



HAL
open science

Ledoit-Wolf linear shrinkage with unknown mean

Benoit Oriol, Alexandre Miot

► **To cite this version:**

| Benoit Oriol, Alexandre Miot. Ledoit-Wolf linear shrinkage with unknown mean. 2023. hal-04071051

HAL Id: hal-04071051

<https://hal.science/hal-04071051>

Preprint submitted on 17 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ledoit-Wolf linear shrinkage with unknown mean

Benoît Oriol *

CEREMADE, Université Paris-Dauphine, PSL, Paris, France
Société Générale Corporate and Investment Banking, Puteaux, France

Alexandre Miot

Société Générale Corporate and Investment Banking, Puteaux, France

14/04/2023

Abstract

This work addresses large dimensional covariance matrix estimation with unknown mean. The empirical covariance estimator fails when dimension and number of samples are proportional and tend to infinity, settings known as Kolmogorov asymptotics. When the mean is known, Ledoit and Wolf (2004) proposed a linear shrinkage estimator and proved its convergence under those asymptotics. To the best of our knowledge, no formal proof has been proposed when the mean is unknown. To address this issue, we propose a new estimator and prove its quadratic convergence under the Ledoit and Wolf assumptions. Finally, we show empirically that it outperforms other standard estimators.

Keywords: covariance matrix estimation, linear shrinkage, Ledoit-Wolf estimator, unknown mean, general asymptotics

Declarations of interest: none

1 Introduction and related work

The covariance matrix plays a major role in numerous machine learning algorithms and statistics. Just to cite a few, the PCA [1] in machine learning, Markowitz portfolio management [2] in finance, or generalized method of moments estimators [3] in statistics. However, those algorithms are designed to use the true covariance matrix, which is often inaccessible. Even if the sample covariance matrix seems to be a simple and appealing choice, it severely fails in many applications: for instance, the use of the sample covariance matrix for Markowitz portfolio management does not beat a naive uniform distribution among the assets [4].

In the context of Kolmogorov asymptotics, where the ratio of the dimension p_n and the number of samples n tends to a finite positive constant $\frac{p_n}{n} \rightarrow c > 0$, this estimator fails to converge quadratically. Moreover, its eigenvalue spectrum is biased: high eigenvalues tends at

*Corresponding author: benoit.oriol@dauphine.eu

B218, CEREMADE, Université Paris-Dauphine, PSL, place du Marechal de Lattre de Tassigny, 75116 Paris, France

being too high, and low ones, too low. The behavior of the eigenvalues is studied in random matrix theory: in the context of the Kolmogorov asymptotics, this topic is widely covered by V. L. Girko [5, 6, 7].

We focus on the shrinkage-type estimators which have suitable asymptotic properties, influenced by the work of Stein on Gaussian mean estimation in 1956 [8]. Due to their simplicity to implement and strong theoretical support, linear methods are widely used, and, for some, implemented in ScikitLearn [9]: Ledoit-Wolf linear shrinkage [10], which will be our main focus, its extension for Gaussian distributions using Rao-Blackwell theorem, named Oracle Approximating Shrinkage (OAS) estimator [11], linear shrinkage with factor models [12], linear shrinkage for elliptical distributions with unknown mean and known radius distribution [13], just to name a few. Non-linear methods propose shrinkage methods where the factor differs from an eigenvalue to an other. Among them, Stein’s covariance estimator [14] works for Gaussian distributions, and several algorithms were developed by Ledoit and Wolf using eigenvalue spectrum analysis from random matrix theory [15, 16, 17]. Further theoretical analysis of those algorithms can be found in [18, 19, 20].

Usually, when estimating the covariance matrix, we don’t know the mean of the distribution. Yet, the extension from known to unknown mean is rarely studied. To extend the empirical covariance with T samples S_T , one uses the unbiased estimator \tilde{S}_T , after removing the empirical mean and dividing by $T - 1$ instead of T . If it seems straightforward for S_T , it can be non trivial for more complex estimators. Ahurbekova explicitly worked in the case of known then unknown mean [13], and the resulting estimators of linear shrinkage for elliptical distributions with known radius distributions are notably different from the ones with a known mean. In the review of their work in 2020 [21], Ledoit and Wolf worked and proved their results in the case where the mean is known, and they claim at the end ”One then simply replaces S_T with \tilde{S}_T and T with $T-1$ in all the previous descriptions and computations in practice” (Section 6: Computational Aspects and Code). However, to the best of our knowledge, there are no proofs in the literature to extend the theoretical results, nor show the optimality of this approach. Moreover, focusing on Ledoit-Wolf linear shrinkage algorithm, one can note that the implementation used in ScikitLearn [9] doesn’t follow the recommendations of Ledoit and Wolf regarding the case of uncentered data. They didn’t change T to $T - 1$ and used $\frac{T-1}{T}\tilde{S}_T$ instead of \tilde{S}_T . Unexpectedly, experiments show notably worse results using Ledoit and Wolf recommendations rather than the ScikitLearn implementation. This remark underlines that the problem is more counter-intuitive than expected, and a closer look at the dependence between the covariance and the mean estimation is required.

We address the lack of theoretical results when the mean is unknown and propose a new Ledoit-Wolf-like linear shrinkage estimator and its theoretical and empirical analysis.

2 Notations, definitions and hypotheses

Let us introduce the following notations.

Notation 1. *In the following we consider a sequence of observation matrices $(X_n)_{n \in \mathbb{N}^*}$ with $X_n \in \mathbb{R}^{p_n \times n}$ of n iid observations on a system of p_n dimensions. Decomposing the covariance matrix, we denote $\Sigma_n = \Gamma_n \Lambda_n \Gamma_n^t$, where Λ_n is a diagonal matrix and Γ_n a rotation matrix. The diagonal elements of Λ_n are the eigenvalues $\lambda_1^n, \dots, \lambda_{p_n}^n$, and the columns of Γ_n are the eigenvectors $\gamma_1^n, \dots, \gamma_{p_n}^n$. $Y_n = \Gamma_n^t X_n$ is a $p_n \times n$ matrix of n iid observations of p_n uncorrelated random variables $(y_1^n, \dots, y_{p_n}^n)$.*

Notation 2. *Let A_n and B_n two $p_n \times p_n$ matrices. We consider the Frobenius norm: $\|A_n\| = \sqrt{\text{tr}(A_n A_n^t)/p_n}$, and the associated inner product: $\langle A_n, B_n \rangle_n = \text{tr}(A_n B_n^t)/p_n$. Dividing by the*

dimension is not standard, it is done to fix the norm of the identity as 1 regardless of the dimension.

Notation 3. Let $(E_n)_n$ a sequence of euclidean spaces with associated norm $\|\cdot\|_n$. The quadratic convergence of a random variable $Z_n \in E_n$, i.e. $\mathbb{E}[\|Z_n\|_n^2] \rightarrow 0$, is denoted as $Z_n \xrightarrow[q.m.]{} 0$.

We describe now several assumptions, the same used in the linear shrinkage of Ledoit and Wolf [10], that will be used in the following.

Assumption 1. There exists a constant K_1 independent of n such that $p_n/n \leq K_1$.

Assumption 2. There exists a constant K_2 independent of n such that $\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E}[(\tilde{y}_{i1}^n)^8] \leq K_2$ where for all $i \in \llbracket 1, p_n \rrbracket$, $\tilde{y}_{i1}^n = y_{i1}^n - \mathbb{E}[y_{i1}^n]$.

Assumption 3.

$$\lim_{n \rightarrow \infty} \frac{p_n^2}{n} \times \frac{\sum_{(i,j,k,l) \in Q_n} (\text{Cov}[\tilde{y}_{i1}^n \tilde{y}_{j1}^n, \tilde{y}_{k1}^n \tilde{y}_{l1}^n])^2}{|Q_n|} = 0,$$

where Q_n denotes the set of all the quadruples that are made of four distinct integers between 1 and p_n , and for all $i \in \llbracket 1, p_n \rrbracket$, $\tilde{y}_{i1}^n = y_{i1}^n - \mathbb{E}[y_{i1}^n]$.

We need some definitions to properly define the problem and the asymptotics.

Definition 1 (Empirical covariance). For an observation matrix X_n of size $p_n \times n$, we define the empirical covariance as:

$$S_n = \tilde{X}_n \tilde{X}_n^t / (n-1),$$

with $(\tilde{X}_n)_{ik} = (X_n)_{ik} - \frac{1}{n} \sum_{k'=1}^n (X_n)_{ik'}$.

Definition 2 (Scalars $(\mu_n, \alpha_n^2, \beta_n^2, \delta_n^2)$). We define four scalars:

$$\mu_n = \langle \Sigma_n, I_{p_n} \rangle_n, \alpha_n^2 = \|\Sigma_n - \mu_n I_{p_n}\|_n^2, \beta_n^2 = \mathbb{E}[\|S_n - \Sigma_n\|_n^2], \delta_n^2 = \mathbb{E}[\|S_n - \mu_n I_{p_n}\|_n^2].$$

(Lemma 2.1 in (Ledoit and Wolf, 2004) [10]) proves that $\alpha_n^2 + \beta_n^2 = \delta_n^2$.

The oracle linear shrinkage estimator is given by the following minimization problem. The following corollary is the central point of the linear shrinkage methods.

Corollary 1 (Corollary of theorem 2.1 from (Ledoit and Wolf, 2004)). Consider the optimization problem:

$$\begin{aligned} & \underset{\rho_1, \rho_2}{\text{minimize}} && \mathbb{E}[\|\Sigma_n^* - \Sigma_n\|_n^2], \\ & \text{s.t. } \Sigma_n^* = \rho_1 I_{p_n} + \rho_2 S_n \end{aligned}$$

where the coefficients ρ_1 and ρ_2 are not random. Its solution Σ_n^* verifies:

$$\Sigma_n^* = \frac{\beta_n^2}{\delta_n^2} \mu_n I_{p_n} + \frac{\alpha_n^2}{\delta_n^2} S_n, \quad \mathbb{E}[\|\Sigma_n^* - \Sigma_n\|_n^2] = \frac{\alpha_n^2 \beta_n^2}{\delta_n^2}.$$

Remark 1. Corollary 1 remains true for any unbiased estimator $\hat{\Sigma}_n$ instead of S_n .

$(\mu_n, \alpha_n^2, \beta_n^2, \delta_n^2)$ depends on the true covariance Σ_n , and thus can't be used directly in the estimation of Σ_n^* . The central issue of this work is to find estimators $(m_n, a_n^2, b_n^2, d_n^2)$ of $(\mu_n, \alpha_n^2, \beta_n^2, \delta_n^2)$ in order to compute an estimation S_n^* of Σ_n^* . As the mean is unknown, those estimators differ from Ledoit and Wolf work [10], particularly when p_n is higher than n .

3 Theoretical results

All the following results extend the work of Ledoit and Wolf [10] in the case where the empirical mean is used as estimator of the mean.

All proofs are shown in appendix A.

Remark 2. *In the following, as all the estimators are invariant by change of mean, resulting from the definition of $\tilde{X} = X - \sum_k X_{\cdot,k}$, we can assume $\mathbb{E}[X] = 0$ for the simplicity of notations.*

We present a sequence of lemmata, that naturally define estimators with suitable asymptotic properties for the scalars $(\mu_n, \alpha_n^2, \beta_n^2, \delta_n^2)$.

Lemma 1. *Under assumptions 1 and 2, $\mu_n, \alpha_n^2, \beta_n^2, \delta_n^2$ remain bounded as $n \rightarrow \infty$.*

Theorem 1. *Under assumptions 1 and 2, define $\theta_n^2 = \mathbb{V} \left[\frac{1}{p_n} \sum_{i=1}^{p_n} (y_{i1}^n)^2 \right]$. θ_n^2 is bounded as $n \rightarrow \infty$, and we have:*

$$\lim_{n \rightarrow \infty} \mathbb{E}[\|S_n - \Sigma_n\|_n^2] - \frac{p_n}{n} (\mu_n^2 + \theta_n^2) = 0.$$

In particular, taking $p_n = p$ constant, we see that $S_n \xrightarrow{q.m.} \Sigma_n$. But, when p_n is of the same order of magnitude than n , the sample covariance generally fails to converge as the error is at least of the same order of magnitude as $\mu_n^2 = \langle \Sigma_n, I_{p_n} \rangle_n^2$.

Lemma 2 (Estimator of μ_n). *Define $m_n = \langle S_n, I_{p_n} \rangle_n$. Then, under assumptions 1 and 2, $\mathbb{E}[m_n] = \mu_n$ for all n , and $m_n - \mu_n$ converges to zero in quartic mean (fourth moment) as n goes to infinity.*

Corollary 2. *Under assumptions 1 and 2, $m_n^2 - \mathbb{E}[m_n^2]$ converges to zero in quadratic mean as n goes to infinity.*

Lemma 3 (Estimator of δ_n^2). *Define $d_n^2 = \|S_n - m_n I_{p_n}\|_n^2$. Then, under assumptions 1, 2 and 3, $d_n^2 - \delta_n^2 \xrightarrow{q.m.} 0$. It follows that $d_n^2 - \mathbb{E}[d_n^2] \xrightarrow{q.m.} 0$.*

The following lemmata aim at defining an unbiased estimator of β_n^2 that quadratically converges. We work around \bar{b}_n^2 , inspired from the estimator of β_n^2 in the case where the mean is known, and write residual terms in the expectation as a combination of m_n and d_n^2 .

Lemma 4. *Define:*

$$\bar{b}_n^2 = \frac{1}{n^2} \sum_{k=1}^n \left\| \frac{n}{n-1} \tilde{x}_{\cdot,k}^n (\tilde{x}_{\cdot,k}^n)^t - S_n \right\|_n^2,$$

where $\tilde{x}_{\cdot,k}^n = x_{\cdot,k}^n - \frac{1}{n} \sum_{k'=1}^n x_{\cdot,k'}^n$, and $(x_{\cdot,1}^n, \dots, x_{\cdot,k}^n, \dots, x_{\cdot,n}^n)$ are the independent samples forming X . Then, under assumption 1,

$$\mathbb{E}[\bar{b}_n^2] = c_0 \beta_n^2 + c_1 \delta_n^2 + c_2 \mu_n^2,$$

with $\gamma_n = \frac{n(n-1)}{n^2-3n+3}$, $\lambda_n = \frac{n^2(n-2)}{(n-1)(n^2-3n+3)}$, $c_0 = \frac{1}{\gamma_n} - \frac{1}{n} - \frac{\lambda_n}{\gamma_n n^2}$, $c_1 = \frac{\lambda_n}{\gamma_n n^2}$, $c_2 = (p+1)c_1$.

Lemma 5. *Under assumption 1, we have: $\mathbb{E}[\bar{b}_n^2] = c_0 \beta_n^2 + c_1 \mathbb{E}[d_n^2] + c_2 \mathbb{E}[m_n^2] + (c_1 - c_2) \mathbb{V}[m_n]$.*

Lemma 6. *Under assumption 1, we have: $\mathbb{V}[\bar{b}_n^2] \xrightarrow{n \rightarrow \infty} 0$.*

We need to compute $\mathbb{V}[m_n]$ which is unknown in the development of $\mathbb{E}[\bar{b}_n^2]$ in Lemma 5.

Lemma 7. Under assumption 1, we have: $\mathbb{V}[m_n] = q_0\beta_n^2 + q_1\delta_n^2 - q_2\mu_n^2$, with $q_0 = \frac{n-2}{p(n-1)}$, $q_1 = \frac{1}{p(n-1)}$, $q_2 = \frac{p-1}{p(n-1)}$.

Lemma 8. Under assumption 1, we have: $\mathbb{V}[m_n] = \frac{1}{1-q_1-q_2} (q_0\beta_n^2 + q_1\mathbb{E}[d_n^2] - q_2\mathbb{E}[m_n^2])$.

Lemma 9. Define:

$$b_n^2 = \frac{1}{c_0^f} \left(\bar{b}_n^2 - c_1^f d_n^2 - c_2^f m_n^2 \right),$$

with $c_0^f = c_0 + (c_1 - c_2)\frac{q_0}{1-q_1-q_2}$, $c_1^f = c_1 + (c_1 - c_2)\frac{q_1}{1-q_1-q_2}$, $c_2^f = c_2 - (c_1 - c_2)\frac{q_2}{1-q_1-q_2}$.

Then, under assumptions 1, 2 and 3, b_n^2 is an unbiased estimator of β_n^2 , i.e. $\mathbb{E}[b_n^2] = \beta_n^2$, and $b_n^2 - \beta_n^2 \xrightarrow[q.m.]{} 0$.

For notation consistency with the estimators in Ledoit-Wolf linear shrinkage [10], we keep the notation b_n^2 even if its value can be negative.

Lemma 10 (Estimator of β_n^2). Define: $b_{n,u}^2 = \min((b_n^2)_+, d_n^2)$ and $a_{n,u}^2 = d_n^2 - b_{n,u}^2$. Under assumptions 1, 2 and 3, $b_{n,u}^2 - \beta_n^2 \xrightarrow[q.m.]{} 0$ and $a_{n,u}^2 - \alpha_n^2 \xrightarrow[q.m.]{} 0$.

We can now define our linear shrinkage estimator S_n^* and prove its asymptotic properties.

Definition 3 (Final estimator LW_u). Let's define our estimator:

$$S_n^* = \frac{b_{n,u}^2}{d_{n,u}^2} m_{n,u} I_{p_n} + \frac{a_{n,u}^2}{d_{n,u}^2} S_n, \text{ with } m_{n,u} = m_n, d_{n,u}^2 = d_n^2.$$

Theorem 2. Under assumptions 1, 2 and 3, $\mathbb{E}[\|S_n^* - \Sigma_n^*\|^2] \rightarrow 0$. As a consequence, S_n^* has the same asymptotic expected loss as Σ_n^* , i.e. $\mathbb{E}[\|S_n^* - \Sigma_n\|_n^2] - \mathbb{E}[\|\Sigma_n^* - \Sigma_n\|_n^2] \rightarrow 0$.

The following lemma gives an asymptotic estimation of the optimal error

$$\mathbb{E} [\|\Sigma_n^* - \Sigma_n\|_n^2] = \frac{\alpha_n^2 \beta_n^2}{\delta_n^2}.$$

Lemma 11.

$$\mathbb{E} \left[\left| \frac{a_{n,u}^2 b_{n,u}^2}{d_n^2} - \frac{\alpha_n^2 \beta_n^2}{\delta_n^2} \right| \right] \rightarrow 0.$$

The last results easily make possible to extend the Theorems 3.3 and 3.4 of (Ledoit and Wolf, 2004) [10] in our situation where the mean is unknown. Previously, we showed that our estimator's loss converge to the optimal one in the class of linear combinations of S_n and I_{p_n} with non random coefficients, the optimal estimator of this class being Σ_n^* . In the following, we show that our estimator is still asymptotically optimal with respect to a bigger class, where the coefficients can be random. Formally, we are looking for the following optimal loss (this time, there is no expectation in the minimization). Let Σ_n^{**} be the linear combination of S_n and I_{p_n} solving:

$$\begin{aligned} & \underset{\rho_1, \rho_2}{\text{minimize}} && \|\Sigma_n^{**} - \Sigma_n\|_n^2. \\ & \text{s.t. } \Sigma_n^{**} = \rho_1 I_{p_n} + \rho_2 S_n \end{aligned}$$

By construction, Σ_n^{**} has a lower loss than S_n^* , but we show that the difference converges to 0.

Theorem 3. S_n^* converges to Σ_n^{**} in quadratic mean, i.e. $\|S_n^* - \Sigma_n^{**}\| \xrightarrow[q.m.]{} 0$. As a consequence, S_n^* has the same asymptotic expected loss as Σ_n^{**} , more precisely we have:

$$\mathbb{E} [\|\|S_n^* - \Sigma_n\|_n^2 - \|\Sigma_n^{**} - \Sigma_n\|_n^2\|] \rightarrow 0.$$

Theorem 4. For any sequence of linear combinations $\hat{\Sigma}_n$ of I_n and S_n , the estimator S_n^* verifies:

$$\lim_{N \rightarrow \infty} \inf_{n \geq N} \left(\mathbb{E} \left[\|\hat{\Sigma}_n - \Sigma_n\|_n^2 \right] - \mathbb{E} \left[\|S_n^* - \Sigma_n\|_n^2 \right] \right) \geq 0.$$

In addition, every $\hat{\Sigma}_n$ that performs as well as S_n^* is identical to S_n^* in the limit:

$$\lim_{N \rightarrow \infty} \left(\mathbb{E} \left[\|\hat{\Sigma}_n - \Sigma_n\|_n^2 \right] - \mathbb{E} \left[\|S_n^* - \Sigma_n\|_n^2 \right] \right) = 0 \iff \|\hat{\Sigma}_n - S_n^*\|_n \xrightarrow{q.m} 0.$$

We introduce three other estimators to compare with, which are implemented, recommended, or natural to define. We prove that their asymptotic behavior is similar, and, through different experiments, show the differences in performance.

