
\sum_{\alpha,k,\beta} Y_{\alpha k} A_1^{\alpha,k,\beta} &= -\sum_{\alpha,k,\beta} Y_{\alpha k} S^{+}_{k\beta} (YS^{+}XX^{\top}SS^{+})_{\alpha\beta} \\
&= -\sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (YS^{+}XX^{\top}SS^{+})_{\alpha\beta} S^{+}_{\beta k} \\
&= -\sum_{\alpha,k} Y_{\alpha k} (YS^{+}XX^{\top}SS^{+}S^{+})_{\alpha k}.
\end{aligned}$$

Then,

$$\begin{aligned}
\sum_{\alpha,k,\beta} Y_{\alpha k} A_1^{\alpha,k,\beta} &= -\sum_{\alpha,k} (YS^{+}XX^{\top}S^{+})_{\alpha k} Y^{T}_{k\alpha} = -\sum_{\alpha} (YS^{+}XX^{\top}S^{+}Y^{\top})_{\alpha\alpha} \\
&= -\mathrm{tr}(YS^{+}XX^{\top}S^{+}Y^{\top}) = -\mathrm{tr}(X^{\top}S^{+}Y^{\top}YS^{+}X).
\end{aligned}$$

Hence,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_1^{\alpha,k,\beta} = -\mathrm{tr}(X^{\top}S^{+}SS^{+}X) = -\mathrm{tr}(X^{\top}S^{+}X) = -F. \tag{A.25}$$

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_2^{\alpha,k,\beta} = -\sum_{\alpha,k,\beta} Y_{\alpha k} (S^{+}Y^{\top})_{k\alpha} (S^{+}XX^{\top}SS^{+})_{\beta\beta}$$

$$= -\sum_{\alpha,k} Y_{\alpha k} (S^{+}Y^{\top})_{k\alpha} \sum_{\beta} (S^{+}XX^{\top}SS^{+})_{\beta\beta} = -\sum_{\alpha,k} Y_{\alpha k} (S^{+}Y^{\top})_{k\alpha} \mathrm{tr}(S^{+}XX^{\top}SS^{+}).$$

Then,

$$\begin{aligned}\sum_{\alpha,k,\beta} Y_{\alpha k} A_2^{\alpha,k,\beta} &= -\mathrm{tr}(X^\top SS^+S^+X) \sum_{\alpha} (YS^+Y^\top)_{\alpha\alpha} \\ &= -\mathrm{tr}(X^\top S^+X)\mathrm{tr}(YS^+Y^\top) = -F\mathrm{tr}(S^+Y^\top Y) = -F\mathrm{tr}(S^+S). \qquad \text{(A.26)}\end{aligned}$$

Further,

$$\begin{aligned}\sum_{\alpha,k,\beta} Y_{\alpha k} A_3^{\alpha,k,\beta} &= \sum_{\alpha,k,\beta} Y_{\alpha k}(I - SS^+)_{k\beta}(YS^+SXX^\top SS^+)_{\alpha\beta} \\ &= \sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (YS^+SXX^\top SS^+)_{\alpha\beta}(I - SS^+)_{\beta k} \\ &= \sum_{\alpha,k} Y_{\alpha k}(YS^+SXX^\top SS^+(I - SS^+))_{\alpha k} = 0. \qquad \text{(A.27)}\end{aligned}$$

Similarly, we have

$$\begin{aligned}\sum_{\alpha,k,\beta} Y_{\alpha k} A_4^{\alpha,k,\beta} &= \sum_{\alpha,k,\beta} Y_{\alpha k}(S^+S^+Y^\top)_{k\alpha}((I - SS^+)XX^\top SS^+)_{\beta\beta} \\ &= \sum_{\alpha,k} Y_{\alpha k}(S^+S^+Y^\top)_{k\alpha} \sum_{\beta} ((I - SS^+)XX^\top SS^+)_{\beta\beta} \\ &= \sum_{\alpha,k} Y_{\alpha k}(S^+S^+Y^\top)_{k\alpha}\mathrm{tr}((I - SS^+)XX^\top SS^+) \\ &= \mathrm{tr}(SS^+(I - SS^+)XX^\top) \sum_{\alpha,k} Y_{\alpha k}(S^+S^+Y^\top)_{k\alpha} = 0. \qquad \text{(A.28)}\end{aligned}$$

We also have

$$\begin{aligned}\sum_{\alpha,k,\beta} Y_{\alpha k} A_5^{\alpha,k,\beta} &= \sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top)_{k\beta}(YS^+)_{\alpha\beta} = \sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (S^+XX^\top)_{k\beta}(S^+Y^\top)_{\beta\alpha} \\ &= \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{k\alpha} = \sum_{\alpha} (YS^+XX^\top S^+Y^\top)_{\alpha\alpha}.\end{aligned}$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_5^{\alpha,k,\beta} = \mathrm{tr}(YS^+XX^\top S^+Y^\top) = \mathrm{tr}(X^\top S^+Y^\top YS^+X)$$
$$= \mathrm{tr}(X^\top S^+SS^+X) = \mathrm{tr}(X^\top S^+X) = F. \quad \text{(A.29)}$$