Definition 4 (Ledoit-Wolf recommended estimators). *The estimators recommended by Ledoit and Wolf [21], indexed by the letter "r", are:*

$$m_{n,r} = m_n, d_{n,r}^2 = d_n^2, b_{n,r}^2 = \min \left(\left(\frac{1}{(n-1)^2} \sum_{k=1}^n \|\tilde{x}_{\cdot k}^n (\tilde{x}_{\cdot k}^n)^t - S_n\|_n^2 \right)_+, d_{n,r}^2 \right), a_{n,r}^2 = d_{n,r}^2 - b_{n,r}^2.$$

Theorem 5 (Ledoit-Wolf recommended estimators). *Under Assumptions 1, 2 and 3, $m_{n,r} - \mu_n$ converges to 0 in quartic mean, and that $d_{n,r}^2 - \delta_n^2$, $b_{n,r}^2 - \beta_n^2$ and $a_{n,r}^2 - \alpha_n^2$ converge in quadratic mean to 0 as n goes to infinity.*

Moreover, the conclusions of Theorem 2 remain true with the estimated matrix $S_{n,r}^* = \frac{b_{n,r}^2}{d_{n,r}^2} m_{n,r} I_{p_n} + \frac{a_{n,r}^2}{d_{n,r}^2} S$, i.e. $\mathbb{E}[\|S_{n,r}^* - \Sigma_n^*\|^2] \rightarrow 0$ and $\mathbb{E}[\|S_{n,r}^* - \Sigma_n\|_n^2 - \|\Sigma_n^{**} - \Sigma_n\|_n^2] \rightarrow 0$.

From the proof, $\bar{b}_{n,r}^2 = \bar{b}_n^2 + \frac{1}{n(n-1)^2} \|S_n\|_n^2$, it is then natural to define the following estimator.

Definition 5 ("Natural" estimators). *The estimators that naturally emerge, indexed by the letter "m", are:*

$$m_{n,m} = m_n, d_{n,m}^2 = d_n^2, b_{n,m}^2 = \min \left((\bar{b}_n^2)_+, d_{n,m}^2 \right), a_{n,m}^2 = d_{n,m}^2 - b_{n,m}^2.$$

Theorem 6 ("Natural" estimators). *Under Assumptions 1, 2 and 3, $m_{n,m} - \mu_n$ converges to 0 in quartic mean, and that $d_{n,m}^2 - \delta_n^2$, $b_{n,m}^2 - \beta_n^2$ and $a_{n,m}^2 - \alpha_n^2$ converge in quadratic mean to 0 as n goes to infinity.*

Moreover, the conclusions of Theorem 2 remain true with the estimated matrix $S_{n,m}^* = \frac{b_{n,m}^2}{d_{n,m}^2} m_{n,m} I_{p_n} + \frac{a_{n,m}^2}{d_{n,m}^2} S$, i.e. $\mathbb{E}[\|S_{n,m}^* - \Sigma_n^*\|^2] \rightarrow 0$ and $\mathbb{E}[\|S_{n,m}^* - \Sigma_n\|_n^2 - \|\Sigma_n^{**} - \Sigma_n\|_n^2] \rightarrow 0$.

Definition 6 (ScikitLearn 1.2.2 estimators). *The estimators implemented in ScikitLearn 1.2.2, indexed by the letter "s", are:*

$$m_{n,s} = \frac{n-1}{n} m_n, d_{n,s}^2 = d_n^2, b_{n,s}^2 = \min \left((\bar{b}_n^2)_+, d_{n,s}^2 \right), a_{n,s}^2 = \frac{n-1}{n} (d_{n,s}^2 - b_{n,s}^2),$$

$$S_{n,s}^* = \frac{n-1}{n} S_{n,m}^*.$$

Theorem 7 (Scikit-Learn 1.2.2 estimators). *Under Assumptions 1, 2 and 3, $m_{n,s} - \mu_n$ converges to 0 in quartic mean, and that $d_{n,s}^2 - \delta_n^2$, $b_{n,s}^2 - \beta_n^2$ and $a_{n,s}^2 - \alpha_n^2$ converge in quadratic mean to 0 as n goes to infinity.*

Moreover, the conclusions of Theorem 2 remain true with the estimated matrix $S_{n,s}^ = \frac{b_{n,s}^2}{d_{n,s}^2} m_{n,s} I_{p_n} + \frac{a_{n,s}^2}{d_{n,s}^2} S$, i.e. $\mathbb{E}[\|S_{n,s}^* - \Sigma_n^*\|^2] \rightarrow 0$ and $\mathbb{E}[\|\|S_{n,s}^* - \Sigma_n\|_n^2 - \|\Sigma_n^{**} - \Sigma_n\|_n^2\|] \rightarrow 0$.*

4 Experimental results

The experimental estimations are compared to the theoretical value of Σ^* in the Ledoit-Wolf setting, the implementation in ScikitLearn 1.2.2, the implementation recommended by Ledoit and Wolf [21], and to the other algorithms implemented in ScikitLearn 1.2.2, for multivariate Gaussian and Student-t distributions.

We first derive the exact values of Σ^* for those two distributions.

4.1 Oracle estimators

4.1.1 Gaussian distribution

Lemma 12. *Let $(X_n)_{.k} \sim \mathcal{N}(0, \Sigma_n)$, $k \in \llbracket 1, n \rrbracket$, n iid samples. Then, the analytical oracle estimators are: $\mu_n = \langle \Sigma_n, I \rangle_n$, $\alpha_n^2 = \|\Sigma_n\|_n^2 - \mu_n^2$, $\beta_n^2 = \frac{p+1}{n-1} \mu_n^2 + \frac{1}{n-1} \alpha_n^2$, $\delta_n^2 = \alpha_n^2 + \beta_n^2$.*

4.1.2 t-distribution oracle estimator

Lemma 13. *Let $(X_n)_{.k} \sim t_\nu(0, \tilde{\Sigma}_n)$, $k \in \llbracket 1, n \rrbracket$, n iid samples with scale matrix $\tilde{\Sigma}_n = \frac{\nu-2}{\nu} \Sigma_n$ and covariance $\mathbb{V}[X_n] = \Sigma_n$. The density of the multivariate t-distribution is:*

$$f_n(x) = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu+p) \nu^{p/2} \pi^{p/2} |\tilde{\Sigma}_n|^{1/2}} \left[1 + \frac{1}{\nu} x^T \tilde{\Sigma}_n^{-1} x \right]^{-(\nu+p)/2}.$$

Then, the analytical oracle estimators are:

$$\mu_n = \langle \Sigma_n, I \rangle, \alpha_n^2 = \|\Sigma_n\|_n^2 - \mu_n^2, \beta_n^2 = \frac{1}{n} \left(\frac{\nu}{\nu-4} + \frac{1}{n-1} \right) (\alpha_n^2 + (p+1)\mu_n^2) - \frac{2p}{n(\nu-4)} \mu_n^2, \delta_n^2 = \alpha_n^2 + \beta_n^2.$$

4.2 Experimental setup

We considered 2 settings:

- a Monte-Carlo computation of the loss on a 2d-grid of the parameters $(p_n, n) \in \llbracket 5, 100 \rrbracket^2$, with a step size of 2, to visualize the effect of changing the ratio p_n/n , and see the domains where our algorithm is most suited;
- a Monte-Carlo computation of the loss with a fixed ratio $p_n/n = c \in \{1, 2, 4\}$, to compare the rate of convergence of each algorithm.

In both cases, the Monte-Carlo is computed with $n_{MC} = 10000$ iterations.

Three different distributions are explored: the multivariate Gaussian, and the t-distribution with $\nu = 10$ and $\nu = 8.5$. Note that we have to ensure $\nu > 8$ to respect Assumption 2.

Two different way of choosing Σ_n are explored: fixing $\Sigma_n = I_{p_n}$ - particular case where the oracle Ledoit-Wolf loss is null -, and drawing at each iteration a covariance matrix Σ_n from a Wishart distribution with p_n degrees of freedom, and normalizing it by $\sqrt{\|\Sigma_n \Sigma_n^T\|_n}$ - to respect the assumption 2. Note that when drawn from a Wishart, $\sqrt{\|\Sigma_n \Sigma_n^T\|_n} > 0$ almost surely.

4.2.1 Assumptions check

For the second study at fixed p_n/n in order to compare the rate of convergence, we check that we are under the three assumptions that guarantee the theoretical results on convergence proved in section 3.

Assumption 1 As we fixed the ratio $c = p_n/n$, Assumption 1 is trivially respected.

Assumption 2 - Gaussian distribution Let $(X_n)_{\cdot,k} \sim \mathcal{N}(0, \Sigma_n)$, $k \in \llbracket 1, n \rrbracket$, n iid samples. As previously, we denote $\Sigma_n = \Gamma_n \Lambda_n \Gamma_n^t$, where Λ_n is a diagonal matrix and Γ_n a rotation matrix, and $Y_n = \Gamma_n^t X_n$ is a $p_n \times n$ matrix of n iid observations of p_n uncorrelated random variables. Using the fact that for $z \sim \mathcal{N}(0, \lambda)$, we have $\mathbb{E}[z^8] = 105\lambda^4$, we deduce:

$$\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E}[y_{i1}^8] = \frac{1}{p_n} \sum_{i=1}^{p_n} 105\lambda_{ii}^4 = 105 \|\Lambda_n \Lambda_n^T\|_n^2 = 105 \|\Sigma_n \Sigma_n^T\|_n^2. \quad (1)$$

In the case where we fix $\Sigma_n = I_{p_n}$, we obviously have $\|\Sigma_n \Sigma_n^T\|_n^2 = 1$, so Assumption 2 is respected.

In the case where we draw Σ_n from a Wishart distribution with p_n degrees of freedom, and normalize it by $\sqrt{\|\Sigma_n \Sigma_n^T\|_n}$, we have by construction $\|\Sigma_n \Sigma_n^T\|_n^2 = 1$, so Assumption 2 is respected here too.

Assumption 2 - t-distribution Let $(X_n)_{\cdot,k} \sim t_\nu(0, \tilde{\Sigma}_n)$, $k \in \llbracket 1, n \rrbracket$, n iid samples with $\nu > 8$, scale matrix $\tilde{\Sigma}_n = \frac{\nu-2}{\nu} \Sigma_n$ and covariance $\mathbb{V}[X_n] = \Sigma_n$. As previously, we denote $\Sigma_n = \Gamma_n \Lambda_n \Gamma_n^t$, where Λ_n is a diagonal matrix and Γ_n a rotation matrix, and $Y_n = \Gamma_n^t X_n$ is a $p_n \times n$ matrix of n iid observations of p_n uncorrelated random variables.

From a characterization of multivariate t-distributions, for each $k \in \llbracket 1, n \rrbracket$, there exist 2 independent random variables U_k and $Z_{\cdot,k}$ such that:

$$U_k \sim \chi_\nu^2, Z_{\cdot,k} \sim \mathcal{N}\left(0, \frac{\nu-2}{\nu} \Lambda_n\right), (Y_n)_{\cdot,k} = \Sigma^{-1/2} (X_n)_{\cdot,k} = \sqrt{\frac{\nu}{U_k}} Z_{\cdot,k}.$$

Moreover, we notice that:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{U_1^4}\right] &= \int_{\mathbb{R}} \frac{(1/2)^{\nu/2-1}}{\Gamma(\nu/2)} x^{(\nu-8)/2-1} e^{-x/2} dx \\ &= \frac{(1/2)^{\nu/2}}{(1/2)^{(\nu-8)/2}} \frac{\Gamma((\nu-8)/2)}{\Gamma(\nu/2)} \underbrace{\int_{\mathbb{R}} \frac{(1/2)^{(\nu-8)/2-1}}{\Gamma((\nu-8)/2)} x^{(\nu-8)/2-1} e^{-x/2} dx}_{\text{pdf of } \chi_{\nu-8}^2} dx \\ \mathbb{E}\left[\frac{1}{U_1^4}\right] &= \frac{1}{(\nu-8)(\nu-6)(\nu-4)(\nu-2)}. \end{aligned} \quad (2)$$

This 2 previous points lead to:

$$\begin{aligned}
\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E}[y_{i1}^8] &= \frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E}[z_{i1}^8] \mathbb{E}\left[\frac{\nu^4}{U_1^4}\right] \\
&= \frac{1}{p_n} \sum_{i=1}^{p_n} 105 \left(\frac{\nu-2}{\nu} \lambda_{ii}\right)^4 \frac{\nu^4}{(\nu-8)(\nu-6)(\nu-4)(\nu-2)} \\
&= 105 \frac{(\nu-8)(\nu-6)(\nu-4)}{(\nu-2)^3} \|\Lambda \Lambda^T\|_n^2 \\
\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E}[y_{i1}^8] &= 105 \frac{(\nu-8)(\nu-6)(\nu-4)}{(\nu-2)^3} \|\Sigma \Sigma^T\|_n^2.
\end{aligned} \tag{3}$$

Similarly as the Gaussian case, when we fix $\Sigma_n = I_{p_n}$, we obviously have $\|\Sigma_n \Sigma_n^T\|_n^2 = 1$, so Assumption 2 is respected, and when we draw Σ_n from a Wishart distribution with p_n degrees of freedom, and normalize it by $\sqrt{\|\Sigma_n \Sigma_n^T\|_n}$, we have by construction $\|\Sigma_n \Sigma_n^T\|_n^2 = 1$, so Assumption 2 is respected here too.

Assumption 3 Let Y a p_n -dimensional random variables drawn from a centered multivariate Gaussian distribution, then for all $(i, j) \in \llbracket 1, p_n \rrbracket^2$, we have the following property:

$$\text{Cov}(Y_i, Y_j) = 0 \implies Y_i, Y_j \text{ independent.}$$

Moreover, when drawing X_n from n iid multivariate Gaussian, we have that Y_n , using the previous notations, is made of n iid samples of an uncorrelated p_n -dimensional centered multivariate Gaussian distribution.

So, for all $(i, j) \in \llbracket 1, p_n \rrbracket^2$, $i \neq j$, we have that y_{i1}, y_{j1} independent.

Finally, for all $(i, j, k, l) \in \llbracket 1, p_n \rrbracket^4$ where i, j, k, l are all different, we have:

$$\text{Cov}[y_{i1}^n y_{j1}^n, y_{k1}^n y_{l1}^n] = 0.$$

In the case where X_n are n p_n -dimensional iid samples drawn from a centered multivariate t-distribution, we have that Y_n , using the previous notations, is made of n iid samples of an uncorrelated p_n -dimensional centered multivariate t-distribution. Then we use the decomposition $Y = \sqrt{\frac{\nu}{U}} Z$ where Z is drawn from a multivariate Gaussian distribution independent from U , drawn from a χ_ν^2 distribution. As for all $i \neq j$, $\text{Cov}(y_{i1}^n, y_{j1}^n) = 0$, then we trivially have $\text{Cov}(z_{i1}^n, z_{j1}^n) = 0$. So z_{i1} and z_{j1} are independent, which immediately leads to the fact that for all $(i, j, k, l) \in \llbracket 1, p_n \rrbracket^4$ where i, j, k, l are all different, we have:

$$\text{Cov}[y_{i1}^n y_{j1}^n, y_{k1}^n y_{l1}^n] = 0.$$

This proves that Assumption 3 is respected in all the experimental cases we studied.

4.3 Results

In the following, we will use abbreviations to refer the different expected losses of each algorithms. Concerning the variants of Ledoit-Wolf shrinkage estimators with unknown mean, we denote:

- LW_u for the estimator we propose in this paper,

- *LW_r* for the implementation recommended by Ledoit and Wolf in 2020 [21],
- *LW_s* for the implementation of ScikitLearn 1.2.2,
- *LW_m* for the natural estimator,
- *LW_ex* for the oracle estimator Σ^* ,
- *LW_op* for the optimal estimator Σ^{**} ,

Concerning the other baseline algorithms implemented in ScikitLearn, we have:

- *EC* for the Empirical Covariance estimator,
- *SC* for the Shrunk Covariance estimator,
- *OAS* for the Oracle Approximated Shrinkage estimator,

We didn't run the Elliptic Envelope, GLasso and MinCovDet estimators present in ScikitLearn, due to time complexity: in our setup, the computing time of those ones exceeds by a factor at least 10 the computing time of the shrinkage estimators listed before. Consequently, for reason of feasibility, we chose not to compare to them, considering that the latter algorithms are part of a different class of estimators.

4.3.1 Constant covariance $\Sigma_n = I_{p_n}$

Study on a grid over (p,n)

As they often show similar behaviors, we only show a subset of the experimental results for brevity. The three estimators *LW_s*, *LW_r* and *LW_m* have a very similar behavior compared to *LW_u* in this scenario, that's why we will only show the comparison with *LW_m*, having the best performance among the three. The results are shown in figure 1. The black contour on the surface plots is the iso-line at level 0, where the expected losses are equal. In this scenario, *LW_u* is constantly better than the other estimators, and the important difference is in the part $p > n$, where the mean estimation affects a lot the overall covariance estimation.

Convergence study

We now fix $c = p_n/n$ and study the convergence of the different algorithms we cited in the experimental setup. We only show the results with $c = 1$ as the other cases only widen the differences but do not change the order. The cases t_{10} and $t_{8.5}$ -distributions are very similar, that's why we show only the $t_{8.5}$ one. The key difference between the Gaussian case and the t-distribution, is that the OAS doesn't converge in the latter, while not being so efficient in the Gaussian case which is tailored for it. The results are shown in figure 2.

4.3.2 Covariance drawn from normalized Wishart

Study on a grid over (p,n)

The three estimators *LW_r* and *LW_m* have a very similar behavior compared to *LW_u* in this scenario, that's why we will only show the comparison with *LW_m*, having the best performance among the two. Moreover, the results between the t_{10} and $t_{8.5}$ -distributions are very similar, so we will only show the $t_{8.5}$ case. The results are shown in figures 3 and 4. The black contour on the surface plots is the iso-line at level 0, here it is where the losses are equal. In the case $p > n$, *LW_u* is far better than the other estimators, where the mean estimation affects a lot the overall covariance estimation. In a finite subset of the part $n > p$, *LW_s* is slightly better. *LW_m* presents no significant advantage compared to the two others.

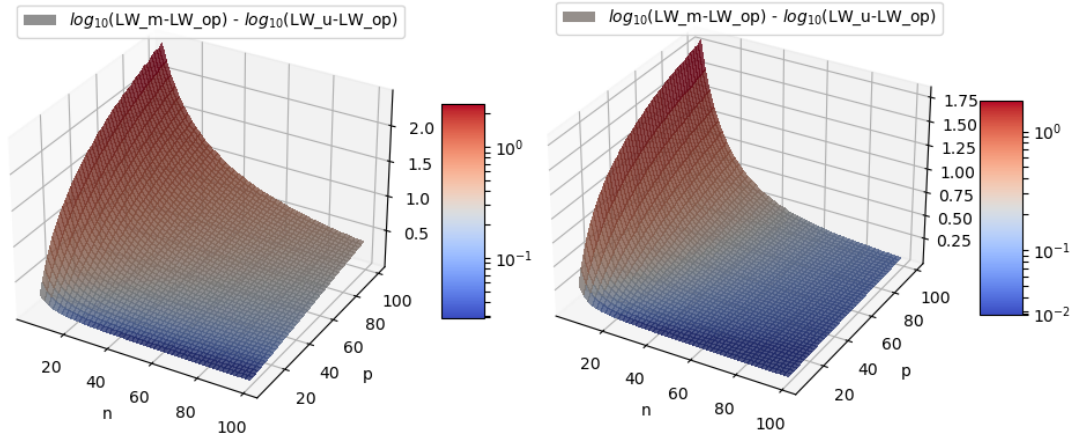


Figure 1: With p the dimension, n the number of samples, the axis z shows the difference of the \log_{10} of the expected losses between LW_m and LW_u. The losses are relative to the theoretical bound LW_op. The samples are drawn with a Gaussian (left)/ $t_{8.5}$ (right) distribution, $\Sigma = I_p$.