Further,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_6^{\alpha,k,\beta} = \sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top Y^\top)_{k\alpha} S^+_{\beta\beta} = \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top Y^T)_{k\alpha} \sum_{\beta} S^+_{\beta\beta}$$
$$= \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top Y^T)_{k\alpha} tr(S^+) = \mathrm{tr}(S^+) \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top Y^\top)_{k\alpha}$$
$$= \mathrm{tr}(S^+) \sum_{\alpha} (YS^+XX^\top Y^\top)_{\alpha\alpha}.$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_6^{\alpha,k,\beta} = \mathrm{tr}(S^+)\mathrm{tr}(YS^+XX^\top Y^\top) = \mathrm{tr}(S^+)\mathrm{tr}(S^+XX^\top Y^\top Y)$$
$$= \mathrm{tr}(S^+)\mathrm{tr}(S^+XX^\top S). \quad \text{(A.30)}$$

We also have

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_7^{\alpha,k,\beta} = -\sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top SS^+)_{k\beta}(YS^+)_{\alpha\beta}$$
$$= -\sum_{\alpha,k} Y_{\alpha k} \sum_{\beta} (S^+XX^\top SS^+)_{k\beta}(S^+Y^\top)_{\beta\alpha} = -\sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top SS^+S^+Y^\top)_{k\alpha}.$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_7^{\alpha,k,\beta} = -\sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{k\alpha} = -\sum_{\alpha} (YS^+XX^\top S^+Y^\top)_{\alpha\alpha}.$$

Hence,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_7^{\alpha,k,\beta} = -\mathrm{tr}(YS^+XX^\top S^+Y^\top) = -\mathrm{tr}(X^\top S^+Y^\top YS^+X)$$
$$= -\mathrm{tr}(X^\top S^+SS^+X) = -\mathrm{tr}(X^\top S^+X) = -F. \tag{A.31}$$

Further, we have

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_8^{\alpha,k,\beta} = -\sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top SS^+Y^\top)_{k\alpha} S^+_{\beta\beta}$$
$$= -\sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top SS^+Y^\top)_{k\alpha} \sum_\beta S^+_{\beta\beta} = -\sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top SS^+Y^\top)_{k\alpha} tr(S^+)$$
$$= -\mathrm{tr}(S^+)\sum_\alpha (YS^+XX^\top SS^+Y^\top)_{\alpha\alpha} = -\mathrm{tr}(S^+)\mathrm{tr}(YS^+XX^\top SS^+Y^\top).$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_8^{\alpha,k,\beta} = -\mathrm{tr}(S^+)\mathrm{tr}(S^+XX^\top SS^+Y^\top Y) = -\mathrm{tr}(S^+)\mathrm{tr}(S^+XX^\top SS^+S)$$
$$= -\mathrm{tr}(S^+)\mathrm{tr}(S^+XX^\top S). \tag{A.32}$$

We also have,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_9^{\alpha,k,\beta} = \sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top S^+Y^T)_{k\alpha}(I - SS^+)_{\beta\beta}$$
$$= \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{k\alpha} \sum_\beta (I - SS^+)_{\beta\beta}$$
$$= \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{k\alpha}\mathrm{tr}(I - SS^+)$$
$$= \sum_{\alpha,k} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{k\alpha}(p - \mathrm{tr}(SS^+)).$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_9^{\alpha,k,\beta} = (p - \mathrm{tr}(SS^+)) \sum_{\alpha} (YS^+XX^\top S^+Y^\top)_{\alpha\alpha}$$
$$= (p - \mathrm{tr}(SS^+))\mathrm{tr}(YS^+XX^\top S^+Y^\top) = (p - \mathrm{tr}(SS^+))\mathrm{tr}(X^\top S^+Y^\top YS^+X)$$
$$= (p - \mathrm{tr}(SS^+))\mathrm{tr}(X^\top S^+SS^+X).$$

Therefore,

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_9^{\alpha,k,\beta} = (p - \mathrm{tr}(SS^+))\mathrm{tr}(X^\top S^+X) = (p - \mathrm{tr}(SS^+))F. \tag{A.33}$$

Therefore, by $(A.25), (A.26), \cdots, (A.33)$, we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_1^{\alpha,k,\beta} + \sum_{\alpha,k,\beta} Y_{\alpha k} A_5^{\alpha,k,\beta} = 0 \tag{A.34}$$
$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_6^{\alpha,k,\beta} + \sum_{\alpha,k,\beta} Y_{\alpha k} A_8^{\alpha,k,\beta} = 0 \tag{A.35}$$
$$\sum_{\alpha,k,\beta} Y_{\alpha k} A_3^{\alpha,k,\beta} = \sum_{\alpha,k,\beta} Y_{\alpha k} A_4^{\alpha,k,\beta} = 0. \tag{A.36}$$

Then, by replacing (A.34),(A.35) and (A.36) in (A.24) together with (A.26), (A.31) and (A.33), we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(A_1^{\alpha,k,\beta} + A_2^{\alpha,k,\beta} + \cdots + A_9^{\alpha,k,\beta}) = \sum_{\alpha,k,\beta} Y_{\alpha k} A_2^{\alpha,k,\beta} + \sum_{\alpha,k,\beta} Y_{\alpha k} A_7^{\alpha,k,\beta} + \sum_{\alpha,k,\beta} Y_{\alpha k} A_9^{\alpha,k,\beta}.$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(A_1^{\alpha,k,\beta} + A_2^{\alpha,k,\beta} + \cdots + A_9^{\alpha,k,\beta}) = F(p - 2\mathrm{tr}(SS^+) - 1).$$

(iv) $\displaystyle\sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = \frac{2r(F)r'(F)}{F^2} \sum_{\alpha,\beta,k} Y_{\alpha k} (\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k}$

$$- \frac{2r^2(F)}{F^3} \sum_{\alpha,\beta,k} Y_{\alpha k} (\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+ XX^\top S^+)_{\beta k}$$
$$+ \frac{r^2(F)}{F^2} \sum_{\alpha,k,\beta} Y_{\alpha k} \frac{\partial}{\partial y_{\alpha\beta}} (SS^+ XX^\top S^+)_{\beta k}$$
$$= \frac{2r(F)r'(F)}{F^2} \left(-2F^2\right) - \frac{2r^2(F)}{F^3} \left(-2F^2\right) + \frac{r^2(F)}{F^2} F(p - 2\mathrm{tr}(SS^+) - 1).$$

Then,

$$\sum_{\alpha,\beta,k} Y_{\alpha k} \frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = -4r(F)r'(F) + \frac{4r^2(F)}{F} + \frac{r^2(F)}{F}(p - 2\mathrm{tr}(SS^+) - 1)$$
$$= -4r(F)r'(F) + \frac{r^2(F)}{F} \left(p - 2\mathrm{tr}(SS^+) + 3\right),$$

which completes the proof. □

Proof of Lemma 2.3.

$$(i) \quad \frac{\partial F}{\partial X_i} = \frac{\partial}{\partial X_i}(X^\top S^+ X) = \frac{\partial}{\partial X_i} \sum_k X^\top_{1k} (S^+ X)_{k1} = \sum_k \frac{\partial}{\partial X_i} (X_{k1} (S^+ X)_{k1}).$$

Then,

$$\frac{\partial F}{\partial X_i} = \sum_k \left\{ (\frac{\partial X_{k1}}{\partial X_i})(S^+ X)_{k1} + X_{k1} (\frac{\partial S^+ X}{\partial X_i})_{k1} \right\}$$
$$= \sum_k \frac{\partial X_{k1}}{\partial X_i} (S^+ X)_{k1} + \sum_k X_{k1} (\frac{\partial S^+ X}{\partial X_i})_{k1}$$
$$= \sum_k \delta_{ki} (S^+ X)_{k1} + \sum_k X_{k1} (\frac{\partial}{\partial X_i} \sum_l S^+_{kl} X_{l1}).$$

This gives

$$\frac{\partial F}{\partial X_i} = (S^+ X)_{i1} + \sum_k X_{k1} (\sum_l S^+_{kl} \frac{\partial X_{l1}}{\partial X_i}) = (S^+ X)_{i1} + \sum_k X_{k1} (\sum_l S^+_{kl} \delta_{li}).$$

Then,

$$\frac{\partial F}{\partial X_i} = (S^+X)_{i1} + \sum_k X_{k1}S^+_{ki} = (S^+X)_{i1} + \sum_k S^+_{ik}X_{k1}.$$

Hence,

$$\frac{\partial F}{\partial X_i} = (S^+X)_{i1} + (S^+X)_{i1} = 2(S^+X)_{i1}.$$

(ii) $$\left(\frac{\partial SS^+X}{\partial X_i}\right)_k = \frac{\partial}{\partial X_i}\sum_\alpha (SS^+)_{k\alpha}X_{\alpha 1} = \sum_\alpha (SS^+)_{k\alpha}\frac{\partial X_\alpha}{\partial X_i} = \sum_\alpha (SS^+)_{k\alpha}\delta_{\alpha i}.$$

Then,

$$\frac{\partial F}{\partial X_i} = (S^+S^+)_{ki}.$$

(iii) By Part (i) and (ii) ,we get

$$\begin{aligned}
&\frac{\partial g_k}{\partial X_i} = (\frac{\partial}{\partial X_i}\frac{r(F)}{F})(SS^+X)_k + \frac{r(F)}{F}(\frac{\partial}{\partial X_i}(SS^+X)_k) \\
&= \frac{r'F - r(F)}{F^2}(\frac{\partial F}{\partial X_i})(SS^+X)_k + \frac{r(F)}{F}(SS^+)_{ki} \\
&= \frac{2(Fr'(F) - r(F))}{F^2}(S^+X)_i(SS^+X)_k + \frac{r(F)}{F}(SS^+)_{ki}.
\end{aligned}$$

(iv) By Part (iii) for $k = i$, we have

$$\begin{aligned}
&\sum_i \frac{\partial g_i}{\partial X_i} = \sum_i \left\{\frac{2Fr'(F) - r(F)}{F^2}(S^+X)_i(SS^+X)_i + \frac{r(F)}{F}(SS^+)_{ii}\right\} \\
&= 2\frac{Fr'(F) - r(F)}{F^2}\sum_i (S^+X)_i(X^\top S^+)_i + \frac{r(F)}{F}\mathrm{tr}(SS^+) \\
&= 2\frac{Fr'(F) - r(F)}{F^2}\sum_i (SS^+XX^\top S^+)_{ii} + \frac{r(F)}{F}\mathrm{tr}(SS^+).
\end{aligned}$$

This gives

$$\sum_i \frac{\partial g_i}{\partial X_i} = 2\frac{Fr'(F) - r(F)}{F^2}\text{tr}(SS^+XX^\top S^+) + \frac{r(F)}{F}\text{tr}(SS^+)$$
$$= 2\frac{Fr'(F) - r(F)}{F^2}\text{tr}(X^\top S^+SS^+X) + \frac{r(F)}{F}\text{tr}(SS^+).$$