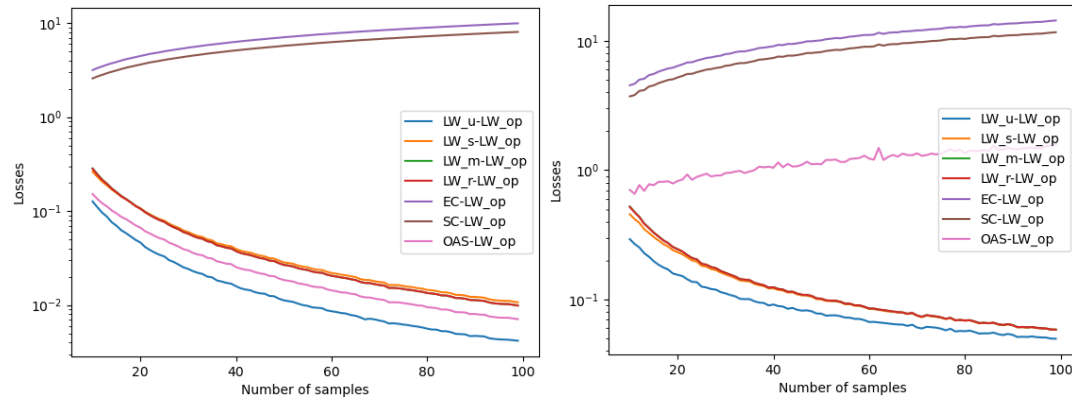


Figure 2: oss comparison to LW_op, Gaussian (left)/ $t_{8.5}$ (right) distribution, $\Sigma = I_p$, $c = 1$. Here, note that LW_ex and LW_op have both null loss, we didn't plot them.

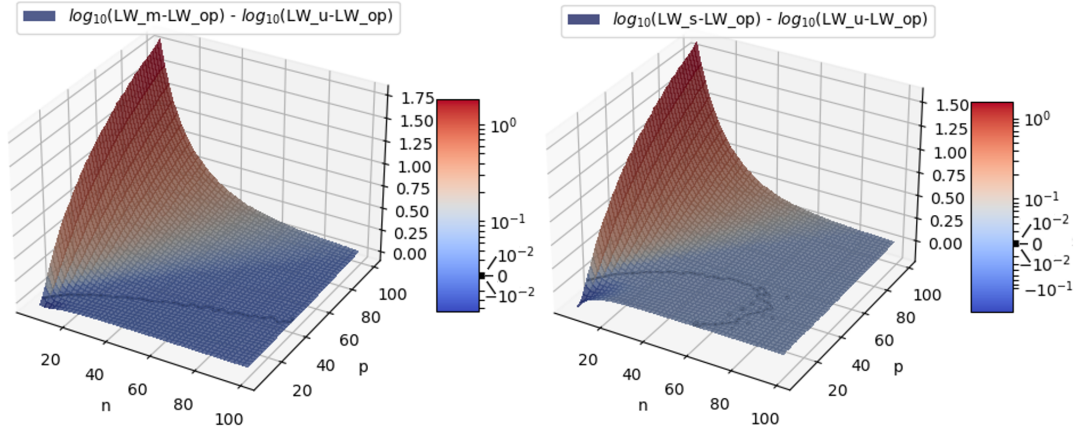


Figure 3: With p the dimension, n the number of samples, the axis z shows the difference of the \log_{10} of the expected losses between LW_m (left)/LW_s (right) and LW_u. The losses are relative to the theoretical bound LW_op. Samples are drawn from a Gaussian distribution, random Σ . The black contour is the iso-line at level 0, where the expected losses are equal.

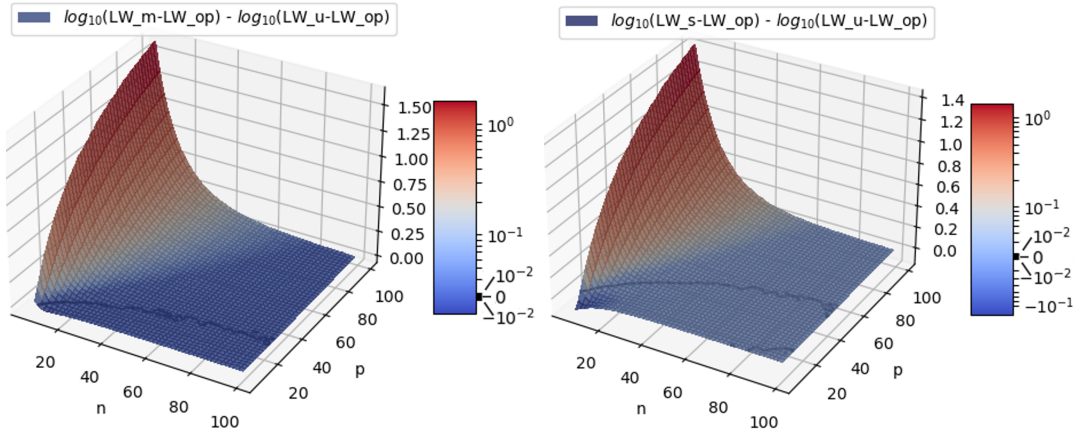


Figure 4: With p the dimension, n the number of samples, the axis z shows the difference of the \log_{10} of the expected losses between LW_m (left)/LW_s (right) and LW_u. The losses are relative to the theoretical bound LW_op. Samples are drawn from a $t_{8.5}$ -distribution, random Σ . The black contour is the iso-line at level 0, where the expected losses are equal.

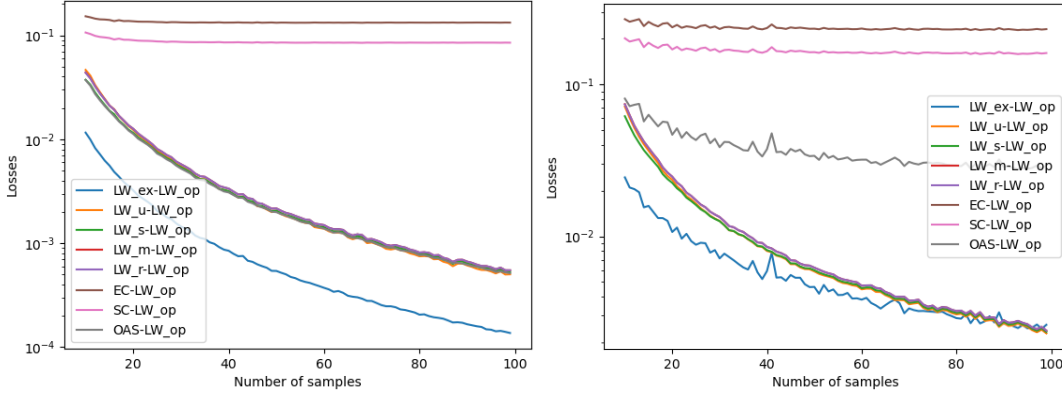


Figure 5: Loss comparison to LW_op, Gaussian (left)/ $t_{8.5}$ (right) distribution, random Σ , $c = 1$

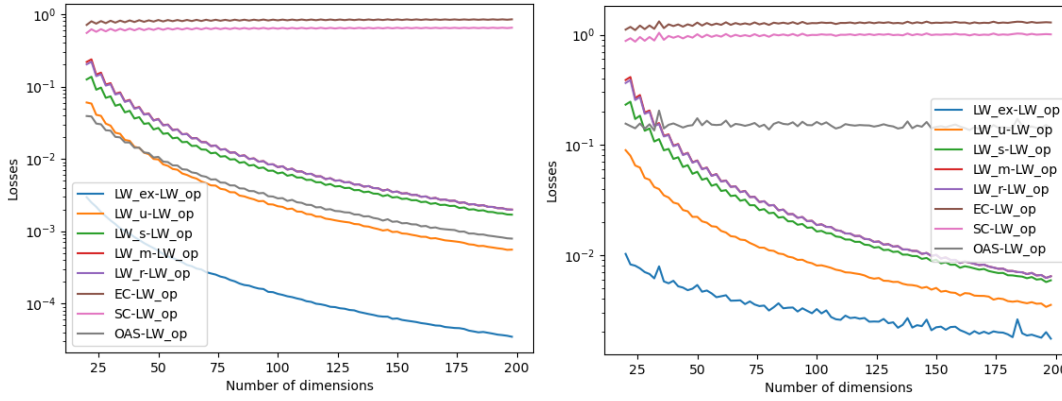


Figure 6: Loss comparison to LW_op, Gaussian (left)/ $t_{8.5}$ (right) distribution, random Σ , $c = 4$. Note that LW_r and LW_m are not discernable.

Convergence study

We fix $c = p_n/n$ and study the convergence of the different algorithms we cited in the experimental setup. The cases t_{10} and $t_{8.5}$ -distributions being very similar, we only show $t_{8.5}$. The key difference between the Gaussian case and the t-distribution, is that the OAS doesn't converge in the latter, while not being so efficient in the Gaussian case which is tailored for it. The results are shown in figures 5 and 6.

5 Conclusion

In this work, we extended the linear shrinkage approach of Ledoit and Wolf [10] for covariance matrix estimation to the case where the mean of the distribution is unknown. Theoretically, we showed that in this case we have similar asymptotic properties as in the situation when the mean is known. Four different estimators emerged, three around those implemented in ScikitLearn or recommended by Ledoit and Wolf, and one naturally emerging from the theoretical proofs. Experimentally, the latter showed improved performances in a large spectrum of situations compared to ScikitLearn 1.2.2 baselines and to the three other estimators presented in the theoretical part. The gain in performance is particularly high when the dimension is bigger than

the number of samples, while the differences are comparably low when dimension is smaller than the number of samples.

Ledoit and Wolf developed several non-linear shrinkage estimators where the mean is known [15, 16, 17]. Work needs to be conducted to investigate if similar approach can be used to extend their non-linear frameworks.

Acknowledgments

We want to thank Gabriel Turinici for its advices all along the work, and Olivier Ledoit and Michael Wolf for the impressive work they have done in the field of covariance estimation for twenty years.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6 Appendix A: Proofs of the technical results

For brevity, we omit the subscript n ; but it is understood that everything depends on n . Coefficients of Λ are denoted λ_{ij} , and if not stated otherwise, sum indices i, j, l, m are in $\llbracket 1, p \rrbracket$ and $k_1, k'_1, k_2, k'_2, \dots$ are in $\llbracket 1, n \rrbracket$. Moreover, we denote $\bar{y} = \frac{1}{n} \sum_k y_{\cdot, k}$. We recall that, from Remark 2, as all the estimators are invariant by change of mean from the definition of $\tilde{X} = X - \sum_k X_{\cdot, k}$, we assume $\mathbb{E}[X] = 0$ for the simplicity of notations.

6.1 Technical lemma

Frequently used identities are proven in preamble of the other proofs here.

6.1.1 Identity 1

Let $(i, j) \in \llbracket 1, p \rrbracket^2$, then:

$$\begin{aligned}
\sum_k (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) &= \sum_k \left(y_{ik} - \frac{1}{n} \sum_{k'} y_{ik'} \right) \left(y_{jk} - \frac{1}{n} \sum_{k''} y_{jk''} \right) \\
&= \frac{1}{n^2} \sum_{k, k', k''} (y_{ik} - y_{ik'}) (y_{jk} - y_{jk''}) \\
&= \frac{1}{n^2} \sum_{k, k', k''} (y_{ik} y_{jk} - y_{ik'} y_{jk} - y_{ik} y_{jk''} + y_{ik'} y_{jk''}) \\
&= \frac{1}{n^2} \sum_{k, k', k''} (y_{ik} y_{jk} - y_{ik} y_{jk'} - y_{ik'} y_{jk} + y_{ik'} y_{jk'}) \quad (\text{re-indexing}) \\
&= \frac{1}{n^2} \sum_{k, k', k''} y_{ik} (y_{jk} - y_{jk'}) \\
\sum_k (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) &= \frac{1}{n} \sum_{k, k' \neq k} y_{ik} (y_{jk} - y_{jk'}).
\end{aligned}$$

(4)

6.1.2 Identity 2

$$\frac{1}{p} \sum_{i,j} \lambda_{ij}^2 = \|\Lambda\|_p^2 = \|\Sigma\|_p^2 = \|\Sigma - \mu I\|_p^2 + \|\mu I\|_p^2 = \alpha^2 + \mu^2. \quad (5)$$

6.1.3 Identity 3

$$\frac{1}{p} \sum_{i,j} \lambda_{ii} \lambda_{jj} = \frac{1}{p} \left(\sum_i \lambda_{ii} \right)^2 = \frac{1}{p} \text{tr}(\Lambda)^2 = \frac{1}{p} \text{tr}(\Sigma)^2 = p \langle \Sigma, I \rangle_p^2 = p \mu^2. \quad (6)$$

6.2 Proof of Lemma 1

We have:

$$\|\Sigma\|^2 = \|\Lambda\|^2 = \frac{1}{p} \sum_{i=1}^p \mathbb{E}[y_{i1}^2]^2 \leq \frac{1}{p} \sum_{i=1}^p \mathbb{E}[y_{i1}^4] \leq \sqrt{\frac{1}{p} \sum_{i=1}^p \mathbb{E}[y_{i1}^8]} \leq \sqrt{K_2}. \quad (7)$$

As $\mu = \langle \Sigma, I \rangle \leq \|\Sigma\|$, μ remains bounded as n goes to infinity.

Also, $\alpha^2 = \|\Sigma - \mu I\|^2 = \|\Sigma\|^2 - \mu^2$, so remains bounded as n goes to infinity too.

For β^2 , we will deeply decompose the expectation. This is not absolutely necessary to prove the boundedness, but the decomposition will be of utter importance in the following proofs. So, we have:

$$\begin{aligned} \beta^2 &= \mathbb{E}[\|S - \Sigma\|^2] \\ &= \frac{1}{p} \sum_{i,j} \mathbb{E} \left[\left(\frac{1}{n-1} \sum_k (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) - \lambda_{ij} \right)^2 \right] \\ &= \frac{1}{p} \sum_{i,j} \mathbb{E} \left[\left(\frac{1}{n(n-1)} \sum_{k,k' \neq k} y_{ik}(y_{jk} - y_{jk'}) - \lambda_{ij} \right)^2 \right] \quad (\text{Identity 1}) \\ \beta^2 &= \frac{1}{p} \sum_{i,j} \frac{1}{n^2(n-1)^2} \sum_{k_1} \sum_{k'_1 \neq k_1} \sum_{k_2} \sum_{k'_2 \neq k_2} \mathbb{E} \left[(y_{ik_1}(y_{jk_1} - y_{jk'_1}) - \lambda_{ij}) (y_{ik_2}(y_{jk_2} - y_{jk'_2}) - \lambda_{ij}) \right]. \end{aligned} \quad (8)$$

We denote, for $k_1, k'_1 \neq k_1, k_2, k'_2 \neq k_2$:

$$E_{ij}(k_1, k'_1, k_2, k'_2) = \mathbb{E} \left[(y_{ik_1}(y_{jk_1} - y_{jk'_1}) - \lambda_{ij}) (y_{ik_2}(y_{jk_2} - y_{jk'_2}) - \lambda_{ij}) \right]. \quad (9)$$

- If $|\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 0$: $E_{ij}(k_1, k'_1, k_2, k'_2) = 0$.
- If $|\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 1$:
 - If $k_1 = k_2$: then $k'_1 \neq k'_2$, and $k_1 \neq k'_2$. So,

$$\begin{aligned} E_{ij}(k_1, k'_1, k_1, k'_2) &= \mathbb{E} \left[(y_{ik_1}(y_{jk_1} - y_{jk'_1}) - \lambda_{ij}) (y_{ik_1}(y_{jk_1} - y_{jk'_2}) - \lambda_{ij}) \right] \\ &= \mathbb{E} \left[(y_{ik_1} y_{jk_1} - \lambda_{ij})^2 \right] \\ &= \mathbb{V}[y_{ik_1} y_{jk_1}] \\ E_{ij}(k_1, k'_1, k_1, k'_2) &= \mathbb{V}[y_{i1} y_{j1}]. \end{aligned} \quad (10)$$

Moreover, the number of terms in the initial sum on $k_1, k'_1 \neq k_1, k_2, k'_2 \neq k_2$ satisfying the conditions of this case on the indices is:

$$\begin{aligned} & |\{(k_1, k'_1, k_2, k'_2) \in \llbracket 1, n \rrbracket^4 \mid k'_1 \neq k_1, k'_2 \neq k_2, |\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 1, k_1 = k_2\}| \\ & = n(n-1)(n-2). \end{aligned} \quad (11)$$

- If $k_1 = k'_2$ (or similarly $k'_1 = k_2$): then $k'_1 \neq k_2$ and $k_1 \neq k_2$. So, $E_{ij}(k_1, k'_1, k_2, k_1) = 0$.
- If $k'_1 = k'_2$: $E_{ij}(k_1, k'_1, k_2, k'_1) = 0$.

- If $|\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 2$:

- If $k_1 = k_2$ and $k'_1 = k'_2$: then $k_1 \neq k'_1$. So,

$$\begin{aligned} E_{ij}(k_1, k'_1, k_1, k'_1) &= \mathbb{E} \left[(y_{ik_1}(y_{jk_1} - y_{jk'_1}) - \lambda_{ij})^2 \right] \\ &= \mathbb{V}[y_{ik_1}y_{jk_1}] + \lambda_{ii}\lambda_{jj} \\ E_{ij}(k_1, k'_1, k_1, k'_1) &= \mathbb{V}[y_{i1}y_{j1}] + \lambda_{ii}\lambda_{jj}. \end{aligned} \quad (12)$$

Moreover, the number of terms in the initial sum on $k_1, k'_1 \neq k_1, k_2, k'_2 \neq k_2$ satisfying the conditions of this case on the indices is:

$$\begin{aligned} & |\{(k_1, k'_1, k_2, k'_2) \in \llbracket 1, n \rrbracket^4 \mid k'_1 \neq k_1, k'_2 \neq k_2, |\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 2, k_1 = k_2, k'_1 = k'_2\}| \\ & = n(n-1). \end{aligned} \quad (13)$$

- If $k_1 = k'_2$ and $k'_1 = k_2$: then $k_1 \neq k'_1$. So,

$$\begin{aligned} E_{ij}(k_1, k'_1, k'_1, k_1) &= \mathbb{E} \left[(y_{ik_1}(y_{jk_1} - y_{jk'_1}) - \lambda_{ij}) (y_{ik'_1}(y_{jk'_1} - y_{jk_1}) - \lambda_{ij}) \right] \\ E_{ij}(k_1, k'_1, k'_1, k_1) &= \lambda_{ij}^2. \end{aligned} \quad (14)$$

Moreover, the number of terms in the initial sum on $k_1, k'_1 \neq k_1, k_2, k'_2 \neq k_2$ satisfying the conditions of this case on the indices is:

$$\begin{aligned} & |\{(k_1, k'_1, k_2, k'_2) \in \llbracket 1, n \rrbracket^4 \mid k'_1 \neq k_1, k'_2 \neq k_2, |\{k_1, k'_1\} \cap \{k_2, k'_2\}| = 2, k_1 = k'_2, k'_1 = k_2\}| \\ & = n(n-1). \end{aligned} \quad (15)$$

Using the latter decomposition on (k_1, k'_1, k_2, k'_2) , we deduce:

$$\begin{aligned} \beta^2 &= \frac{1}{p} \sum_{i,j} \frac{1}{n^2(n-1)^2} [n(n-1)(n-2)\mathbb{V}[y_{i1}y_{j1}] + n(n-1)(\mathbb{V}[y_{i1}y_{j1}] + \lambda_{ii}\lambda_{jj}) + n(n-1)\lambda_{ij}^2] \\ \beta^2 &= \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{p+1}{n(n-1)}\mu^2 + \frac{1}{n(n-1)}\alpha^2 \text{ (Identities 2 \& 3)}. \end{aligned} \quad (16)$$

- μ^2 is bounded when n goes to infinity, so $\frac{p+1}{n(n-1)}\mu^2 \leq \frac{K_1+1}{n-1}\mu^2$ remains bounded too.
- α^2 is bounded when n goes to infinity, so $\frac{1}{n(n-1)}\alpha^2$ remains bounded too.