Then,

$$\sum_i \frac{\partial g_i}{\partial X_i} = 2\frac{Fr'(F) - r(F)}{F^2}\text{tr}(X^\top S^+X) + \frac{r(F)}{F}\text{tr}(SS^+) = 2\frac{Fr'(F) - r(F)}{F^2}F + \frac{r(F)}{F}\text{tr}(SS^+).$$

Hence,

$$\sum_i \frac{\partial g_i}{\partial X_i} = 2r'(F) - 2\frac{r(F)}{F} + \frac{r(F)}{F}\text{tr}(SS^+) = 2r'(F) + \frac{r(F)}{F}(\text{tr}(SS^+) - 2),$$

which completes the proof. □

A.4 Some Technical Proofs of Chapter 3

Proof of Lemma 3.1.

$$(i) \quad \left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \frac{\partial}{\partial Y_{\alpha\beta}}\sum_i Y_{ki}^\top Y_{il} = \sum_i \frac{\partial}{\partial Y_{\alpha\beta}}(Y_{ik}Y_{il}) = \sum_i \left[(\frac{\partial Y_{ik}}{\partial Y_{\alpha\beta}})Y_{il} + Y_{qk}(\frac{\partial Y_{il}}{\partial Y_{\alpha\beta}})\right].$$

Then,

$$\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \sum_i (\frac{\partial Y_{ik}}{\partial Y_{\alpha\beta}})Y_{il} + \sum_i Y_{ik}(\frac{\partial Y_{il}}{\partial Y_{\alpha\beta}}) = \sum_i \delta_{\beta k}Y_{il} + \sum_i Y_{ik}\delta_{\beta l}.$$

Therefore,

$$\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{kl} = \delta_{\beta k}Y_{\alpha l} + \delta_{\beta l}Y_{\alpha k}.$$

(ii) By Part (i), we get

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = \sum_j (A\frac{\partial S}{\partial Y_{\alpha\beta}})_{kj}B_{jl} = \sum_j \left(\sum_i A_{ki}\left(\frac{\partial S}{\partial Y_{\alpha\beta}}\right)_{ij}\right) B_{jl}.$$

Then,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = \sum_j \left(\sum_i A_{ki}\left\{\delta_{\beta i}Y_{\alpha j} + \delta_{\beta j}Y_{\alpha i}\right\}\right) B_{jl} = \sum_j \left(\sum_i A_{ki}\delta_{\beta i}Y_{\alpha j} + \sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right) B_{jl}.$$

Then,

$$\begin{aligned}\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} &= \sum_j \left(A_{k\beta}Y_{\alpha j} + \sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right) B_{jl} \\ &= \sum_j A_{k\beta}Y_{\alpha j}B_{jl} + \sum_j \left(\sum_i A_{ki}\delta_{\beta j}Y_{\alpha i}\right) B_{jl} \\ &= A_{k\beta}\sum_j Y_{\alpha j}B_{jl} + \sum_i A_{ki}Y_{\alpha i}\sum_j \delta_{\beta j}B_{jl}.\end{aligned}$$

Hence,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + \sum_i A_{ki}Y_{\alpha i}B_{\beta l} = A_{k\beta}(YB)_{\alpha l} + B_{\beta l}\sum_i^p A_{ki}Y_{i\alpha}^\top.$$

Hence,

$$\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl} = A_{k\beta}(YB)_{\alpha l} + B_{\beta l}(AY^\top)_{k\alpha} = A_{k\beta}(YB)_{\alpha l} + (AY^\top)_{k\alpha}B_{\beta l}.$$

(iii) $$\left(\frac{\partial}{\partial Y_{\alpha\beta}}(X^\top S^+ X)\right)_{kk} = \left(X^\top(\frac{\partial S^+}{\partial Y_{\alpha\beta}})X\right)_{kk}.$$

From Proposition A.2, we get

$$\left(X^\top(\frac{\partial S^+}{\partial Y_{\alpha\beta}})X\right)_{kk}$$

$$= \left(X^\top \left(- S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ + (I - SS^+) \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ S^+ + S^+ S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} (I - SS^+) \right) X \right)_{kk}$$
$$= \left(-X^\top S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ X + X^\top (I - SS^+) \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ S^+ X + X^\top S^+ S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} (I - SS^+) X \right)_{kk}$$
$$= - \left(X^\top S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ X \right)_{kk} + \left(X^\top (I - SS^+) \frac{\partial S}{\partial Y_{\alpha\beta}} S^+ S^+ X \right)_{kk} + \left(X^\top S^+ S^+ \frac{\partial S}{\partial Y_{\alpha\beta}} (I - SS^+) X \right)_{kk}.$$

Now by Part (ii) for $l = k$, we get

$$\left(\frac{\partial X^\top S^+ X}{\partial Y_{\alpha\beta}} \right)_{kk} = -(X^\top S^+)_{k\beta} (Y S^+ X)_{\alpha k} - (X^\top S^+ Y^\top)_{k\alpha} (S^+ X)_{\beta k}$$
$$+ (X^\top (I - SS^+))_{k\beta} (Y SS^+ X)_{\alpha k} + (X^\top (I - SS^+) Y^\top)_{k\alpha} (SS^+ X)_{\beta k}$$
$$+ (X^\top S^+ S^+)_{k\beta} (Y (I - SS^+) X)_{\alpha k} + (X^\top S^+ S^+ Y^\top)_{k\alpha} ((I - SS^+) X)_{\beta k}.$$