- Finally, following (Ledoit and Wolf, 2004) [10] proof of Lemma 3.1,

$$\begin{aligned}
\frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] &\leq \frac{1}{pn} \sum_{i,j} \mathbb{E}[y_{i1}^2 y_{j1}^2] \\
&\leq \frac{1}{pn} \sum_{i,j} \sqrt{\mathbb{E}[y_{i1}^4] \mathbb{E}[y_{j1}^4]} \quad (\text{Cauchy-Schwarz}) \\
&\leq \frac{1}{pn} \sum_{i,j} \sqrt[4]{\mathbb{E}[y_{i1}^8] \mathbb{E}[y_{j1}^8]} \quad (\text{Cauchy-Schwarz}) \\
&= \frac{p}{n} \left(\sum_i \frac{1}{p} \sqrt[4]{\mathbb{E}[y_{i1}^8]} \right)^2 \\
&\leq \frac{p}{n} \sqrt{\left(\sum_i \frac{1}{p} \mathbb{E}[y_{i1}^8] \right)} \quad (\text{Jensen for } \sqrt[4]{\cdot} \text{ concave}) \\
\frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] &\leq K_1 \sqrt{K_2}.
\end{aligned} \tag{17}$$

So, β^2 remains bounded as n goes to infinity. Finally, $\delta^2 = \alpha^2 + \beta^2$ is also bounded as n goes to infinity, which conclude the proof of the lemma.

6.3 Proof of Theorem 1

We have:

$$\begin{aligned}
\mu^2 + \theta^2 &= \left(\mathbb{E} \left[\frac{1}{p} \sum_i y_{i1}^2 \right] \right)^2 + \mathbb{V} \left[\frac{1}{p} \sum_i y_{i1}^2 \right] \\
&= \mathbb{E} \left[\left(\frac{1}{p} \sum_i y_{i1}^2 \right)^2 \right] \\
&= \frac{1}{p^2} \sum_{i,j} \mathbb{E} [y_{i1}^2 y_{j1}^2] \\
&= \frac{1}{p^2} \sum_{i,j} \mathbb{V} [y_{i1}y_{j1}] + \lambda_{ij}^2 \\
&= \frac{1}{p} (\alpha^2 + \mu^2) + \frac{1}{p^2} \sum_{i,j} \mathbb{V} [y_{i1}y_{j1}] \quad (\text{Identity 2}) \\
\mu^2 + \theta^2 &\leq \frac{1}{p} (\alpha^2 + \mu^2) + \sqrt{K_2} \quad (\text{Proof of Lemma 1}).
\end{aligned} \tag{18}$$

As α^2 and μ^2 remains bounded as n goes to infinity, so is θ^2 . And, from the proof of Lemma 1:

$$\beta^2 = \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{p+1}{n(n-1)} \mu^2 + \frac{1}{n(n-1)} \alpha^2. \tag{19}$$

So,

$$\begin{aligned}\beta^2 &= \frac{p}{n} \left(\mu^2 + \theta^2 - \frac{1}{p}(\alpha^2 + \mu^2) \right) + \frac{p+1}{n(n-1)}\mu^2 + \frac{1}{n(n-1)}\alpha^2 \\ &= \frac{p}{n} (\mu^2 + \theta^2) - \frac{1}{n(n-1)}\alpha^2 + \frac{p-n+2}{n(n-1)}\mu^2.\end{aligned}\tag{20}$$

As α^2 , μ^2 and $\frac{p}{n}$ remains bounded as n goes to infinity, immediately we have:

$$\beta^2 - \frac{p}{n} (\mu^2 + \theta^2) \xrightarrow{n \rightarrow \infty} 0.\tag{21}$$

6.4 Proof of Lemma 2

By linearity of the inner product, we trivially have: $\mathbb{E}[m] = \mu$. For the quartic mean convergence, we write:

$$\begin{aligned}\mathbb{E}[(m - \mu)^4] &= \mathbb{E} \left[\left(\frac{1}{p} \sum_i \frac{1}{n-1} \sum_k \left((y_{ik} - \bar{y}_i)^2 - \frac{n-1}{n} \lambda_{ii} \right) \right)^4 \right] \\ &= \mathbb{E} \left[\left(\frac{1}{n-1} \sum_k \frac{1}{p} \sum_i \left((y_{ik} - \bar{y}_i)^2 - \frac{n-1}{n} \lambda_{ii} \right) \right)^4 \right].\end{aligned}\tag{22}$$

Using Identity 1, we obtain:

$$\begin{aligned}\mathbb{E}[(m - \mu)^4] &= \frac{1}{(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \mathbb{E} \left[\left[\frac{1}{p} \sum_i \frac{1}{n} (y_{ik_1}^2 - \lambda_{ii}) - \frac{1}{n} (y_{ik_1} y_{ik'_1} - \delta_{k_1=k'_1} \lambda_{ii}) \right] \right. \\ &\quad \times \left[\frac{1}{p} \sum_i \frac{1}{n} (y_{ik_2}^2 - \lambda_{ii}) - \frac{1}{n} (y_{ik_2} y_{ik'_2} - \delta_{k_2=k'_2} \lambda_{ii}) \right] \\ &\quad \times \left[\frac{1}{p} \sum_i \frac{1}{n} (y_{ik_3}^2 - \lambda_{ii}) - \frac{1}{n} (y_{ik_3} y_{ik'_3} - \delta_{k_3=k'_3} \lambda_{ii}) \right] \\ &\quad \left. \times \left[\frac{1}{p} \sum_i \frac{1}{n} (y_{ik_4}^2 - \lambda_{ii}) - \frac{1}{n} (y_{ik_4} y_{ik'_4} - \delta_{k_4=k'_4} \lambda_{ii}) \right] \right].\end{aligned}\tag{23}$$

$$\begin{aligned}
\mathbb{E}[(m - \mu)^4] = & \frac{1}{(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \mathbb{E} \left[\prod_{s=1}^4 \left(\frac{1}{pn} \sum_i (y_{ik_s}^2 - \lambda_{ii}) \right) \right] \\
& - 4\mathbb{E} \left[\prod_{s=1}^3 \left(\frac{1}{pn} \sum_i (y_{ik_s}^2 - \lambda_{ii}) \right) \left(\frac{1}{pn} \sum_i (y_{ik_4} y_{ik'_4} - \delta_{k_4=k'_4} \lambda_{ii}) \right) \right] \\
& + 6\mathbb{E} \left[\prod_{s=1}^2 \left(\frac{1}{pn} \sum_i (y_{ik_s}^2 - \lambda_{ii}) \right) \prod_{t=3}^4 \left(\frac{1}{pn} \sum_i (y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii}) \right) \right] \\
& - 4\mathbb{E} \left[\left(\frac{1}{pn} \sum_i (y_{ik_1}^2 - \lambda_{ii}) \right) \prod_{t=2}^4 \left(\frac{1}{pn} \sum_i (y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii}) \right) \right] \\
& + \mathbb{E} \left[\prod_{t=1}^4 \left(\frac{1}{pn} \sum_i (y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii}) \right) \right].
\end{aligned} \tag{24}$$

If $k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4$ are all different, then the expectation in the sum equals 0. So there are at most $28n^7 + O(n^6)$ non-zero terms in the sum.

Now, let's find a bound of those expectations. Let's first note that, for all $(k, k') \in \llbracket 1, n \rrbracket^2$ and $i \in \llbracket 1, p \rrbracket$:

$$\begin{aligned}
\mathbb{E} \left[(y_{ik} y_{ik'} - \delta_{k=k'} \lambda_{ii})^4 \right] &= \mathbb{E} \left[(y_{ik} y_{ik'} - \mathbb{E}[y_{ik} y_{ik'}])^4 \right] = \\
& \mathbb{E}[(y_{ik} y_{ik'})^4] - 4\mathbb{E}[(y_{ik} y_{ik'})^3] \mathbb{E}[y_{ik} y_{ik'}] + 6\mathbb{E}[(y_{ik} y_{ik'})^2] \mathbb{E}[y_{ik} y_{ik'}]^2 - 3\mathbb{E}[y_{ik} y_{ik'}]^4.
\end{aligned} \tag{25}$$

If $k = k'$ then $\mathbb{E}[(y_{ik} y_{ik'})^3] \mathbb{E}[y_{ik} y_{ik'}] \geq 0$ and if $k \neq k'$, $\mathbb{E}[(y_{ik} y_{ik'})^3] \mathbb{E}[y_{ik} y_{ik'}] = 0 \geq 0$, so:

$$\begin{aligned}
\mathbb{E} \left[(y_{ik} y_{ik'} - \delta_{k=k'} \lambda_{ii})^4 \right] &\leq \mathbb{E}[(y_{ik} y_{ik'})^4] + 6\mathbb{E}[(y_{ik} y_{ik'})^2] \mathbb{E}[y_{ik} y_{ik'}]^2 \\
&\leq \mathbb{E}[(y_{ik} y_{ik'})^4] + 6\mathbb{E}[(y_{ik} y_{ik'})^4] && \text{(Cauchy-Schwarz)} \\
\mathbb{E} \left[(y_{ik} y_{ik'} - \delta_{k=k'} \lambda_{ii})^4 \right] &\leq 7\mathbb{E}[(y_{ik} y_{ik'})^4].
\end{aligned} \tag{26}$$

Back to the bound of our expectation, let $N \in \llbracket 1, 4 \rrbracket$, and we have:

$$\begin{aligned}
& \left| \mathbb{E} \left[\prod_{s=1}^N \left(\frac{1}{pn} \sum_i (y_{ik_s}^2 - \lambda_{ii}) \right) \prod_{t=N+1}^4 \left(\frac{1}{pn} \sum_i (y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii}) \right) \right] \right| \\
& \leq \frac{1}{n^4} \prod_{s=1}^N \sqrt[4]{\mathbb{E} \left[\left(\frac{1}{p} \sum_i (y_{ik_s}^2 - \lambda_{ii}) \right)^4 \right]} \prod_{t=N+1}^4 \sqrt[4]{\mathbb{E} \left[\left(\frac{1}{p} \sum_i (y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii}) \right)^4 \right]} \quad (2 \text{ Cauchy-Schwarz}) \\
& \leq \frac{1}{n^4} \prod_{s=1}^N \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[(y_{ik_s}^2 - \lambda_{ii})^4]} \prod_{t=N+1}^4 \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[(y_{ik_t} y_{ik'_t} - \delta_{k_t=k'_t} \lambda_{ii})^4]} \quad (\text{Jensen}) \\
& \leq \frac{7}{n^4} \prod_{s=1}^N \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[y_{ik_s}^8]} \prod_{t=N+1}^4 \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[y_{ik_t}^4 y_{ik'_t}^4]} \quad (\text{Previous remark}) \\
& \leq \frac{7}{n^4} \prod_{s=1}^N \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[y_{ik_s}^8]} \prod_{t=N+1}^4 \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[y_{ik_t}^8]} \quad (\text{Jensen if } k_t \neq k'_t) \\
& \leq \frac{7}{n^4} \prod_{s=1}^4 \sqrt[4]{\frac{1}{p} \sum_i \mathbb{E}[y_{ik_s}^8]} \\
& \leq \frac{7}{n^4} \frac{1}{p} \sum_i \mathbb{E}[y_{ik_1}^8] \\
& \leq \frac{7K_2}{n^4}.
\end{aligned} \tag{27}$$

So, in conclusion of this proof,

$$\mathbb{E}[(m - \mu)^4] \leq 7(1 + 4 + 6 + 4 + 1) \frac{28n^7 + O(n^6)}{n^4(n-1)^4} K_2 \xrightarrow{n \rightarrow \infty} 0. \tag{28}$$

6.5 Proof of Corollary 2

We have:

$$\begin{aligned}
\mathbb{V}[m^2] &= \mathbb{E}[m^4] - \mathbb{E}[m^2]^2 \\
&= \mathbb{E}[(m - \mu + \mu)^4] - \mathbb{E}[(m - \mu + \mu)^2]^2 \\
&= \mathbb{E}[(m - \mu)^4] + 4\mathbb{E}[(m - \mu)^3]\mu + 6\mathbb{E}[(m - \mu)^2]\mu^2 + \mu^4 - \mathbb{E}[(m - \mu)^2]^2 - 2\mathbb{E}[(m - \mu)^2]\mu^2 - \mu^4 \\
\mathbb{V}[m^2] &= \mathbb{E}[(m - \mu)^4] + 4\mathbb{E}[(m - \mu)^3]\mu + 4\mathbb{E}[(m - \mu)^2]\mu^2 - \mathbb{E}[(m - \mu)^2]^2.
\end{aligned} \tag{29}$$

And, from Lemma 2, $\mathbb{E}[(m - \mu)^4] \rightarrow 0$, and so goes for the smaller moments by Jensen inequality, and μ is bounded from Lemma 1. So, $\mathbb{V}[m^2] \rightarrow 0$.

6.6 Proof of Lemma 3

6.6.1 Preliminary combinatorial result

Let $K \in \mathbb{N}^*$, and K indices $(k_1, \dots, k_K) \in \llbracket 1, n \rrbracket^K$.

Let's associate a graph with K vertices $\mathcal{V} = \{1, \dots, K\}$ to this set of indices. The set of edges \mathcal{E} is built as following: there is an edge between the node $a \in \mathcal{V}$ and $b \in \mathcal{V}$, $a \neq b$ (we don't allow self-loops), if the corresponding indices are equal, i.e if $k_a = k_b$. We finally define our graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Proposition 1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a graph with K vertices generated from some indices $(k_1^{(0)}, \dots, k_K^{(0)}) \in \llbracket 1, n \rrbracket^K$ with the procedure described previously. Suppose \mathcal{G} has $C \in \llbracket 1, K \rrbracket$ connected components. Then, there are $\prod_{i=0}^{C-1} (n-i)$ set of indices $(k_1, \dots, k_K) \in \llbracket 1, n \rrbracket^K$ which have the associated graph \mathcal{G} .*

For each node $v \in \mathcal{V}$, v belongs to a unique connected component that we denote $c(v) \in \llbracket 1, C \rrbracket$. Then, the function:

$$x \in \{x \in \llbracket 1, n \rrbracket^C, x_1, \dots, x_C \text{ all different}\} \mapsto (x_{c(1)}, \dots, x_{c(K)})$$

is a bijection between $\{x \in \llbracket 1, n \rrbracket^C, x_1, \dots, x_C \text{ all different}\}$ and $\{(k_1, \dots, k_K) \in \llbracket 1, n \rrbracket^K \text{ which have the associated graph } \mathcal{G}\}$. Immediately, we deduce that its cardinal is equal to $\prod_{i=0}^{C-1} (n-i)$.

6.6.2 Proof of Lemma 3

From the proof of Lemma 3.3 in (Ledoit and Wolf, 2004) [10], we have:

$$d^2 - \delta^2 = -(m - \mu)^2 + 2\mu(\mu - m) + (\|S\|^2 - \mathbb{E}[\|S\|^2]). \quad (30)$$

The first two terms converge to 0 in quadratic mean thanks to Lemma 2. Let's show that the last term converges to 0 in quadratic mean too, i.e $\mathbb{V}[\|S\|^2] \xrightarrow{n \rightarrow \infty} 0$.

Decomposing $\|S\|^2$, we have:

$$\begin{aligned}
\|S\|^2 &= \frac{1}{p} \left\| \frac{1}{n-1} \sum_k (X_{\cdot k} - \bar{X}) (X_{\cdot k} - \bar{X})^T \right\|^2 \\
&= \frac{1}{p} \left\| \frac{1}{n-1} \sum_k (Y_{\cdot k} - \bar{Y}) (Y_{\cdot k} - \bar{Y})^T \right\|^2 \\
&= \frac{1}{p(n-1)^2} \sum_{i,j} \left(\sum_k (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) \right)^2 \\
&= \frac{1}{p(n-1)^2} \sum_{i,j} \left(\frac{1}{n} \sum_{k,k'} y_{ik}(y_{jk} - y_{jk'}) \right)^2 \quad (\text{Identity 1}) \\
&= \frac{1}{pn^2(n-1)^2} \sum_{i,j} \left(\sum_{k_1, k'_1 \neq k_1} y_{ik_1}(y_{jk_1} - y_{jk'_1}) \right) \left(\sum_{k_2, k'_2 \neq k_2} y_{ik_2}(y_{jk_2} - y_{jk'_2}) \right) \\
&= \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \sum_{k_2 = k'_1, k'_2 \neq k_2} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} (y_{jk_1} y_{jk_2} - y_{jk_1} y_{jk'_2} - y_{jk'_1} y_{jk_2} + y_{jk'_1} y_{jk'_2}) \\
&\quad + \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \sum_{k_2 \neq k'_1, k'_2 \neq k_2} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} (y_{jk_1} y_{jk_2} - y_{jk_1} y_{jk'_2} - y_{jk'_1} y_{jk_2} + y_{jk'_1} y_{jk'_2}) \\
&= \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \sum_{k_2 = k'_1, k'_2 \neq k_2} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} (y_{jk_1} y_{jk'_1} - y_{jk_1} y_{jk'_2} - y_{jk'_1}^2 + y_{jk'_1} y_{jk'_2}) \\
&\quad + \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \sum_{k_2 \neq k'_1, k'_2 \neq k_2} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} (y_{jk_1} y_{jk_2} - y_{jk_1} y_{jk'_2} - y_{jk'_1} y_{jk_2} + y_{jk'_1} y_{jk'_2}) \\
&= \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \left(\left(\frac{1}{p} \sum_i y_{ik_1} y_{ik'_1} \right)^2 - \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} (y_{jk_1} y_{jk'_2} + y_{jk'_1}^2 - y_{jk'_1} y_{jk'_2}) \right) \\
&\quad + \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \left(\left(\frac{1}{p} \sum_i y_{ik_1} y_{ik_2} \right)^2 - \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} (y_{jk_1} y_{jk'_2} + y_{jk'_1} y_{jk_2} - y_{jk'_1} y_{jk'_2}) \right).
\end{aligned} \tag{31}$$

$$\|S\|^2 = \frac{p}{n^2(n-1)} \sum_{k_1, k'_1 \neq k_1} \left(\frac{1}{p} \sum_i y_{ik_1} y_{ik'_1} \right)^2 \quad (t1)$$

$$- \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2} \quad (t2)$$

$$- \frac{p}{n^2(n-1)} \sum_{k_1, k'_1 \neq k_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk'_1}^2 \quad (t3)$$

$$+ \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk'_1} y_{jk'_2} \quad (t4)$$

$$+ \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \left(\frac{1}{p} \sum_i y_{ik_1} y_{ik_2} \right)^2 \quad (t5) \quad (32)$$

$$- \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk_1} y_{jk'_2} \quad (t6)$$

$$- \frac{p}{n^2(n-1)} \sum_{k_1, k'_1 \neq k_1, k_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk'_1} y_{jk_2} \quad (t7)$$

$$+ \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk'_1} y_{jk'_2} \quad (t8).$$

We notice that, re-injecting the missing terms into the sum:

$$(t2) = - \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}$$

$$(t2) = - \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1, k'_2} \frac{1}{p^2} \sum_{i,j} \left(y_{ik_1} y_{ik'_1} y_{jk_1} \left(y_{jk'_2} - \frac{1}{n} y_{jk'_1} \right) - \frac{1}{n} y_{ik_1}^2 y_{jk_1} \left(y_{jk'_2} - \frac{1}{n} y_{jk_1} \right) \right). \quad (33)$$

Similarly, we have:

$$(t4) = \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk'_1} y_{jk'_2}$$

$$(t4) = \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1, k'_2} \frac{1}{p^2} \sum_{i,j} \left(y_{ik_1} y_{ik'_1} y_{jk_1} \left(y_{jk'_2} - \frac{1}{n} y_{jk_1} \right) - \frac{1}{n} y_{ik_1}^2 y_{jk_1} \left(y_{jk'_2} - \frac{1}{n} y_{jk_1} \right) \right). \quad (34)$$

So,

$$(t2) + (t4) = \frac{p}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk_1} (y_{jk'_1} - y_{jk_1}) \quad (35)$$

$$(t2) + (t4) = \frac{1}{n-1} ((1) + (3)).$$

And,

$$\begin{aligned}
(t5) &= \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \left(\frac{1}{p} \sum_i y_{ik_1} y_{ik_2} \right)^2 \\
&= \frac{p(n-2)}{n^2(n-1)} \sum_{k_1, k_2 \neq k_1} \left(\frac{1}{p} \sum_i y_{ik_1} y_{ik_2} \right)^2 + \frac{p}{n^2} \sum_{k_1} \left(\frac{1}{p} \sum_i y_{ik_1}^2 \right)^2 \\
(t5) &= (n-2) \times (t1) + \frac{p}{n^2} \sum_{k_1} \left(\frac{1}{p} \sum_i y_{ik_1}^2 \right)^2.
\end{aligned} \tag{36}$$

Moreover,

$$\begin{aligned}
(t6) &= -\frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk_1} y_{jk'_2} \\
&= -\frac{p(n-2)}{n^2(n-1)^2} \sum_{\substack{k_1, k_2 \neq k_1 \\ k'_2 \neq k_2}} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk_1} y_{jk'_2} - \frac{p}{n^2(n-1)} \sum_{\substack{k_1 \\ k'_2 \neq k_1}} \frac{1}{p^2} \sum_{i,j} y_{ik_1}^2 y_{jk_1} y_{jk'_2} \\
(t6) &= (n-2) \times (t2) + (t3).
\end{aligned} \tag{37}$$

And,

$$\begin{aligned}
(t7) &= -\frac{p}{n^2(n-1)} \sum_{k_1, k'_1 \neq k_1, k_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk'_1} y_{jk_2} \\
&= -\frac{p}{n^2(n-1)} \sum_{k_2, k'_1 \neq k_2, k_1 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk'_1} y_{jk_2} \text{ (re-indexing)} \\
(t7) &= (n-1) \times (t2).
\end{aligned} \tag{38}$$

Finally, we obtain,

$$\begin{aligned}
\|S\|^2 &= \frac{p}{n^2} \sum_{k_1} \left(\frac{1}{p} \sum_i y_{ik_1}^2 \right)^2 & (t1') \\
&+ \frac{p(n^2 - 2n + 2)}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \left(\frac{1}{p} \sum_i y_{ik_1} y_{ik'_1} \right)^2 & (t2') \\
&- \frac{p(2n-3)}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1, k'_2 \neq k'_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2} & (t3') \\
&- \frac{p(2n-1)}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik'_1} y_{jk_1}^2 & (t4') \\
&+ \frac{p}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \frac{1}{p^2} \sum_{i,j} y_{ik_1} y_{ik_2} y_{jk'_1} y_{jk'_2} & (t5').
\end{aligned} \tag{39}$$

It is sufficient to show that the variance of each of the 5 term converges to 0 as n goes to infinity in order to prove that $\mathbb{V}[\|S\|^2]$ converges to 0 as n goes to infinity.