Since

$$(X^\top S^+)_{k\beta} (Y S^+ X)_{\alpha k} = (X^\top S^+ Y^\top)_{k\alpha} (S^+ X)_{\beta k},$$
$$(X^\top (I - SS^+))_{k\beta} (Y SS^+ X)_{\alpha k} = (X^\top S^+ S^+ Y^\top)_{k\alpha} ((I - SS^+) X)_{\beta k} \text{ and}$$
$$Y (I - SS^+) = (I - SS^+) Y^\top = 0,$$

then,

$$\left(\frac{\partial X^\top S^+ X}{\partial Y_{\alpha\beta}} \right)_{kk} = -2 (X^\top S^+ Y^\top)_{k\alpha} (S^+ X)_{\beta k} + 2 (X^\top S^+ S^+ Y^\top)_{k\alpha} ((I - SS^+) X)_{\beta k}.$$

(iv) By Part (ii) we get

$$\frac{\partial F}{\partial Y_{\alpha\beta}} = \frac{\partial}{\partial Y_{\alpha\beta}} \sum_k (X^\top S^+ X)_{kk} = \sum_k \frac{\partial (X^\top S^+ X)_{kk}}{\partial Y_{\alpha\beta}}$$
$$= \sum_k \left\{ -2 (X^\top S^+ Y^\top)_{k\alpha} (S^+ X)_{\beta k} + 2 (X^\top S^+ S^+ Y^\top)_{k\alpha} ((I - SS^+) X)_{\beta k} \right\}$$

$$= -2 \left(S^+ X X^\top S^+ Y^\top \right)_{\beta\alpha} + 2 \left((I - SS^+) X X^\top S^+ S^+ Y^\top \right)_{\beta\alpha}.$$

(v) $\left(\dfrac{\partial S^+XX^\top SS^+}{\partial Y_{\alpha\beta}}\right)_{kl}$

$$= \left(\frac{\partial S^+}{\partial Y_{\alpha\beta}}XX^\top SS^+\right)_{kl} + \left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} + \left(S^+XX^\top S\frac{\partial S^+}{\partial Y_{\alpha\beta}}\right)_{kl}$$

$$= \left((-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+ + (I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+ + S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+))XX^\top SS^+\right)_{kl}$$

$$+ \left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl}$$

$$+ \left(S^+XX^\top S(-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+ + (I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+ + S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+))\right)_{kl}.$$

Then,

$$\begin{aligned}\left(\frac{\partial S^+XX^\top SS^+}{\partial Y_{\alpha\beta}}\right)_{kl} &= \left(-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+XX^\top SS^+\right)_{kl} + \left((I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+XX^\top SS^+\right)_{kl}\\ &+ \left(S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)XX^\top SS^+\right)_{kl} + \left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl}\\ &- \left(S^+XX^\top SS^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} + \left(S^+XX^\top S(I-SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+\right)_{kl}\\ &+ \left(S^+XX^\top SS^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)\right)_{kl}.\end{aligned}$$

Now, by using Part (ii), we get

$$\left(-S^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+XX^\top SS^+\right)_{kl} = -S^+_{k\beta}(YS^+XX^\top SS^+)_{\alpha l} - (S^+Y^\top)_{k\alpha}(S^+XX^\top SS^+)_{\beta l}. \tag{A.37}$$

Further,

$$\left((I - SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}} S^+S^+XX^\top SS^+\right)_{kl}$$
$$= (I - SS^+)_{k\beta}(YS^+S^+XX^\top SS^+)_{\alpha l} + ((I - SS^+)Y^\top)_{k\alpha}(S^+S^+XX^\top SS^+)_{\beta l}$$
$$= (I - SS^+)_{k\beta}(YS^+S^+XX^\top SS^+)_{\alpha l}. \tag{A.38}$$

Since $Y(I - SS^+) = 0$ (see Corollary A.1), we get

$$\left(S^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I - SS^+)XX^\top SS^+\right)_{kl}$$
$$= (S^+S^+)_{k\beta}(Y(I - SS^+)XX^\top SS^+)_{\alpha l} + (S^+S^+Y^\top)_{k\alpha}((I - SS^+)XX^\top SS^+)_{\beta l}$$
$$= (S^+S^+Y^\top)_{k\alpha}((I - SS^+)XX^\top SS^+)_{\beta l}. \tag{A.39}$$

We also have

$$\left(S^+XX^\top\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} = (S^+XX^\top)_{k\beta}(YS^+)_{\alpha l} + (S^+XX^\top Y^\top)_{k\alpha}S^+_{\beta l}. \tag{A.40}$$

Further,

$$\left(-S^+XX^\top SS^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} = -(S^+XX^\top SS^+)_{k\beta}(YS^+)_{\alpha l}(S^+XX^\top SS^+Y^\top)_{k\alpha}S^+_{\beta l}.$$

Since $SS^+Y^\top = Y^\top$ (see Corollary A.1), we get

$$\left(-S^+XX^\top SS^+\frac{\partial S}{\partial Y_{\alpha\beta}}S^+\right)_{kl} = -(S^+XX^\top SS^+)_{k\beta}(YS^+)_{\alpha l}(S^+XX^\top Y^\top)_{k\alpha}S^+_{\beta l}. \tag{A.41}$$

Since $S(I - SS^+) = S^+(I - SS^+) = 0$, we get

$$\left(S^+XX^\top S(I - SS^+)\frac{\partial S}{\partial Y_{\alpha\beta}}S^+S^+\right)_{kl} = 0. \tag{A.42}$$

Further,

$$\left(S^+XX^\top SS^+S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)\right)_{kl}=\left(S^+XX^\top S^+\frac{\partial S}{\partial Y_{\alpha\beta}}(I-SS^+)\right)_{kl}$$
$$=\left(A\frac{\partial S}{\partial Y_{\alpha\beta}}B\right)_{kl}=A_{k\beta}(YB)_{\alpha l}+(AY^\top)_{k\alpha}B_{\beta l}$$
$$=(S^+XX^\top S^+)_{k\beta}(Y(I-SS^+))_{\alpha l}+(S^+XX^\top S^+Y^\top)_{k\alpha}(I-SS^+)_{\beta l}$$
$$=(S^+XX^\top S^+Y^\top)_{k\alpha}(I-SS^+)_{\beta l}. \tag{A.43}$$

Therefore by $(A.37),(A.38),\cdots,(A.43)$, we get

$$\left(\frac{\partial S^+XX^\top SS^+}{\partial Y_{\alpha\beta}}\right)_{kl}=-S^+_{k\beta}(YS^+XX^\top SS^+)_{\alpha l}-(S^+Y^T)_{k\alpha}(S^+XX^\top SS^+)_{\beta l}$$
$$+(I-SS^+)_{k\beta}(YS^+SXX^\top SS^+)_{\alpha l}+(S^+S^+Y^\top)_{k\alpha}((I-SS^+)XX^\top SS^+)_{\beta l}$$
$$+(S^+XX^\top)_{k\beta}(YS^+)_{\alpha l}+(S^+XX^\top Y^\top)_{k\alpha}S^+_{\beta l}-(S^+XX^\top SS^+)_{k\beta}(YS^+)_{\alpha l}$$
$$-(S^+XX^\top SS^+Y^\top)_{k\alpha}S^+_{\beta l}+(S^+XX^\top S^+Y^\top)_{k\alpha}(I-SS^+)_{\beta l},$$

which completes the proof. □

Proof of Lemma 3.2.

$$(i)\quad \frac{\partial G_{kl}}{\partial Y_{\alpha\beta}}=\frac{\partial}{\partial Y_{\alpha\beta}}\left\{\frac{r^2(F)}{F^2}(S^+XX^\top SS^+)_{kl}\right\}=\frac{\partial}{\partial Y_{\alpha\beta}}(\frac{r^2(F)}{F^2})(S^+XX^\top SS^+)_{kl}$$
$$+\frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl}.$$

Then,

$$\frac{\partial G_{kl}}{\partial Y_{\alpha\beta}}=\frac{2r(F)r'(F)}{F^2}(\frac{\partial F}{\partial y_{\alpha\beta}})(S^+XX^\top SS^+)_{kl}-\frac{2r^2(F)}{F^3}(\frac{\partial F}{\partial y_{\alpha\beta}})(S^+XX^\top SS^+)_{kl}$$
$$+\frac{r^2(F)}{F^2}\frac{\partial}{\partial Y_{\alpha\beta}}(S^+XX^\top SS^+)_{kl}.$$

(ii) By Part (iii) of Lemma 3.1, we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k}$$

$$= -2\sum_{\alpha,k,\beta} Y_{\alpha k}(S^+XX^\top S^+Y^\top)_{\beta\alpha}(SS^+XX^\top S^+)_{\beta k}$$

$$+ 2\sum_{\alpha,k,\beta} Y_{\alpha k}((I-SS^+)XX^\top S^+S^+Y^\top)_{\beta\alpha}(SS^+XX^\top S^+)_{\beta k}$$

$$= -2\sum_{\alpha,k} Y_{\alpha k}\sum_{\beta}(YS^+XX^\top S^+)_{\alpha\beta}(SS^+XX^T S^+)_{\beta k}$$

$$+ 2\sum_{\alpha,k} Y_{\alpha k}\sum_{\beta}(YS^+S^+XX^\top(I-SS^+))_{\alpha\beta}(SS^+XX^\top S^+)_{\beta k}$$

$$= -2\sum_{\alpha,k} Y_{\alpha k}(YS^+XX^\top S^+SS^+XX^\top S^+)_{\alpha k}$$

$$+ 2\sum_{\alpha,k} Y_{\alpha k}(YS^+S^+XX^\top(I-SS^+)SS^+XX^\top S^+)_{\alpha k}.$$

Since $(I-SS^+)SS^+=0$ then we get

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k} = -2\sum_{\alpha,k} Y^\top_{k\alpha}(YS^+XX^\top S^+XX^\top S^+)_{\alpha k}$$

$$= -2\sum_{k}(Y^\top YS^+XX^\top S^+XX^\top S^+)_{kk} = -2\sum_{k}(SS^+XX^\top S^+XX^\top S^+)_{kk}$$

$$= -2\mathrm{tr}(SS^+XX^\top S^+XX^\top S^+) = -2\mathrm{tr}(X^\top S^+XX^\top S^+SS^+X).$$

Then,

$$\sum_{\alpha,k,\beta} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k} = -2\mathrm{tr}(X^\top S^+XX^\top S^+X) = -2\mathrm{tr}((X^\top S^+X)^2).$$

(iii) Similar to the proof of Part (iii) of Lemma 2.2.