- $\mathbb{V}[(t1')] \rightarrow 0$ immediately from the proof of Lemma 3.3 in (Ledoit and Wolf, 2004) [10].
- $\mathbb{V}[(t2')] \rightarrow 0$ immediately from the proof of Lemma 3.3 in (Ledoit and Wolf, 2004) [10].
- Let's prove that $\mathbb{V}[(t3')] \rightarrow 0$.

$$\mathbb{V}[(t3')] = \frac{p^2(2n-3)^2}{n^4(n-1)^4} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k'_2 \neq k'_1}} \sum_{\substack{k_3, k'_3 \neq k_3 \\ k'_4 \neq k'_3}} \sum_{i, j, l, m} \frac{1}{p^4} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}). \quad (40)$$

Let $(k_1, k'_1, k'_2, k_3, k'_3, k'_4) \in \llbracket 1, n \rrbracket^6$ respecting the conditions given in the sums. Suppose there exists $(i, j, l, m) \in \llbracket 1, p \rrbracket^4$ such that $\text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) \neq 0$. Let consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ built from $(k_1, k'_1, k'_2, k_3, k'_3, k'_4)$ following the procedure described in the preliminary combinatorial result.

As for some $(i, j, l, m) \in \llbracket 1, p \rrbracket^4$ we have $\text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) \neq 0$, by independence, the nodes 1', 2', 3', and 4' can't be isolated. As a consequence, \mathcal{G} has at least 2 edges.

– When the graph \mathcal{G} has only 2 edges, we have either of the following conditions, that we denote the (*) conditions:

- * $(k'_1 = k'_3) \wedge (k'_2 = k'_4) \wedge (k_1 \neq k_3) \wedge (k_1 \neq k'_2) \wedge (k_3 \neq k'_4)$,
- * $(k'_1 = k'_4) \wedge (k'_3 = k'_2) \wedge (k_1 \neq k_3) \wedge (k_1 \neq k'_2) \wedge (k_3 \neq k'_4)$.

Note that the case where $(k'_1 = k'_2) \wedge (k'_3 = k'_4)$ is impossible due to the constraints on the indices in the sum. In the first case, we have:

$$\begin{aligned} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) &= \mathbb{E} [y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2} y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1} y_{lk_3} y_{mk_3} y_{ik'_1} y_{lk'_3} y_{jk'_2} y_{mk'_4}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1} y_{lk_3} y_{mk_3} y_{ik'_1} y_{lk'_3} y_{jk'_2} y_{mk'_4}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1}] \mathbb{E} [y_{lk_3} y_{mk_3}] \mathbb{E} [y_{ik'_1} y_{lk'_3}] \mathbb{E} [y_{jk'_2} y_{mk'_4}] \\ \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) &= \lambda_{ij} \lambda_{lm} \lambda_{il} \lambda_{jm}. \end{aligned} \quad (41)$$

And in the second case, we have:

$$\begin{aligned} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) &= \mathbb{E} [y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2} y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1} y_{lk_3} y_{mk_3} y_{ik'_1} y_{mk'_4} y_{jk'_2} y_{lk'_3}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1} y_{lk_3} y_{mk_3} y_{ik'_1} y_{mk'_4} y_{jk'_2} y_{lk'_3}] \\ &= \mathbb{E} [y_{ik_1} y_{jk_1}] \mathbb{E} [y_{lk_3} y_{mk_3}] \mathbb{E} [y_{ik'_1} y_{mk'_4}] \mathbb{E} [y_{jk'_2} y_{lk'_3}] \\ \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) &= \lambda_{ij} \lambda_{lm} \lambda_{im} \lambda_{jl}. \end{aligned} \quad (42)$$

Using the fact that $i \neq j \implies \lambda_{ij} = 0$, in both cases we have:

$$\begin{aligned}
\sum_{i,j,l,m} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) &= \sum_i \lambda_{ii}^4 \\
&= p \sum_i \frac{1}{p} \mathbb{E}[y_{i1}^2]^4 \\
&\leq p \sum_i \frac{1}{p} \mathbb{E}[y_{i1}^8] \\
&\leq p \sum_i \frac{1}{p} \mathbb{E}[y_{i1}^8]
\end{aligned} \tag{43}$$

$$\sum_{i,j,l,m} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) \leq pK_2.$$

Under the conditions (*), \mathcal{G} has exactly 4 connected components. So, from the preliminary combinatorial result, there are $2n(n-1)(n-2)(n-3)$ different 6-uples of indices k respecting the (*) conditions. We finally have:

$$\begin{aligned}
&\frac{p^2(2n-3)^2}{n^4(n-1)^4} \sum_{\substack{(k_1, k'_1, k'_2, k_3, k'_3, k'_4) \\ \text{under (*) conditions}}} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_2}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_4}) \\
&\leq \frac{p^2(2n-3)^2}{n^4(n-1)^4} \times 2n(n-1)(n-2)(n-3) \frac{1}{p^4} pK_2 \\
&\leq 2 \frac{(2n-3)^2}{pn^3(n-1)} K_2 \\
&\leq \frac{8}{n(n-1)} K_2 \\
&\xrightarrow{n \rightarrow \infty} 0.
\end{aligned} \tag{44}$$

- Otherwise, \mathcal{G} has 3 edges or more: we denote it as the (**) condition. As it has only 6 vertices, there are at most 3 connected component. So, from the preliminary combinatorial result, there are $n(n-1)(n-2) + n(n-1) + n = n((n-1)^2 + n)$ different 6-uples of indices k that have \mathcal{G} as associated graph. Moreover, there are a finite number $N \in \mathbb{N}$ (independent of n) of graphs with 6 vertices that have at most 3 connected components.

So, there are at most $Nn((n-1)^2 + n)$ different 6-uples of indices k such that the associated graph \mathcal{G} has 3 edges or more.

And we have:

$$\begin{aligned}
& \sum_{i,j,l,m} \frac{1}{p^4} |\text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_1}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_3})| \\
& \leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{V}[y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_1}] \mathbb{V}[y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_3}]} \\
& \leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{E}[y_{ik_1}^2 y_{ik'_1}^2 y_{jk_1}^2 y_{jk'_1}^2] \mathbb{E}[y_{lk_3}^2 y_{lk'_3}^2 y_{mk_3}^2 y_{mk'_3}^2]} \\
& \leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{E}[y_{ik_1}^8] \mathbb{E}[y_{ik'_1}^8] \mathbb{E}[y_{jk_1}^8] \mathbb{E}[y_{jk'_1}^8] \mathbb{E}[y_{lk_3}^8] \mathbb{E}[y_{lk'_3}^8] \mathbb{E}[y_{mk_3}^8] \mathbb{E}[y_{mk'_3}^8]} \\
& \leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{E}[y_{i1}^8] \mathbb{E}[y_{j1}^8] \mathbb{E}[y_{l1}^8] \mathbb{E}[y_{m1}^8]} \\
& \leq \left(\sum_i \frac{1}{p} \sqrt{\mathbb{E}[y_{i1}^8]} \right)^4 \\
& \leq \sum_i \frac{1}{p} \mathbb{E}[y_{i1}^8] \\
& \leq K_2.
\end{aligned} \tag{45}$$

So, using both of the previous inequalities,

$$\begin{aligned}
& \frac{p^2(2n-3)^2}{n^4(n-1)^4} \sum_{\substack{(k_1, k'_1, k_2, k'_2, k_3, k'_3) \\ \text{under (**) condition}}} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1} y_{jk'_1}, y_{lk_3} y_{lk'_3} y_{mk_3} y_{mk'_3}) \\
& \leq \frac{p^2(2n-3)^2}{n^4(n-1)^4} \times Nn((n-1)^2 + n)K_2 \\
& \leq \frac{(2n-3)^2}{n^2(n-1)^4} \times Nn((n-1)^2 + n)K_1^2 K_2 \\
& \xrightarrow{n \rightarrow \infty} 0.
\end{aligned} \tag{46}$$

So, from both previous cases, we immediately have:

$$\mathbb{V}[(t3')] \xrightarrow{n \rightarrow \infty} 0. \tag{47}$$

- Let's prove that $\mathbb{V}[(t4')] \rightarrow 0$.

$$\mathbb{V}[(t4')] = \frac{p^2(2n-1)^2}{n^4(n-1)^4} \sum_{k_1} \sum_{k'_1 \neq k_1} \sum_{k_3} \sum_{k'_3 \neq k_3} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1}^2, y_{lk_3} y_{lk'_3} y_{mk_3}^2). \tag{48}$$

Similarly as the previous case, let $(k_1, k'_1, k_3, k'_3) \in \llbracket 1, n \rrbracket^4$ respecting the conditions given in the sums. Suppose there exists $(i, j, l, m) \in \llbracket 1, p \rrbracket^4$ such that $\text{Cov}(y_{ik_1} y_{ik'_1} y_{jk_1}^2, y_{lk_3} y_{lk'_3} y_{mk_3}^2) \neq 0$. Let consider the associated graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

By independence, the nodes 1 and 3 can't be isolated. As a consequence, \mathcal{G} has at most 3 connected components.

Denoting N the number of graphs with at most 3 connected components and 4 vertices,

from the preliminary combinatorial result, there are at most $Nn((n-1)^2+1)$ different indices $(k_1, k'_1, k_3, k'_3) \in \llbracket 1, n \rrbracket^4$ such that $\sum_{i,j,l,m} \text{Cov}(y_{ik_1}y_{ik'_1}y_{jk'_1}^2, y_{lk_3}y_{lk'_3}y_{mk'_3}^2) \neq 0$.
Moreover,

$$\begin{aligned}
\sum_{i,j,l,m} \frac{1}{p^4} |\text{Cov}(y_{ik_1}y_{ik'_1}y_{jk'_1}^2, y_{lk_3}y_{lk'_3}y_{mk'_3}^2)| &\leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{V}[y_{jk'_1}^2 y_{ik_1} y_{ik'_1}] \mathbb{V}[y_{mk'_3}^2 y_{lk_3} y_{lk'_3}]} \\
&\leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt{\mathbb{E}[y_{jk'_1}^4 y_{ik_1}^2 y_{ik'_1}^2] \mathbb{E}[y_{mk'_3}^4 y_{lk_3}^2 y_{lk'_3}^2]} \\
&\leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt[4]{\mathbb{E}[y_{jk'_1}^8] \sqrt{\mathbb{E}[y_{ik_1}^8] \mathbb{E}[y_{ik'_1}^8] \mathbb{E}[y_{mk_3}^8] \mathbb{E}[y_{lk_3}^8] \mathbb{E}[y_{lk'_3}^8]}} \\
&\leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt[4]{\mathbb{E}[y_{j1}^8] \sqrt{\mathbb{E}[y_{i1}^8] \mathbb{E}[y_{i1}^8] \mathbb{E}[y_{m1}^8] \mathbb{E}[y_{l1}^8] \mathbb{E}[y_{l1}^8]}} \\
&\leq \sum_{i,j,l,m} \frac{1}{p^4} \sqrt[4]{\mathbb{E}[y_{i1}^8] \mathbb{E}[y_{j1}^8] \mathbb{E}[y_{l1}^8] \mathbb{E}[y_{m1}^8]} \\
&\leq \left(\sum_i \frac{1}{p} \sqrt[4]{\mathbb{E}[y_{i1}^8]} \right)^4 \\
&\leq \sum_i \frac{1}{p} \mathbb{E}[y_{i1}^8]
\end{aligned}$$

$$\sum_{i,j,l,m} \frac{1}{p^4} |\text{Cov}(y_{ik_1}y_{ik'_1}y_{jk'_1}^2, y_{lk_3}y_{lk'_3}y_{mk'_3}^2)| \leq K_2. \tag{49}$$

So,

$$\begin{aligned}
\mathbb{V}[(t4')] &\leq \frac{p^2(2n-1)^2}{n^4(n-1)^4} Nn((n-1)^2+1)K_2 \\
&\quad \frac{(2n-1)^2}{n(n-1)^4} N((n-1)^2+1)K_1^2K_2 \\
\mathbb{V}[(t4')] &\xrightarrow{n \rightarrow \infty} 0.
\end{aligned} \tag{50}$$

- Let's prove that $\mathbb{V}[(t5')] \rightarrow 0$.

$$\mathbb{V}[(t5')] = \frac{p^2}{n^4(n-1)^4} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2 \neq k'_1, k'_2 \neq k_2}} \sum_{\substack{k_3, k'_3 \neq k_3 \\ k_4 \neq k'_3, k'_4 \neq k_4}} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov}(y_{ik_1}y_{jk'_1}y_{ik_2}y_{jk'_2}, y_{lk_3}y_{mk'_3}y_{lk_4}y_{mk'_4}). \tag{51}$$

Let $(k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4) \in \llbracket 1, n \rrbracket^8$ and $(i, j, l, m) \in \llbracket 1, p \rrbracket^4$ such that

$\text{Cov}(y_{ik_1}y_{jk'_1}y_{ik_2}y_{jk'_2}, y_{lk_3}y_{mk'_3}y_{lk_4}y_{mk'_4}) \neq 0$.

Let's consider the associated graph \mathcal{G} , built following the procedure described in the preliminary combinatorial result.

For each node $a \in \mathcal{V}$, the connected component of \mathcal{G} containing a contains at least 2 nodes (a and at least an other one). Otherwise, a is isolated, which means that k_a is different from

all the other indices, and so by independence, $\text{Cov}(y_{ik_1}y_{jk'_1}y_{ik_2}y_{jk'_2}, y_{lk_3}y_{mk'_3}y_{lk_4}y_{mk'_4}) = 0$, which is in contradiction with our hypothesis.

As each connected component contains at least 2 nodes, there are at most 4 connected components in \mathcal{G} .

So, from the preliminary combinatorial result, there are at most $n(n-1)(n-2)(n-3) + n(n-1)(n-2) + n(n-1) + n = n((n-1)(n-2)^2 + n)$ different $(k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4) \in \llbracket 1, n \rrbracket^8$ which have the same associated graph \mathcal{G} .

Moreover, there is a finite number $N \in \mathbb{N}$ (independent of n) of graphs with 8 nodes and at least 4 connected components.

Finally, combining the previous 2 counting, we deduce that there are at most $Nn((n-1)(n-2)^2 + n)$ terms such that $\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov}(y_{ik_1}y_{jk'_1}y_{ik_2}y_{jk'_2}, y_{lk_3}y_{mk'_3}y_{lk_4}y_{mk'_4}) \neq 0$.

As previously, we have also,

$$\sum_{i,j,l,m} \frac{1}{p^4} |\text{Cov}(y_{ik_1}y_{jk'_1}y_{ik_2}y_{jk'_2}, y_{lk_3}y_{mk'_3}y_{lk_4}y_{mk'_4})| \leq K_2. \quad (52)$$

Finally,

$$\begin{aligned} \mathbb{V}[(t5')] &\leq \frac{p^2}{n^4(n-1)^4} Nn((n-1)(n-2)^2 + n) K_2 \\ &\leq \frac{N((n-1)(n-2)^2 + n)}{n(n-1)^3} K_1^2 K_2 \\ \mathbb{V}[(t5')] &\xrightarrow[n \rightarrow \infty]{} 0. \end{aligned} \quad (53)$$

We showed that each of the 5 terms of $\|S\|^2$ have a variance that converges to 0 as n goes to infinity.

So,

$$\mathbb{V}[\|S\|^2] \xrightarrow[n \rightarrow \infty]{} 0. \quad (54)$$

Which concludes the proof of the first part of the lemma:

$$d^2 - \delta^2 \xrightarrow[n \rightarrow \infty]{q.m} 0. \quad (55)$$

Finally, by property of the expectation, we have that $\mathbb{E}[(d^2 - \mathbb{E}[d^2])^2] \leq \mathbb{E}[(d^2 - \delta^2)^2]$, so it follows that:

$$d^2 - \mathbb{E}[d^2] \xrightarrow[n \rightarrow \infty]{q.m} 0. \quad (56)$$

6.7 Proof of Lemma 4

$$\begin{aligned}
\mathbb{E}[\bar{b}^2] &= \mathbb{E} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} (\tilde{x}_{\cdot k})^t - S \right\|^2 \right] \\
&= \mathbb{E} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} (\tilde{x}_{\cdot k})^t - \Sigma + \Sigma - S \right\|^2 \right] \\
&= \mathbb{E} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} (\tilde{x}_{\cdot k})^t - \Sigma \right\|^2 \right] - \frac{1}{n} \mathbb{E} [\|S - \Sigma\|^2] \\
&= \mathbb{E} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{y}_{\cdot k} (\tilde{y}_{\cdot k})^t - \Lambda \right\|^2 \right] - \frac{1}{n} \beta^2 \\
&= \frac{1}{pn^2} \sum_{i,j} \sum_k \mathbb{E} \left[\left(\frac{n}{n-1} (y_{ik} - \bar{y}_i) (y_{jk} - \bar{y}_j) - \lambda_{ij} \right)^2 \right] - \frac{1}{n} \beta^2 \\
&= \frac{n}{p(n-1)^2} \sum_{i,j} \mathbb{E} \left[\left((y_{i1} - \bar{y}_i) (y_{j1} - \bar{y}_j) - \frac{n-1}{n} \lambda_{ij} \right)^2 \right] - \frac{1}{n} \beta^2 \\
&= \frac{n}{p(n-1)^2} \sum_{i,j} \mathbb{E} \left[\left((y_{i1} y_{j1} - \lambda_{ij}) - \left(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij} \right) - \left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) + \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right)^2 \right] - \frac{1}{n} \beta^2 \\
\mathbb{E}[\bar{b}^2] &= \frac{n}{p(n-1)^2} \sum_{i,j} \left(\mathbb{V}[y_{i1} y_{j1}] + \mathbb{V}[\bar{y}_i y_{j1}] + \mathbb{V}[\bar{y}_j y_{i1}] + \mathbb{V}[\bar{y}_i \bar{y}_j] \right. \\
&\quad - 2\mathbb{E} \left[(y_{i1} y_{j1} - \lambda_{ij}) \left(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij} \right) \right] - 2\mathbb{E} \left[(y_{i1} y_{j1} - \lambda_{ij}) \left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) \right] \\
&\quad + 2\mathbb{E} \left[(y_{i1} y_{j1} - \lambda_{ij}) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] + 2\mathbb{E} \left[\left(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) \right] \\
&\quad \left. - 2\mathbb{E} \left[\left(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] - 2\mathbb{E} \left[\left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \right) - \frac{1}{n} \beta^2.
\end{aligned} \tag{57}$$

We simplify each of the terms into manageable quantities:

- $\mathbb{V}[\bar{y}_i y_{j1}] = \frac{1}{n^2} \mathbb{V}[y_{i1} y_{j1}] + \frac{n-1}{n^2} \lambda_{ii} \lambda_{jj},$
- $\mathbb{V}[\bar{y}_j y_{i1}] = \frac{1}{n^2} \mathbb{V}[y_{i1} y_{j1}] + \frac{n-1}{n^2} \lambda_{ii} \lambda_{jj},$
- $\mathbb{V}[\bar{y}_i \bar{y}_j] = \frac{1}{n^3} (\mathbb{V}[y_{i1} y_{j1}] + (n-1)(\lambda_{ii} \lambda_{jj} + \lambda_{ij}^2)),$
- $\mathbb{E}[(y_{i1} y_{j1} - \lambda_{ij}) (\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij})] = \frac{1}{n} \mathbb{V}[y_{i1} y_{j1}],$
- $\mathbb{E}[(y_{i1} y_{j1} - \lambda_{ij}) (\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij})] = \frac{1}{n} \mathbb{V}[y_{i1} y_{j1}],$
- $\mathbb{E}[(y_{i1} y_{j1} - \lambda_{ij}) (\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij})] = \frac{1}{n^2} \mathbb{V}[y_{i1} y_{j1}],$
- $\mathbb{E}[(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij}) (\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij})] = \frac{1}{n^2} \mathbb{V}[y_{i1} y_{j1}] + \frac{n-1}{n^2} \lambda_{ij}^2,$
- $\mathbb{E}[(\bar{y}_i y_{j1} - \frac{1}{n} \lambda_{ij}) (\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij})] = \frac{1}{n^3} (\mathbb{V}[y_{i1} y_{j1}] + (n-1)(\lambda_{ii} \lambda_{jj} + \lambda_{ij}^2)),$

- $\mathbb{E} \left[\left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] = \frac{1}{n^3} \left(\mathbb{V}[y_{i1} y_{j1}] + (n-1)(\lambda_{ii} \lambda_{jj} + \lambda_{ij}^2) \right).$

Adding up all the terms, we obtain:

$$\begin{aligned} \mathbb{E}[\bar{b}^2] &= \frac{1}{p(n-1)} \sum_{i,j} \left(\frac{n^2 - 3n + 3}{n^2} \mathbb{V}[y_{i1} y_{j1}] + \frac{2n-3}{n^2} \lambda_{ii} \lambda_{jj} + \frac{2n-3}{n^2} \lambda_{ij}^2 \right) - \frac{1}{n} \beta^2 \\ &= \frac{1}{p(n-1)} \left(\frac{n^2 - 3n + 3}{n^2} \sum_{i,j} \mathbb{V}[y_{i1} y_{j1}] + p^2 \frac{2n-3}{n^2} \langle \Sigma, I \rangle^2 + p \frac{2n-3}{n^2} \|\Sigma\|^2 \right) - \frac{1}{n} \beta^2 \quad (58) \\ \mathbb{E}[\bar{b}^2] &= \frac{1}{p(n-1)} \left(\frac{n^2 - 3n + 3}{n^2} \sum_{i,j} \mathbb{V}[y_{i1} y_{j1}] + p^2 \frac{2n-3}{n^2} \mu^2 + p \frac{2n-3}{n^2} (\alpha^2 + \mu^2) \right) - \frac{1}{n} \beta^2. \end{aligned}$$

From the proof of Lemma 1, we have:

$$\frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1} y_{j1}] = \beta^2 - \frac{p+1}{n(n-1)} \mu^2 - \frac{1}{n(n-1)} \alpha^2. \quad (59)$$

Denoting: $\gamma_n = \frac{n(n-1)}{n^2-3n+3}$, $\lambda_n = \frac{n^2(n-2)}{(n-1)(n^2-3n+3)}$, $c_0 = \frac{1}{\gamma_n} - \frac{1}{n} - \frac{\lambda_n}{\gamma_n n^2}$, $c_1 = \frac{\lambda_n}{\gamma_n n^2}$, $c_2 = (p+1)c_1$, we notice that:

$$\begin{aligned} c_0 + c_1 &= \frac{n^2 - 3n + 3}{n(n-1)} - \frac{1}{n}, \\ c_1 &= -\frac{n^2 - 3n + 3}{n^2(n-1)^2} + \frac{2n-3}{n^2(n-1)}, \\ c_2 &= -\frac{(n^2 - 3n + 3)(p+1)}{n^2(n-1)^2} + \frac{(2n-3)(p+1)}{n^2(n-1)}. \end{aligned} \quad (60)$$

We obtain from the previous lines: $\mathbb{E}[\bar{b}^2] = (c_0 + c_1)\beta^2 + c_1\alpha^2 + c_2\mu^2$. So, we can conclude using $\delta^2 = \alpha^2 + \beta^2$: $\mathbb{E}[\bar{b}^2] = c_0\beta^2 + c_1\delta^2 + c_2\mu^2$.

6.8 Proof of Lemma 5

We compute the expectations of m^2 and d^2 .

$$\mathbb{E}[m^2] = \mu^2 + \mathbb{V}[m]. \quad (61)$$

And,

$$\begin{aligned} \mathbb{E}[d^2] &= \mathbb{E}[\|S - mI\|^2] \\ &= \mathbb{E}[\|S - \mu I\|^2 + \|\mu I - mI\|^2 + 2\langle S - \mu I, \mu I - mI \rangle] \\ &= \delta^2 + \mathbb{V}[m] - 2\mathbb{V}[m] \\ \mathbb{E}[d^2] &= \delta^2 - \mathbb{V}[m]. \end{aligned} \quad (62)$$

Moreover, from Lemma 4, we have: $\mathbb{E}[\bar{b}^2] = c_0\beta_n^2 + c_1\delta_n^2 + c_2\mu_n^2$.

So, combining the last 3 equations, we obtain: $\mathbb{E}[\bar{b}^2] = c_0\beta^2 + c_1\mathbb{E}[d^2] + c_2\mathbb{E}[m^2] + (c_1 - c_2)\mathbb{V}[m]$.

6.9 Proof of Lemma 6

$$\begin{aligned} \bar{b}^2 - \mathbb{E}[\bar{b}^2] &= \left[\left(\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - \Sigma \right\|^2 \right) - \mathbb{E}[\bar{b}^2] \right] \\ &\quad + \left[\left(\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - S \right\|^2 \right) - \left(\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - \Sigma \right\|^2 \right) \right]. \end{aligned} \quad (63)$$

Firstly, we want to show that the variance of the first term converges to 0 as n goes to infinity. Following the decomposition developed in Lemma 4, we have:

$$\begin{aligned} &\mathbb{V} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - \Sigma \right\|^2 \right] \\ &= \mathbb{V} \left[\frac{1}{(n-1)^2} \sum_k \left\| \tilde{y}_{\cdot k} \tilde{y}_{\cdot k}^T - \frac{n-1}{n} \Lambda \right\|^2 \right] \\ &= \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left((y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) - \frac{n-1}{n} \lambda_{ij} \right)^2 \right] \\ &= \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left((y_{ik}y_{jk} - \lambda_{ij}) - (\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij}) - (y_{ik} \bar{y}_j - \frac{1}{n} \lambda_{ij}) + (\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij}) \right)^2 \right] \\ &= \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left([y_{ik}y_{jk} - \lambda_{ij}]^2 + [\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij}]^2 + [\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij}]^2 + [\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij}]^2 \right. \right. \\ &\quad \left. \left. - 2 [y_{ik}y_{jk} - \lambda_{ij}] \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) - 2 [y_{ik}y_{jk} - \lambda_{ij}] \left(\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij} \right) \right. \right. \\ &\quad \left. \left. + 2 [y_{ik}y_{jk} - \lambda_{ij}] \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) + 2 \left[\left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij} \right) \right] \right. \right. \\ &\quad \left. \left. - 2 \left[\left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] - 2 \left[\left(\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \right) \right]. \end{aligned} \quad (64)$$

We then prove that the variance of each of the 10 separated sums converges to 0. For that, let's show a useful inequality.

6.9.1 Preliminary inequality

Let $(k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4) \in \llbracket 1, n \rrbracket^8$. Then, using multiple Cauchy-Schwarz and Jensen inequalities, we have:

$$\begin{aligned}
& \text{Cov} \left((y_{ik_1} y_{jk'_1} - \lambda_{ij} 1_{k_1=k'_1}) (y_{ik_2} y_{jk'_2} - \lambda_{ij} 1_{k_2=k'_2}), (y_{lk_3} y_{mk'_3} - \lambda_{lm} 1_{k_3=k'_3}) (y_{lk_4} y_{mk'_4} - \lambda_{lm} 1_{k_4=k'_4}) \right) \\
= & \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \lambda_{ij} 1_{k_1=k'_1} \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \lambda_{ij} 1_{k_2=k'_2} \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \lambda_{lm} 1_{k_3=k'_3} \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_4} y_{mk'_4} \right) \\
& - \lambda_{lm} 1_{k_4=k'_4} \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} \right) \\
& + \lambda_{ij} \lambda_{lm} 1_{k_1=k'_1} 1_{k_3=k'_3} \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_4} y_{mk'_4} \right) \\
& + \lambda_{ij} \lambda_{lm} 1_{k_1=k'_1} 1_{k_4=k'_4} \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} \right) \\
& + \lambda_{ij} \lambda_{lm} 1_{k_2=k'_2} 1_{k_3=k'_3} \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_4} y_{mk'_4} \right) \\
& + \lambda_{ij} \lambda_{lm} 1_{k_2=k'_2} 1_{k_4=k'_4} \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_3} y_{mk'_3} \right) \\
= & \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \mathbb{E}[y_{ik_1} y_{jk'_1}] \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \mathbb{E}[y_{ik_2} y_{jk'_2}] \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_3} y_{mk'_3} y_{lk_4} y_{mk'_4} \right) \\
& - \mathbb{E}[y_{lk_3} y_{mk'_3}] \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_4} y_{mk'_4} \right) \\
& - \mathbb{E}[y_{lk_4} y_{mk'_4}] \text{Cov} \left(y_{ik_1} y_{jk'_1} y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} \right) \\
& + \mathbb{E}[y_{ik_1} y_{jk'_1}] \mathbb{E}[y_{lk_3} y_{mk'_3}] \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_4} y_{mk'_4} \right) \\
& + \mathbb{E}[y_{ik_1} y_{jk'_1}] \mathbb{E}[y_{lk_4} y_{mk'_4}] \text{Cov} \left(y_{ik_2} y_{jk'_2}, y_{lk_3} y_{mk'_3} \right) \\
& + \mathbb{E}[y_{ik_2} y_{jk'_2}] \mathbb{E}[y_{lk_3} y_{mk'_3}] \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_4} y_{mk'_4} \right) \\
& + \mathbb{E}[y_{ik_2} y_{jk'_2}] \mathbb{E}[y_{lk_4} y_{mk'_4}] \text{Cov} \left(y_{ik_1} y_{jk'_1}, y_{lk_3} y_{mk'_3} \right) \\
\leq & 9 \sqrt[4]{\mathbb{E}[y_{i_1}^8] \mathbb{E}[y_{j_1}^8] \mathbb{E}[y_{l_1}^8] \mathbb{E}[y_{m_1}^8]}.
\end{aligned} \tag{65}$$

So, for all $(k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4) \in \llbracket 1, n \rrbracket^8$, we have,

$$\begin{aligned}
& \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1} y_{jk'_1} - \lambda_{ij} 1_{k_1=k'_1}) (y_{ik_2} y_{jk'_2} - \lambda_{ij} 1_{k_2=k'_2}), \right. \\
& \left. (y_{lk_3} y_{mk'_3} - \lambda_{lm} 1_{k_3=k'_3}) (y_{lk_4} y_{mk'_4} - \lambda_{lm} 1_{k_4=k'_4}) \right) \leq 9K_2.
\end{aligned} \tag{66}$$

6.9.2 Variance of the first term

Now, we can prove that the variance of each of the 10 separated sums converges to 0.

- We have:

$$\begin{aligned} & \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij})^2 \right] \\ &= \frac{p^2}{(n-1)^4} \sum_{k_1, k_2} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1}y_{jk_1} - \lambda_{ij})^2, (y_{lk_2}y_{mk_2} - \lambda_{lm})^2 \right). \end{aligned} \quad (67)$$

If $k_1 \neq k_2$, the covariance is trivially null. So,

$$\begin{aligned} & \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij})^2 \right] \\ &= \frac{p^2}{(n-1)^4} \sum_{k_1} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1}y_{jk_1} - \lambda_{ij})^2, (y_{lk_1}y_{mk_1} - \lambda_{lm})^2 \right) \\ &\leq \frac{K_1^2 n^3}{(n-1)^4} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{i1}y_{j1} - \lambda_{ij})^2, (y_{l1}y_{m1} - \lambda_{lm})^2 \right). \end{aligned} \quad (68)$$

And we have from the preliminary inequality,

$$\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{i1}y_{j1} - \lambda_{ij})^2, (y_{l1}y_{m1} - \lambda_{lm})^2 \right) \leq 9K_2. \quad (69)$$

Finally, we can conclude,

$$\begin{aligned} & \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij})^2 \right] \leq \frac{n^3}{(n-1)^4} \times 9K_1^2 K_2 \\ & \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij})^2 \right] \xrightarrow{n \rightarrow \infty} 0. \end{aligned} \quad (70)$$

- We have:

$$\begin{aligned} & \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right)^2 \right] \\ &= \frac{p^2}{(n-1)^4 n^4} \sum_{\substack{k_1, k'_1, k''_1 \\ k_2, k'_2, k''_2}} \sum_{i,j,l,m} \\ & \frac{1}{p^4} \text{Cov} \left((y_{ik'_1}y_{jk_1} - \lambda_{ij} \mathbf{1}_{k'_1=k_1}) (y_{ik''_1}y_{jk_1} - \lambda_{ij} \mathbf{1}_{k''_1=k_1}), (y_{lk'_2}y_{mk_2} - \lambda_{lm} \mathbf{1}_{k'_2=k_2}) (y_{lk''_2}y_{mk_2} - \lambda_{lm} \mathbf{1}_{k''_2=k_2}) \right). \end{aligned} \quad (71)$$

If $k_1, k'_1, k''_1, k_2, k'_2, k''_2$ are all different, then, the covariance is null. So there are at most $n^6 - n(n-1)(n-2)(n-3)(n-4)(n-5)$ non-zero terms in the sum over k -indices.

And we have from the preliminary inequality,

$$\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik'_1} y_{jk_1} - \lambda_{ij} \mathbf{1}_{k'_1=k_1}) (y_{ik'_1} y_{jk_1} - \lambda_{ij} \mathbf{1}_{k'_1=k_1}), \right. \\ \left. (y_{lk'_2} y_{mk_2} - \lambda_{lm} \mathbf{1}_{k'_2=k_2}) (y_{lk'_2} y_{mk_2} - \lambda_{lm} \mathbf{1}_{k'_2=k_2}) \right) \leq 9K_2. \quad (72)$$

So,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right)^2 \right] \leq \frac{p^2(n^6 - n(n-1)(n-2)(n-3)(n-4)(n-5))}{(n-1)^4 n^4} \times 9K_2 \\ \leq \frac{n^2(n^6 - n(n-1)(n-2)(n-3)(n-4)(n-5))}{(n-1)^4 n^4} \times 9K_1^2 K_2 \\ \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right)^2 \right] \xrightarrow{n \rightarrow \infty} 0. \quad (73)$$

- Immediately from the previous point,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij} \right)^2 \right] \xrightarrow{n \rightarrow \infty} 0. \quad (74)$$

- We have:

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right)^2 \right] \\ = \frac{p^2}{(n-1)^4 n^6} \sum_{\substack{k_1, k'_1, k_2, k'_2 \\ k_3, k'_3, k_4, k'_4}} \sum_{i,j,l,m} \\ \frac{1}{p^4} \text{Cov} \left((y_{ik_1} y_{jk'_1} - \lambda_{ij} \mathbf{1}_{k_1=k'_1}) (y_{ik_2} y_{jk'_2} - \lambda_{ij} \mathbf{1}_{k_2=k'_2}), (y_{lk_3} y_{mk'_3} - \lambda_{lm} \mathbf{1}_{k_3=k'_3}) (y_{lk_4} y_{mk'_4} - \lambda_{lm} \mathbf{1}_{k_4=k'_4}) \right). \quad (75)$$

If $k_1, k'_1, k_2, k'_2, k_3, k'_3, k_4, k'_4$ are all different, then, the covariance is null. So there are at most $n^8 - n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)(n-7)$ non-zero terms in the sum over k -indices.

And we have from the preliminary inequality,

$$\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1} y_{jk'_1} - \lambda_{ij} \mathbf{1}_{k_1=k'_1}) (y_{ik_2} y_{jk'_2} - \lambda_{ij} \mathbf{1}_{k_2=k'_2}), \right. \\ \left. (y_{lk_3} y_{mk'_3} - \lambda_{lm} \mathbf{1}_{k_3=k'_3}) (y_{lk_4} y_{mk'_4} - \lambda_{lm} \mathbf{1}_{k_4=k'_4}) \right) \leq 9K_2. \quad (76)$$

So,

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right)^2 \right] \\
& \leq \frac{p^2 (n^8 - n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)(n-7))}{(n-1)^4 n^6} \times 9K_2 \quad (77) \\
& \leq \frac{n^2 (n^8 - n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)(n-7))}{(n-1)^4 n^6} \times 9K_1^2 K_2 \\
& \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

• We have:

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik} y_{jk} - \lambda_{ij}) \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \right] \\
& = \frac{p^2}{(n-1)^4 n^2} \sum_{k_1, k'_1, k_2, k'_2} \sum_{i,j,l,m} \\
& \quad \frac{1}{p^4} \text{Cov} \left((y_{ik_1} y_{jk_1} - \lambda_{ij}) (y_{ik'_1} y_{jk_1} - \lambda_{ij} 1_{k_1=k'_1}), (y_{lk_2} y_{mk_2} - \lambda_{lm}) (y_{lk'_2} y_{mk_2} - \lambda_{lm} 1_{k_2=k'_2}) \right). \quad (78)
\end{aligned}$$

If $k'_1 \notin \{k_1, k_2, k'_2\}$, then, the covariance is null. So there are at most $n^4 - n(n-1)^3$ non-zero terms in the sum over k -indices.

And we have from the preliminary inequality,

$$\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1} y_{jk_1} - \lambda_{ij}) (y_{ik'_1} y_{jk_1} - \lambda_{ij} 1_{k_1=k'_1}), (y_{lk_2} y_{mk_2} - \lambda_{lm}) (y_{lk'_2} y_{mk_2} - \lambda_{lm} 1_{k_2=k'_2}) \right) \leq 9K_2. \quad (79)$$

So,

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik} y_{jk} - \lambda_{ij}) \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \right] \\
& \leq \frac{p^2 (n^4 - n(n-1)^3)}{(n-1)^4 n^2} \times 9K_2 \quad (80) \\
& \leq \frac{n^2 (n^4 - n(n-1)^3)}{(n-1)^4 n^2} \times 9K_1^2 K_2 \\
& \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

• Immediately from the previous point,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik} y_{jk} - \lambda_{ij}) \left(\bar{y}_j y_{i1} - \frac{1}{n} \lambda_{ij} \right) \right] \xrightarrow{n \rightarrow \infty} 0. \quad (81)$$

- We have:

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij}) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \\
&= \frac{p^2}{(n-1)^4 n^4} \sum_{\substack{k_1, k'_1, k''_1 \\ k_2, k'_2, k''_2}} \sum_{i,j,l,m} \\
& \frac{1}{p^4} \text{Cov} \left((y_{ik_1}y_{jk_1} - \lambda_{ij}) (y_{ik'_1}y_{jk'_1} - \lambda_{ij} \mathbf{1}_{k'_1=k'_1}), (y_{lk_2}y_{mk_2} - \lambda_{lm}) (y_{lk'_2}y_{mk'_2} - \lambda_{lm} \mathbf{1}_{k'_2=k'_2}) \right).
\end{aligned} \tag{82}$$

If $k'_1 \notin \{k_1, k'_1, k_2, k'_2, k''_2\}$, then, the covariance is null. So there are at most $n^6 - n(n-1)^5$ non-zero terms in the sum over k -indices.