$$(iv)\quad \sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = \frac{2r(F)r'(F)}{F^2}\sum_{\alpha,\beta,k} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k}$$
$$-\frac{2r^2(F)}{F^3}\sum_{\alpha,\beta,k} Y_{\alpha k}(\frac{\partial F}{\partial y_{\alpha\beta}})(SS^+XX^\top S^+)_{\beta k} + \frac{r^2(F)}{F^2}\sum_{\alpha,k,\beta} Y_{\alpha k}\frac{\partial}{\partial y_{\alpha\beta}}(SS^+XX^\top S^+)_{\beta k}$$
$$= \frac{2r(F)r'(F)}{F^2}\left(-2\mathrm{tr}\left((X^\top S^+X)^2\right)\right) - \frac{2r^2(F)}{F^3}\left(-2\mathrm{tr}\left((X^\top S^+X)^2\right)\right)$$
$$+\frac{r^2(F)}{F^2}F(p-2\mathrm{tr}(SS^+)-1).$$

Then,

$$\sum_{\alpha,\beta,k} Y_{\alpha k}\frac{\partial G_{k\beta}}{\partial Y_{\alpha\beta}} = -\frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+X)^2\right) + \frac{4r^2(F)\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^3}$$
$$+\frac{r^2(F)}{F}(p-2\mathrm{tr}(SS^+)-1)$$
$$= -\frac{4r(F)r'(F)}{F^2}\mathrm{tr}\left((X^\top S^+X)^2\right)$$
$$+\frac{r^2(F)}{F}\left(\frac{4\mathrm{tr}\left((X^\top S^+X)^2\right)}{F^2} + p - 2\mathrm{tr}(SS^+) - 1\right),$$

which completes the proof. □

Proof of Lemma 3.3.

$$(i)\quad \frac{\partial F}{\partial X_{ij}} = \frac{\partial}{\partial X_{ij}}\sum_k (X^\top S^+X)_{kk} = \frac{\partial}{\partial X_{ij}}\sum_{k,\alpha,\beta} X^\top_{k\alpha}S^+_{\alpha\beta}X_{\beta k}$$
$$= \sum_{k,\alpha,\beta}\left(\frac{\partial}{\partial X_{ij}}X^T_{k\alpha}\right)S^+_{\alpha\beta}X_{\beta k} + \sum_{k,\alpha,\beta}X^\top_{k\alpha}S^+_{\alpha\beta}\left(\frac{\partial}{\partial X_{ij}}X_{\beta k}\right)$$
$$= \sum_\beta S^+_{i\beta}X_{\beta j} + \sum_\alpha X^\top_{j\alpha}S^+_{\alpha i}$$
$$= (S^+X)_{ij} + (X^\top S^+)_{ji} = (S^+X)_{ij} + (S^+X)_{ij} = 2(S^+X)_{ij}.$$

(ii) $$\left(\frac{\partial SS^+X}{\partial X_{ij}}\right)_{kl} = \frac{\partial}{\partial X_{ij}}\sum_\alpha (SS^+)_{k\alpha}X_{\alpha l} = \sum_\alpha (SS^+)_{k\alpha}\frac{\partial X_{\alpha l}}{\partial X_{ij}} = \sum_\alpha (SS^+)_{k\alpha}\delta_{\alpha i}\delta_{lj}.$$

Then,

$$\left(\frac{\partial SS^+X}{\partial X_{ij}}\right)_{kl} = (SS^+)_{ki}\delta_{lj}.$$

(iii) By Parts (i) and (ii), we get

$$\begin{aligned}
\frac{\partial g_{kl}}{\partial X_{i,j}} &= \left(\frac{\partial}{\partial X_{ij}}\frac{r(F)}{F}\right)(SS^+X)_{kl} + \frac{r(F)}{F}\left(\frac{\partial}{\partial X_{ij}}(SS^+X)_{kl}\right)\\
&= \frac{r'(F)F - r(F)}{F^2}\left(\frac{\partial F}{\partial X_{ij}}\right)(SS^+X)_{kl} + \frac{r(F)}{F}\left(\frac{\partial}{\partial X_{ij}}(SS^+X)_{kl}\right)\\
&= \frac{2(Fr'(F) - r(F))}{F^2}(S^+X)_{ij}(SS^+X)_{kl} + \frac{r(F)}{F}(SS^+)_{ki}\delta_{lj}.
\end{aligned}$$

(iv) By Part (iii), we have

$$\begin{aligned}
\sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}} &= \sum_{i,j}\left\{\frac{2Fr'(F) - r(F)}{F^2}(S^+X)_{ij}(SS^+X)_{ij} + \frac{r(F)}{F}(SS^+)_{ii}\right\}\\
&= 2\frac{Fr'(F) - r(F)}{\mathrm{tr}^2(F)}\sum_{i,j}(S^+X)_{ij}(X^\top S^+)_{ji} + \frac{r(F)}{F}\mathrm{tr}(SS^+)\\
&= 2\frac{Fr'(F) - r(F)}{F^2}\sum_i (SS^+XX^\top S^+)_{ii} + q\frac{r(F)}{F}\mathrm{tr}(SS^+)\\
&= 2\frac{Fr'(F) - r(F)}{F^2}\mathrm{tr}(SS^+XX^\top S^+) + q\frac{r(F)}{F}\mathrm{tr}(SS^+).
\end{aligned}$$

This gives

$$\sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}} = 2\frac{Fr'(F) - r(F)}{F^2}\mathrm{tr}(X^\top S^+SS^+X) + q\frac{r(F)}{F}\mathrm{tr}(SS^+)$$

$$= 2\frac{Fr'(F) - r(F)}{F^2}\mathrm{tr}(X^\top S^+ X) + q\frac{r(F)}{F}\mathrm{tr}(SS^+).$$