And we have from the preliminary inequality,

$$\begin{aligned}
& \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1}y_{jk_1} - \lambda_{ij}) (y_{ik'_1}y_{jk'_1} - \lambda_{ij} \mathbf{1}_{k'_1=k'_1}), \right. \\
& \left. (y_{lk_2}y_{mk_2} - \lambda_{lm}) (y_{lk'_2}y_{mk'_2} - \lambda_{lm} \mathbf{1}_{k'_2=k'_2}) \right) \leq 9K_2.
\end{aligned} \tag{83}$$

So,

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (y_{ik}y_{jk} - \lambda_{ij}) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \\
& \leq \frac{p^2(n^6 - n(n-1)^5)}{(n-1)^4 n^4} \times 9K_2 \\
& \leq \frac{n^2(n^6 - n(n-1)^5)}{(n-1)^4 n^4} \times 9K_1^2 K_2 \\
& \xrightarrow{n \rightarrow \infty} 0.
\end{aligned} \tag{84}$$

- We have:

$$\begin{aligned}
& \mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \left(y_{ik} \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \\
&= \frac{p^2}{(n-1)^4 n^4} \sum_{\substack{k_1, k'_1, k''_1 \\ k_2, k'_2, k''_2}} \sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik'_1}y_{jk_1} - \lambda_{ij} \mathbf{1}_{k'_1=k_1}) (y_{ik_1}y_{jk'_1} - \lambda_{ij} \mathbf{1}_{k_1=k'_1}), \right. \\
& \left. (y_{lk'_2}y_{mk_2} - \lambda_{lm} \mathbf{1}_{k'_2=k_2}) (y_{lk_2}y_{mk'_2} - \lambda_{lm} \mathbf{1}_{k_2=k'_2}) \right).
\end{aligned} \tag{85}$$

If $k'_1 \notin \{k_1, k'_1, k_2, k'_2, k''_2\}$, then, the covariance is null. So there are at most $n^6 - n(n-1)^5$ non-zero terms in the sum over k -indices.

And we have from the preliminary inequality,

$$\sum_{i,j,l,m} \frac{1}{p^4} \text{Cov} \left((y_{ik_1'} y_{jk_1} - \lambda_{ij} \mathbf{1}_{k_1'=k_1}) (y_{ik_1} y_{jk_1'} - \lambda_{ij} \mathbf{1}_{k_1=k_1'}) \right. \\ \left. (y_{lk_2'} y_{mk_2} - \lambda_{lm} \mathbf{1}_{k_2'=k_2}) (y_{lk_2} y_{mk_2'} - \lambda_{lm} \mathbf{1}_{k_2=k_2'}) \right) \leq 9K_2. \quad (86)$$

So,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} (\bar{y}_i y_{jk} - \lambda_{ij}) \left(y_{ik} \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \\ \leq \frac{p^2(n^6 - n(n-1)^5)}{(n-1)^4 n^4} \times 9K_2 \quad (87) \\ \leq \frac{n^2(n^6 - n(n-1)^5)}{(n-1)^4 n^4} \times 9K_1^2 K_2 \\ \xrightarrow{n \rightarrow \infty} 0.$$

- Immediately from the fourth point,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_i y_{jk} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \xrightarrow{n \rightarrow \infty} 0. \quad (88)$$

- Immediately from the previous point,

$$\mathbb{V} \left[\frac{1}{p(n-1)^2} \sum_k \sum_{i,j} \left(\bar{y}_j y_{ik} - \frac{1}{n} \lambda_{ij} \right) \left(\bar{y}_i \bar{y}_j - \frac{1}{n} \lambda_{ij} \right) \right] \xrightarrow{n \rightarrow \infty} 0. \quad (89)$$

So, the first term has its variance converging to 0:

$$\mathbb{V} \left[\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - \Sigma \right\|^2 \right] \xrightarrow{n \rightarrow \infty} 0. \quad (90)$$

6.9.3 Variance of the second term

The variance of the second term converges to 0 as n goes to infinity. Indeed, from the proof of Lemma 3.4 in (Ledoit and Wolf, 2004) [10]:

$$\left(\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - S \right\|^2 \right) - \left(\frac{1}{n^2} \sum_k \left\| \frac{n}{n-1} \tilde{x}_{\cdot k} \tilde{x}_{\cdot k}^T - \Sigma \right\|^2 \right) = \frac{1}{n} \|S - \Sigma\|^2. \quad (91)$$

And we have:

$$\mathbb{E} [\|S - \Sigma\|^4] \leq \mathbb{E} [(\|S - \mu I\| + \|\mu I - \Sigma\|)^4]. \quad (92)$$

$\mathbb{E}[\|S - \mu I\|^4]$ and $\|\mu I - \Sigma\|$ are bounded, from Lemma 2 and Lemma 1 respectively, so $\mathbb{E}[\|S - \Sigma\|^4]$ is bounded. Consequently, $\mathbb{V} \left[\frac{1}{n} \|S - \Sigma\|^2 \right] \xrightarrow{n \rightarrow \infty} 0$. To conclude, $\mathbb{V} [b^2] \xrightarrow{n \rightarrow \infty} 0$.

6.10 Proof of Lemma 7

$$\begin{aligned}
\mathbb{V}[m] &= \mathbb{E} \left[\left(\frac{1}{p} \sum_i \frac{1}{n-1} \sum_k (y_{ik} - \bar{y}_i)^2 - \frac{1}{p} \sum_i \lambda_{ii} \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{1}{p} \sum_i \frac{1}{n-1} \sum_k \left((y_{ik} - \bar{y}_i)^2 - \frac{n-1}{n} \lambda_{ii} \right) \right)^2 \right] \\
&= \frac{1}{(n-1)^2} \mathbb{E} \left[\sum_{k_1} \sum_{k_2} \left(\frac{1}{p} \sum_i \left((y_{ik_1} - \bar{y}_i)^2 - \frac{n-1}{n} \lambda_{ii} \right) \right) \left(\frac{1}{p} \sum_j \left((y_{jk_2} - \bar{y}_j)^2 - \frac{n-1}{n} \lambda_{jj} \right) \right) \right] \\
&= \frac{1}{p^2(n-1)^2} \sum_{k_1, k_2} \sum_{i, j} \mathbb{E} \left[\left(y_{ik_1}^2 - 2y_{ik_1} \bar{y}_i + \bar{y}_i^2 - \frac{n-1}{n} \lambda_{ii} \right) \left(y_{jk_2}^2 - 2y_{jk_2} \bar{y}_j + \bar{y}_j^2 - \frac{n-1}{n} \lambda_{jj} \right) \right].
\end{aligned} \tag{93}$$

Fully developing the terms, we obtain,

$$\mathbb{V}[m] = \frac{1}{p^2(n-1)^2} \sum_{k_1} \sum_{k_2} \sum_{i, j} \mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) (y_{jk_2}^2 - \lambda_{jj}) \right] \tag{t1}$$

$$- 2\mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) \left(y_{jk_2} \bar{y}_j - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t2}$$

$$+ \mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) \left(\bar{y}_j^2 - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t3}$$

$$- 2\mathbb{E} \left[\left(y_{ik_1} \bar{y}_i - \frac{1}{n} \lambda_{ii} \right) (y_{jk_2}^2 - \lambda_{jj}) \right] \tag{t4}$$

$$+ 4\mathbb{E} \left[\left(y_{ik_1} \bar{y}_i - \frac{1}{n} \lambda_{ii} \right) \left(y_{jk_2} \bar{y}_j - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t5} \tag{94}$$

$$- 2\mathbb{E} \left[\left(y_{ik_1} \bar{y}_i - \frac{1}{n} \lambda_{ii} \right) \left(\bar{y}_j^2 - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t6}$$

$$+ \mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n} \lambda_{ii} \right) (y_{jk_2}^2 - \lambda_{jj}) \right] \tag{t7}$$

$$- 2\mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n} \lambda_{ii} \right) \left(y_{jk_2} \bar{y}_j - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t8}$$

$$+ \mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n} \lambda_{ii} \right) \left(\bar{y}_j^2 - \frac{1}{n} \lambda_{jj} \right) \right] \tag{t9}.$$

We compute the expectations separately for simplicity.

$$\begin{aligned}
(t1) : \mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) (y_{jk_2}^2 - \lambda_{jj}) \right] &= \mathbb{V}[y_{i1}y_{j1}] + \lambda_{ij}^2 - \lambda_{ii}\lambda_{jj} \text{ if } k_1 = k_2, \\
&= 0 \text{ otherwise.} \\
(t2) : \mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) \left(y_{jk_2}\bar{y}_j - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n} (\mathbb{V}[y_{i1}y_{j1}] + \lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}) \text{ if } k_1 = k_2, \\
&= 0 \text{ otherwise.} \\
(t3) : \mathbb{E} \left[(y_{ik_1}^2 - \lambda_{ii}) \left(\bar{y}_j^2 - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n^2} (\mathbb{V}[y_{i1}y_{j1}] + \lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}) \text{ if } k_1 = k_2, \\
&= 0 \text{ otherwise.} \\
(t4) : \mathbb{E} \left[\left(y_{ik_1}\bar{y}_i - \frac{1}{n}\lambda_{ii} \right) (y_{jk_2}^2 - \lambda_{jj}) \right] &= \frac{1}{n} (\mathbb{V}[y_{i1}y_{j1}] + \lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}) \text{ if } k_1 = k_2, \\
&= 0 \text{ otherwise.} \\
(t5) : \mathbb{E} \left[\left(y_{ik_1}\bar{y}_i - \frac{1}{n}\lambda_{ii} \right) \left(y_{jk_2}\bar{y}_j - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n^2} (\mathbb{V}[y_{i1}y_{j1}] + n\lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}) \text{ if } k_1 = k_2, \\
&= \frac{1}{n^2}\lambda_{ij}^2 \text{ otherwise.} \\
(t6) : \mathbb{E} \left[\left(y_{ik_1}\bar{y}_i - \frac{1}{n}\lambda_{ii} \right) \left(\bar{y}_j^2 - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n^3} (\mathbb{V}[y_{i1}y_{j1}] + (2n-1)\lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}). \\
(t7) : \mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n}\lambda_{ii} \right) (y_{jk_2}^2 - \lambda_{jj}) \right] &= \frac{1}{n^2} (\mathbb{V}[y_{i1}y_{j1}] + \lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}). \\
(t8) : \mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n}\lambda_{ii} \right) \left(y_{jk_2}\bar{y}_j - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n^3} (\mathbb{V}[y_{i1}y_{j1}] + (2n-1)\lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}). \\
(t9) : \mathbb{E} \left[\left(\bar{y}_i^2 - \frac{1}{n}\lambda_{ii} \right) \left(\bar{y}_j^2 - \frac{1}{n}\lambda_{jj} \right) \right] &= \frac{1}{n^3} (\mathbb{V}[y_{i1}y_{j1}] + (2n-1)\lambda_{ij}^2 - \lambda_{ii}\lambda_{jj}).
\end{aligned} \tag{95}$$

Injecting those expectations in the main equation, we obtain:

$$\begin{aligned}
\mathbb{V}[m] &= \frac{1}{p^2(n-1)^2} \sum_{i,j} \left[\frac{n^2-2n+1}{n} \mathbb{V}[y_{i1}y_{j1}] + \frac{n^2-1}{n} \lambda_{ij}^2 - \frac{n^2-2n+1}{n} \lambda_{ii}\lambda_{jj} \right] \\
&= \frac{1}{p^2n(n-1)} \sum_{i,j} [(n-1)\mathbb{V}[y_{i1}y_{j1}] + (n+1)\lambda_{ij}^2 - (n-1)\lambda_{ii}\lambda_{jj}] \\
&= \frac{1}{p^2n} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{n+1}{pn(n-1)} \|\Sigma\|^2 - \frac{1}{n} \langle \Sigma, I \rangle^2 \\
&= \frac{1}{p^2n} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{n+1}{pn(n-1)} (\alpha^2 + \mu^2) - \frac{1}{n} \mu^2 \\
\mathbb{V}[m] &= \frac{1}{p^2n} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{n+1}{pn(n-1)} (\delta^2 - \beta^2 + \mu^2) - \frac{1}{n} \mu^2.
\end{aligned} \tag{96}$$

From the proof Lemma 1, we have:

$$\beta^2 = \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{p+1}{n(n-1)} \mu^2 + \frac{1}{n(n-1)} (\delta^2 - \beta^2). \tag{97}$$

So, combining those two equations, we obtain:

$$\begin{aligned}\mathbb{V}[m] &= \frac{1}{p} \left(\beta^2 - \frac{p+1}{n(n-1)} \mu^2 - \frac{1}{n(n-1)} (\delta^2 - \beta^2) \right) + \frac{n+1}{pn(n-1)} (\delta^2 - \beta^2 + \mu^2) - \frac{1}{n} \mu^2 \\ &= \frac{n-2}{p(n-1)} \beta^2 + \frac{1}{p(n-1)} \delta^2 - \frac{p-1}{p(n-1)} \mu^2 \\ \mathbb{V}[m] &= q_0 \beta^2 + q_1 \delta^2 - q_2 \mu^2.\end{aligned}\tag{98}$$

which concludes the proof.

6.11 Proof of Lemma 8

From the proof of Lemma 5, we have:

$$\mathbb{E}[m^2] = \mu^2 + \mathbb{V}[m], \quad \mathbb{E}[d^2] = \delta^2 - \mathbb{V}[m].\tag{99}$$

And from Lemma 7, we have:

$$\mathbb{V}[m] = q_0 \beta^2 + q_1 \delta^2 - q_2 \mu^2.\tag{100}$$

Which immediately finishes the proof:

$$\mathbb{V}[m] = \frac{1}{1 - q_1 - q_2} (q_0 \beta^2 + q_1 \mathbb{E}[d^2] - q_2 \mathbb{E}[m^2]).\tag{101}$$

6.12 Proof of Lemma 9

From Lemma 5:

$$\mathbb{E}[\bar{b}^2] = c_0 \beta^2 + c_1 \mathbb{E}[d^2] + c_2 \mathbb{E}[m^2] + (c_1 - c_2) \mathbb{V}[m].\tag{102}$$

And from Lemma 8:

$$\mathbb{V}[m] = \frac{1}{1 - q_1 - q_2} (q_0 \beta^2 + q_1 \mathbb{E}[d^2] - q_2 \mathbb{E}[m^2]).\tag{103}$$

So,

$$\begin{aligned}\mathbb{E}[\bar{b}^2] &= c_0 \beta^2 + c_1 \mathbb{E}[d^2] + c_2 \mathbb{E}[m^2] + \frac{c_1 - c_2}{1 - q_1 - q_2} (q_0 \beta^2 + q_1 \mathbb{E}[d^2] - q_2 \mathbb{E}[m^2]) \\ &= c_0^f \beta^2 + c_1^f \mathbb{E}[d^2] + c_2^f \mathbb{E}[m^2].\end{aligned}\tag{104}$$

Then, we deduce:

$$\begin{aligned}\mathbb{E}[b^2] &= \frac{1}{c_0^f} \left((c_0^f \beta^2 + c_1^f \mathbb{E}[d^2] + c_2^f \mathbb{E}[m^2]) - c_1^f \mathbb{E}[d^2] - c_2^f \mathbb{E}[m^2] \right) \\ \mathbb{E}[b^2] &= \beta^2.\end{aligned}\tag{105}$$

So, b^2 is an unbiased estimator of β^2 .

Concerning the quadratic mean convergence, we use the fact that the variances of m^2 , d^2 and \bar{b}^2 converge to 0 as n goes to infinity, from Corollary 2, and Lemma 3 and 6 respectively. So,

$$b^2 - \beta^2 = b^2 - \mathbb{E}[b^2] \xrightarrow[q.m.]{} 0.\tag{106}$$

6.13 Proof of Lemma 10

For the upper bound:

$$\begin{aligned}
b_u^2 - \beta^2 &= \min(b_+^2, d^2) - \beta^2 \\
&\leq b_+^2 - \beta^2 \leq |b^2 - \beta^2| \\
&\leq \max(|b^2 - \beta^2|, |d^2 - \delta^2|).
\end{aligned} \tag{107}$$

For the lower bound:

$$\begin{aligned}
b_u^2 - \beta^2 &= \min(b_+^2 - \beta^2, d^2 - \beta^2) \\
&\geq \min(b_+^2 - \beta^2, d^2 - \delta^2) \\
&\geq \min(-|b^2 - \beta^2|, -|d^2 - \delta^2|) \\
&\geq -\max(|b^2 - \beta^2|, |d^2 - \delta^2|).
\end{aligned} \tag{108}$$

So,

$$\begin{aligned}
\mathbb{E}[(b_u^2 - \beta^2)^2] &\leq \mathbb{E}[\max(|b^2 - \beta^2|, |d^2 - \delta^2|)^2] \\
&\leq \mathbb{E}[(b^2 - \beta^2)^2] + \mathbb{E}[(d^2 - \delta^2)^2].
\end{aligned} \tag{109}$$

So, from Lemma 3 and 9,

$$\mathbb{E}[(b_u^2 - \beta^2)^2] \longrightarrow 0. \tag{110}$$

Which leads trivially to:

$$\mathbb{E}[(a_u^2 - \alpha^2)^2] \longrightarrow 0. \tag{111}$$

6.14 Proof of Theorem 2

We will use the proof of Theorem 3.2 in (Ledoit and Wolf, 2004) [10] to prove ours. We check that we have the set of hypotheses required by the proof to work:

- $\alpha^2, \beta^2, \delta^2$ are non-negative, bounded, and $\alpha^2 + \beta^2 = \delta^2$,
- $m - \mu$ converges to 0 in quartic mean,
- d_u^2 is nonnegative, and $d_u^2 - \delta^2 \xrightarrow{q.m} 0$,
- $0 \leq a_u^2 \leq d_u^2$ and $a_u^2 - \alpha^2 \xrightarrow{q.m} 0$,
- $a_u^2 + b_u^2 = d_u^2$, with $b_u^2 \geq 0$.

Then, we can apply the result of the theorem 3.2 from (Ledoit and Wolf, 2004) [10], so $\mathbb{E}[\|S_n^* - \Sigma_n^*\|^2] \rightarrow 0$ and S_n^* has the same asymptotic expected loss as Σ_n^* , i.e. $\mathbb{E}[\|S_n^* - \Sigma_n\|_n^2] - \mathbb{E}[\|\Sigma_n^* - \Sigma_n\|_n^2] \rightarrow 0$.

6.15 Proof of Lemma 11

The proof from (Ledoit and Wolf, 2004) [10] can be applied here, as the hypotheses of their Lemma A.1 [10] are verified for $u^2 = |a_u^2 b_u^2 \delta^2 - \alpha^2 \beta^2 d_u^2|$, $\tau_1 = 2$ and $\tau_2 = 2$, from the same arguments they used.