Hence,

$$\sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}} = 2\frac{Fr'(F) - r(F)}{F^2}F + q\frac{r(F)}{F}\mathrm{tr}(SS^+) = 2r'(F) - 2\frac{r(F)}{F} + q\frac{r(F)}{F}\mathrm{tr}(SS^+).$$

Therefore,

$$\sum_{i,j}\frac{\partial g_{ij}}{\partial X_{ij}} = 2r'(F) + \frac{r(F)}{F}(q\mathrm{tr}(SS^+) - 2),$$

which completes the proof. □

Appendix B

R code

```
####Important Libraries

library(MASS)

library(corpcor) # to calculate Moore - Penrose inverse

library(matrixsampling) # to simulate a Random matrix normal

library(ggplot2)

library(tidyr)

library(dplyr)

library(gridExtra)


########################################

##Defining Trace Function

########################################

trace = function(x) {

  dim(x)[1]

  if (is.null(dim(x)[1])==TRUE){

    return(x)

  }
```

```
  else{
    tr = 0
    for (i in 1:dim(x)[1]) {
      s = x[i,i]
      tr = tr + s
    }
    return(tr[[1]])
  }
}
#############################################
##Defining function r
#############################################
r = function(f){
  library(psych)
  a = 1/(1+exp(-trace(f)))
  return(a)
}
########################################
##Defining our proposed estimator
########################################
JS_est <- function(x, sigma) {
  f = trace(t(x)%*%pseudoinverse(sigma)%*%x)
  est = x - r(f)*sigma%*%pseudoinverse(sigma)%*%x/f
  return(est)
}

set.seed(13144)
g = 0
```

```
q = 3
e = 0
#to store norm of theta
ntheta = c()
#to store all mean risk difference
df_p = matrix(0,nrow = 4,ncol = 44)
#to store the risk of usual estimator
R_L_n = matrix(0,nrow = 30,ncol = 4)
#to store the risk of proposed estimator
R_J_n = matrix(0,nrow = 30,ncol = 4)
#Starting simulation:
for(p in c(24,32,56,104)){
  #Defining the covariance matrix for each choice of p
  cov_matrix = diag(p)
  # to store mean of risk difference after each 10 repetitions
  md = c()
  #to create 11 different theta for each choice of p:
  for(l in seq(0,10,1)){
    e = e+1
    k= 0
    g=g+1
    theta = matrix(l, nrow = p, ncol = q)
    ntheta[e] = norm(theta, type = "F")
    #different sample sizes for each choice of p
    for(n in c(p/8,p/4,p-1,2*p)){
      k = k+1
      d = c()
      R_L = c()
```

```
      R_J = c()
      #30 Repetitions for each sample size
      for(i in 1:30){
        Z = rmatrixnormal(n, M = theta,U = diag(p),
                          V =  diag(q),keep = FALSE)
        s = matrix(0, nrow = p, ncol = q)
        for (h in 1:n){
          s = s + Z[,,h]
        }
        X = s/n
        Q = matrix(0, nrow = p, ncol = p)
        for(h in 1:n){
          Q = Q+(Z[,,h]-X)%*%t(Z[,,h]-X)
        }
        S = Q/n
        theta_J= JS_est(X,S)
        R_L[i] = trace(t(X-theta)%*%cov_matrix%*%(X-theta))
        R_J[i] = trace(t(theta_J - theta)%*%
                          cov_matrix%*%(theta_J - theta))
        d[i] = R_J[i] - R_L[i]
      }
      #storing the mean of risk differences in each 30 repetitions
      md[k] = mean(d)
    }
    df_p[,g]=md
  }
}
df_p = data.frame(df_p)
```

```
colnames(df_p) = rep(c("24","32","56","104"),11)
df_long_p = as.data.frame(df_p%>%pivot_longer(cols = everything(),
                                              names_to = "P")%>%
  mutate(theta_norm = rep(ntheta,4),"P" = as.numeric(P)) %>%
  mutate("Sample size" = rep(c("p/8","p/4","p-1","2p"), each = 44),
                     P = replace(P, P == 24, "p=24"),
                     P = replace(P, P == 32, "p=32"),
                     P = replace(P, P == 56, "p=56"),
                     P = replace(P, P == 104, "p=104")))
ggplot(df_long_p,aes(x = theta_norm, y = value))+
  geom_line(size = 1,aes(linetype= `Sample size`,
                         color = `Sample size`))+
 facet_wrap(~factor(P,levels = c("p=24","p=32","p=56","p=104")),
           scales="free")+
  theme_bw()+
     labs(x = expression(paste('||', theta,'||')),
          y = "Risk diffference")
```

Bibliography

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist. 41*, 642–645.

Chételat, D. and M. T. Wells (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n. *The Annals of Statistics 40*(6), 3137–3160.

Golub, G. H. and V. Pereyra (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal. 10*, 413–432.

Srivastava, M. (2003). Singular Wishart and multivariate beta distributions. *The Annals of Statistics 31*(5), 1537 – 1560.

Srivastava, M. S. and C. G. Khatri (1979). *An introduction to multivariate statistics*. North-Holland, New York-Oxford.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Volume 3, pp. 197–207. University of California Press.

Stein, C. (1960). Estimation with quadratic loss.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist. 9*(6), 1135–1151.

Vita Auctoris

Mr. Arash Aghaei Foroushani was born in 1998, Tehran, Iran. He graduated from Isfahan University of Technology in 2021 with a Bachelor of Science degree in Statistics. For a peruse of continuous education, he came to Canada. He is currently a candidate for the Master of Science degree in Statistics at the University of Windsor.