6.16 Proof of Theorem 3

Solving the convex minimization problem, we obtain easily that $\Sigma^{**} = \mu I + \frac{\alpha_2}{d^2}(S - mI)$, with $\alpha_2 = \langle S, \Sigma \rangle - m\mu$, is a minimizer. We have then:

$$\begin{aligned} \|S^* - \Sigma^{**}\|^2 &= \left\| (m - \mu)I + \frac{a_u^2 - \alpha_2}{d^2}(S - mI) \right\|^2 \\ \|S^* - \Sigma^{**}\|^2 &= (m - \mu)^2 + \frac{(a_u^2 - \alpha_2)^2}{d^2}. \end{aligned} \tag{112}$$

$(m - \mu)^2$ converges to 0 in quadratic mean by Lemma 2. For the second term, we will use the Lemma A.1 from Ledoit and Wolf [10] with $u^2 = (a_u^2 - \alpha_2)^2$, $\tau_1 = 2$ and $\tau_2 = 0$. In the following, we check the assumptions of the Lemma A.1, i.e. $\frac{(a_u^2 - \alpha_2)^2}{d^2} \leq 2d^2 + 2\delta^2$ and $\mathbb{E}[(a_u^2 - \alpha_2)^2] \rightarrow 0$. We notice:

$$|\alpha_2| = |\langle \Sigma - \mu I, S - mI \rangle| \leq \|\Sigma - \mu I\| \|S - mI\| = \delta d. \tag{113}$$

It comes that:

$$\begin{aligned} \frac{(a_u^2 - \alpha_2)^2}{d^2} &= \frac{a_u^4 + \alpha_2^2 - 2a_u^2\alpha_2}{d^2} \\ &\leq \frac{2a_u^4 + 2\alpha_2^2}{d^2} \\ &\leq 2d^2 + 2\delta^2. \end{aligned} \tag{114}$$

In order to prove $\mathbb{E}[(a_u^2 - \alpha_2)^2] \rightarrow 0$, let's show that $\mathbb{E}[(\alpha_2 - \alpha^2)^2] = \mathbb{V}[\alpha_2] \rightarrow 0$.

$$\begin{aligned} \mathbb{V}[\alpha_2] &= \mathbb{V}[\langle S, \Sigma \rangle - m\mu] \\ &= \mathbb{V}[\langle S, \Sigma \rangle] + \mathbb{V}[m\mu] - 2\text{Cov}(\langle S, \Sigma \rangle, m\mu) \\ &\leq 2\mathbb{V}[\langle S, \Sigma \rangle] + 2\mu^2\mathbb{V}[m]. \end{aligned} \tag{115}$$

μ is bounded by Lemma 1 and $\mathbb{V}[m] \rightarrow 0$ by Lemma 2, so $\mu^2 \mathbb{V}[m] \rightarrow 0$. Considering the other term, we have:

$$\begin{aligned}
\mathbb{V}[\langle S, \Sigma \rangle] &= \mathbb{V} \left[\left\langle \frac{1}{n-1} \tilde{Y} \tilde{Y}^T, \Lambda \right\rangle \right] \\
&= \mathbb{V} \left[\frac{1}{p(n-1)} \sum_i \lambda_{ii} \sum_k (y_{ik} - \bar{y}_i)^2 \right] \\
&= \mathbb{V} \left[\frac{1}{pn(n-1)} \sum_i \lambda_{ii} \sum_{k, k' \neq k} y_{ik} (y_{ik} - y_{ik'}) \right] \quad (\text{Identity 1}) \\
&= \frac{1}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2, k'_2 \neq k_2}} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1} (y_{ik_1} - y_{ik'_1}), \frac{1}{p} \sum_i \lambda_{ii} y_{ik_2} (y_{ik_2} - y_{ik'_2}) \right) \\
&= \frac{1}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2, k'_2 \neq k_2}} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1}^2, \frac{1}{p} \sum_i \lambda_{ii} y_{ik_2}^2 \right) \\
&\quad + \frac{1}{n^2(n-1)^2} \sum_{\substack{k_1, k'_1 \neq k_1 \\ k_2, k'_2 \neq k_2}} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1} y_{ik'_1}, \frac{1}{p} \sum_i \lambda_{ii} y_{ik_2} y_{ik'_2} \right) \\
&= \frac{1}{n^2} \sum_{k_1} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1}^2, \frac{1}{p} \sum_i \lambda_{ii} y_{ik_1}^2 \right) \\
&\quad + \frac{1}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1} y_{ik'_1}, \frac{1}{p} \sum_i \lambda_{ii} y_{ik_1} y_{ik'_1} \right) \\
&\quad + \frac{1}{n^2(n-1)^2} \sum_{k_1, k'_1 \neq k_1} \text{Cov} \left(\frac{1}{p} \sum_i \lambda_{ii} y_{ik_1} y_{ik'_1}, \frac{1}{p} \sum_i \lambda_{ii} y_{ik'_1} y_{ik_1} \right) \\
&= \frac{1}{n} \mathbb{V} \left[\frac{1}{p} \sum_i \lambda_{ii} y_{i1}^2 \right] + \frac{2}{n(n-1)} \mathbb{V} \left[\frac{1}{p} \sum_i \lambda_{ii} y_{i1} y_{i2} \right] \\
&\leq \frac{1}{n} \mathbb{E} \left[\left(\frac{1}{p} \sum_i \lambda_{ii} y_{i1}^2 \right)^2 \right] + \frac{2}{n(n-1)} \mathbb{E} \left[\left(\frac{1}{p} \sum_i \lambda_{ii} y_{i1} y_{i2} \right)^2 \right] \\
&\leq \frac{1}{n} \mathbb{E} \left[\left(\frac{1}{p} \sum_i \lambda_{ii}^2 \right) \left(\frac{1}{p} \sum_i y_{i1}^4 \right) \right] + \frac{2}{n(n-1)} \mathbb{E} \left[\left(\frac{1}{p} \sum_i \lambda_{ii}^2 \right) \left(\frac{1}{p} \sum_i y_{i1}^2 y_{i2}^2 \right) \right] \\
&= \frac{1}{n} \left(\frac{1}{p} \sum_i \mathbb{E} [y_{i1}^2]^2 \right) \left(\frac{1}{p} \sum_i \mathbb{E} [y_{i1}^4] \right) + \frac{2}{n(n-1)} \left(\frac{1}{p} \sum_i \mathbb{E} [y_{i1}^2]^2 \right)^2 \\
&\leq \frac{n+1}{n(n-1)} \left(\frac{1}{p} \sum_i \mathbb{E} [y_{i1}^4] \right)^2 \\
&\leq \frac{n+1}{n(n-1)} \frac{1}{p} \sum_i \mathbb{E} [y_{i1}^8] \\
\mathbb{V}[\langle S, \Sigma \rangle] &\leq \frac{n+1}{n(n-1)} K_2.
\end{aligned}$$

(116)

So, $\mathbb{V}[\langle S, \Sigma \rangle] \rightarrow 0$, and so $\mathbb{E}[(\alpha_2 - \alpha^2)^2] \rightarrow 0$. As $\mathbb{E}[(a_u^2 - \alpha^2)^2] \rightarrow 0$ by Lemma 9, it comes that $\mathbb{E}[(a_u^2 - \alpha_2)^2] \rightarrow 0$. Therefore, the assumptions of Lemma A.1 of (Ledoit and Wolf, 2004) [10] are verified by $u^2 = (a_u^2 - \alpha_2)^2$, $\tau_1 = 2$ and $\tau_2 = 0$. It proves that:

$$\mathbb{E} \left[\frac{(a_u^2 - \alpha_2)^2}{d^2} \right] \rightarrow 0. \quad (117)$$

Backing up, we have shown that $\mathbb{E}[\|S^* - \Sigma^{**}\|^2] \rightarrow 0$. We complete the proof of the theorem with the following inequality:

$$\begin{aligned} \mathbb{E} [\|S^* - \Sigma\|^2 - \|\Sigma^{**} - \Sigma\|^2] &= \mathbb{E} [\langle S^* - \Sigma^{**}, S^* + \Sigma^{**} - 2\Sigma \rangle] \\ &\leq \sqrt{\mathbb{E} [\|S^* - \Sigma^{**}\|^2]} \sqrt{\mathbb{E} [\|S^* + \Sigma^{**} - 2\Sigma\|^2]}. \end{aligned} \quad (118)$$

The first term converges to 0 as we showed above, and the second term is bounded because $\mathbb{E}[\|S^* - \Sigma\|^2]$ is bounded. So, the product converges to 0, which completes the proof.

6.17 Proof of Theorem 4

The proof from (Ledoit and Wolf, 2004) [10] can be applied as it is, because it uses only the results of Theorem 3 which are the same as Theorem 3.3 in [10].

6.18 Proof of Theorem 5

Respectively from Lemma 2 and 3, $m_r - \mu$ converges to 0 in quartic mean and $d_r^2 - \delta^2$ converges to 0 in quadratic mean.

Let's define:

$$\bar{b}_r^2 = \frac{1}{(n-1)^2} \sum_{k=1}^n \|\tilde{x}_{\cdot k}(\tilde{x}_{\cdot k})^t - S\|^2. \quad (119)$$

Let's show that $\bar{b}_r^2 - \beta^2$ converges to 0 in quadratic mean. We use the following decomposition:

$$\begin{aligned} \bar{b}_r^2 &= \frac{1}{(n-1)^2} \sum_k \left\| \tilde{x}_k \tilde{x}_k^t - \frac{n-1}{n} S - \frac{1}{n} S \right\|^2 \\ &= \bar{b}^2 - \frac{2}{(n-1)^2} \sum_k \left\langle \tilde{x}_k \tilde{x}_k^t - \frac{n-1}{n} S, \frac{1}{n} S \right\rangle + \frac{1}{n(n-1)^2} \|S\|^2 \\ \bar{b}_r^2 &= \bar{b}^2 + \frac{1}{n(n-1)^2} \|S\|^2. \end{aligned} \quad (120)$$

$\bar{b}^2 - \beta^2$ converges to 0 in quadratic mean by Lemma 6. Moreover, $\|S\|^2 - \mathbb{E}[\|S\|^2]$ converges to 0 in quadratic mean by Lemma 3, and $\mathbb{E}[\|S\|^2] = \beta^2 + \alpha^2 + \mu^2$, so $\mathbb{E}[\|S\|^2]$ is bounded by Lemma 1. So, $\frac{1}{n(n-1)^2} \|S\|^2$ converges to 0 in quadratic mean, which finally proves that $\bar{b}_r^2 - \beta^2$ converges to 0 in quadratic mean.

Following the idea of proof of Lemma 10, we have:

$$- \max(|\bar{b}_r^2 - \beta^2|, |d_r^2 - \delta^2|) \leq \bar{b}_r^2 - \beta^2 \leq \max(|\bar{b}_r^2 - \beta^2|, |d_r^2 - \delta^2|). \quad (121)$$

So,

$$\begin{aligned}\mathbb{E}[(b_r^2 - \beta^2)^2] &\leq \mathbb{E}[\max(|\bar{b}_r^2 - \beta^2|, |d_r^2 - \delta^2|)^2] \\ &\leq \mathbb{E}[(\bar{b}_r^2 - \beta^2)^2] + \mathbb{E}[(d_r^2 - \delta^2)^2].\end{aligned}\tag{122}$$

Which, as previously, leads to the 2 following results: $\mathbb{E}[(b_r^2 - \beta^2)^2] \rightarrow 0$, $\mathbb{E}[(a_r^2 - \alpha^2)^2] \rightarrow 0$. Finally, we check that we have the set of hypotheses required by the proof of Theorem 2 to work:

- $\alpha^2, \beta^2, \delta^2$ are non-negative, bounded, and $\alpha^2 + \beta^2 = \delta^2$,
- $m_r - \mu$ converges to 0 in quartic mean,
- d_r^2 is nonnegative, and $d_r^2 - \delta^2 \xrightarrow{q.m} 0$,
- $0 \leq a_r^2 \leq d_s^2$ and $a_r^2 - \alpha^2 \xrightarrow{q.m} 0$,
- $a_r^2 + b_r^2 = d_r^2$, with $b_r^2 \geq 0$.

Then, we can apply the result of the theorem 3.2 from (Ledoit and Wolf, 2004) [10], so identically we have that $\mathbb{E}[\|S_r^* - \Sigma^*\|^2] \rightarrow 0$ and S_r^* has the same asymptotic expected loss as $\mathbb{E}[\|S_r^* - \Sigma\|_n^2 - \|\Sigma^{**} - \Sigma\|^2] \rightarrow 0$.

6.19 Proof of Theorem 6

Respectively from Lemma 2 and 3, $m_m - \mu$ converges to 0 in quartic mean and $d_m^2 - \delta^2$ converges to 0 in quadratic mean.

$\bar{b}^2 - \beta^2$ converges to 0 in quadratic mean by Lemma 6, and following the idea of proof of Lemma 10, we have:

$$-\max(|\bar{b}^2 - \beta^2|, |d_m^2 - \delta^2|) \leq b_m^2 - \beta^2 \leq \max(|\bar{b}^2 - \beta^2|, |d_m^2 - \delta^2|).\tag{123}$$

So,

$$\begin{aligned}\mathbb{E}[(b_m^2 - \beta^2)^2] &\leq \mathbb{E}[\max(|\bar{b}^2 - \beta^2|, |d_m^2 - \delta^2|)^2] \\ &\leq \mathbb{E}[(\bar{b}^2 - \beta^2)^2] + \mathbb{E}[(d_m^2 - \delta^2)^2].\end{aligned}\tag{124}$$

Which, as previously, leads to the 2 following results: $\mathbb{E}[(b_m^2 - \beta^2)^2] \rightarrow 0$, $\mathbb{E}[(a_m^2 - \alpha^2)^2] \rightarrow 0$. Finally, we check that we have the set of hypotheses required by the proof of Theorem 2 to work:

- $\alpha^2, \beta^2, \delta^2$ are non-negative, bounded, and $\alpha^2 + \beta^2 = \delta^2$,
- $m_m - \mu$ converges to 0 in quartic mean,
- d_m^2 is nonnegative, and $d_m^2 - \delta^2 \xrightarrow{q.m} 0$,
- $0 \leq a_m^2 \leq d_s^2$ and $a_m^2 - \alpha^2 \xrightarrow{q.m} 0$,
- $a_m^2 + b_m^2 = d_m^2$, with $b_m^2 \geq 0$.

Then, we can apply the result of the theorem 3.2 from (Ledoit and Wolf, 2004) [10], so identically we have that $\mathbb{E}[\|S_m^* - \Sigma^*\|^2] \rightarrow 0$ and S_m^* has the same asymptotic expected loss as Σ^* , i.e. $\mathbb{E}[\|S_m^* - \Sigma\|^2 - \|\Sigma^* - \Sigma\|^2] \rightarrow 0$.

6.20 Proof of Theorem 7

As $d_s^2 = d^2$ and $b_s^2 = b_m^2$, we trivially have the quadratic convergence to δ^2 and β^2 respectively with Lemma 3 and Theorem 6. $m_s = \frac{n-1}{n}m$, and $m - \mu$ converges in quartic mean to 0 with μ bounded, so $m_s - \mu$ converges in quartic mean to 0. Similarly, $a_s^2 = \frac{n-1}{n}a_m^2$ and $a_m^2 - \alpha^2$ converges in quadratic mean with α^2 bounded, so $a_s^2 - \alpha^2$ converges in quadratic mean. Finally, as $S_s^* = \frac{n-1}{n}S_m^*$, we have:

$$\begin{aligned}\mathbb{E}[\|S_s^* - \Sigma^*\|^2] &= \frac{(n-1)^2}{n^2} \mathbb{E} \left[\left\| S_m^* - \Sigma^* - \frac{1}{n-1} \Sigma^* \right\|^2 \right] \\ &= \frac{(n-1)^2}{n^2} \mathbb{E} \left[\|S_m^* - \Sigma^*\|^2 - \frac{2}{n-1} \langle S_m^* - \Sigma^*, \Sigma^* \rangle + \frac{1}{n^2} \|\Sigma^*\|^2 \right] \\ \mathbb{E}[\|S_s^* - \Sigma^*\|^2] &\leq \frac{(n-1)^2}{n^2} \mathbb{E} \left[\|S_m^* - \Sigma^*\|^2 + \frac{1}{n^2} \|\Sigma^*\|^2 \right] + \frac{2(n-1)}{n^2} \sqrt{\mathbb{E}[\|S_m^* - \Sigma^*\|^2] \mathbb{E}[\|\Sigma^*\|^2]}.\end{aligned}\tag{125}$$

Using Theorem 6 and that $\mathbb{E}[\|\Sigma^*\|^2]$ bounded, we have that $\mathbb{E}[\|S_s^* - \Sigma^*\|^2] \rightarrow 0$ and S_s^* has the same asymptotic expected loss as Σ^* , i.e. $\mathbb{E}[\|\|S_s^* - \Sigma\|^2 - \|\Sigma^* - \Sigma\|^2\|] \rightarrow 0$.

6.21 Proof of Lemma 12

Let $X_{\cdot,k} \sim \mathcal{N}(0, \Sigma)$, $k \in \llbracket 1, n \rrbracket$, n iid samples. Trivially, $\mu = \frac{1}{p} \text{tr}(\Sigma)$, $\alpha^2 = \|\Sigma - \mu I\|^2$. For β^2 , we will use the equation from the proof of Lemma 1.

$$\beta^2 = \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{p+1}{n(n-1)} \mu^2 + \frac{1}{n(n-1)} \alpha^2.\tag{126}$$

As X is Gaussian, we have for all $(i, j) \in \llbracket 1, p \rrbracket^2$, $\mathbb{V}[y_{i1}y_{j1}] = \lambda_{ii}\lambda_{jj} + \lambda_{ij}^2$. So,

$$\frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] = \frac{1}{n} \left(\|\Sigma\|^2 + \frac{1}{p} \text{tr}(\Sigma)^2 \right) = \frac{1}{n} (\alpha^2 + (p+1)\mu^2).\tag{127}$$

Which finally leads to,

$$\begin{aligned}\beta^2 &= \frac{p+1}{n(n-1)} \mu^2 + \frac{1}{n(n-1)} \alpha^2 + \frac{1}{n} (\alpha^2 + (p+1)\mu^2) \\ \beta^2 &= \frac{p+1}{n-1} \mu^2 + \frac{1}{n-1} \alpha^2.\end{aligned}\tag{128}$$

And, of course, $\delta^2 = \alpha^2 + \beta^2$.

6.22 Proof of Lemma 13

Let $\nu > 8$, Σ a covariance matrix, $X_{\cdot,k} \sim t_\nu(0, \tilde{\Sigma})$, $k \in \llbracket 1, n \rrbracket$, n iid samples, with scale matrix $\tilde{\Sigma} = \frac{\nu-2}{\nu} \Sigma$.

With this setup, we have as expected: $\mathbb{V}[X] = \Sigma$. Obviously, $\mu = \frac{1}{p} \text{tr}(\Sigma)$, $\alpha^2 = \|\Sigma - \mu I\|^2$.

For β^2 , we will use the equation from the proof of Lemma 1, as in the Gaussian case.

$$\beta^2 = \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] + \frac{p+1}{n(n-1)} \mu^2 + \frac{1}{n(n-1)} \alpha^2.\tag{129}$$

To easily compute the variance term, we will use a characterization of multivariate t -distributions. In fact, as for all $k \in \llbracket 1, n \rrbracket$, $X_{\cdot,k} \sim t_\nu(0, \tilde{\Sigma})$, there exists two independant random variables U_k and $Z_{\cdot,k}$ such that:

$$U_k \sim \chi_\nu^2, Z_{\cdot,k} \sim \mathcal{N}\left(0, \frac{\nu-2}{\nu}\Lambda\right), Y_{\cdot,k} = \Sigma^{-1/2}X_{\cdot,k} = \sqrt{\frac{\nu}{U_k}}Z_{\cdot,k}. \quad (130)$$

We deduce then,

$$\begin{aligned} \mathbb{V}[y_{i1}y_{j1}] &= \nu^2 \mathbb{V}\left[\frac{Z_i Z_j}{U}\right] \\ &= \nu^2 \mathbb{E}[Z_i^2 Z_j^2] \mathbb{E}\left[\frac{1}{U^2}\right] - \mathbb{E}[Y_i Y_j]^2 \end{aligned} \quad (131)$$

$$\mathbb{V}[y_{i1}y_{j1}] = (\nu-2)^2 (\lambda_{ii}\lambda_{jj} + 2\lambda_{ij}^2) \mathbb{E}\left[\frac{1}{U^2}\right] - \lambda_{ij}^2.$$

We have:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{U^2}\right] &= \frac{(1/2)^{\nu/2-1}}{\Gamma(\nu/2)} \int_{\mathbb{R}} x^{\frac{\nu-4}{2}-1} e^{-x/2} dx \\ &= \frac{(1/2)^{\nu/2} \Gamma((\nu-4)/2)}{\Gamma(\nu/2) (1/2)^{(\nu-4)/2}} \times \underbrace{\int_{\mathbb{R}} \frac{(1/2)^{(\nu-4)/2-1}}{\Gamma((\nu-4)/2)} x^{\frac{\nu-4}{2}-1} e^{-x/2} dx}_{\text{pdf of } \chi_{\nu-4}^2} \\ &= \frac{(1/2)^{\nu/2} \Gamma((\nu-4)/2)}{\Gamma(\nu/2) (1/2)^{(\nu-4)/2}} \\ &= \frac{1}{4} \frac{1}{(\nu/2-1)(\nu/2-2)} \\ \mathbb{E}\left[\frac{1}{U^2}\right] &= \frac{1}{(\nu-2)(\nu-4)}. \end{aligned} \quad (132)$$

So,

$$\mathbb{V}[y_{i1}y_{j1}] = \frac{\nu-2}{\nu-4} (\lambda_{ii}\lambda_{jj} + 2\lambda_{ij}^2) - \lambda_{ij}^2. \quad (133)$$

And,

$$\begin{aligned} \frac{1}{pn} \sum_{i,j} \mathbb{V}[y_{i1}y_{j1}] &= \frac{1}{n} \left(\left(2\frac{\nu-2}{\nu-4} - 1\right) \|\Sigma\|^2 + \frac{\nu-2}{p(\nu-4)} \text{tr}(\Sigma)^2 \right) \\ &= \frac{1}{n} \left(\frac{\nu}{\nu-4} (\alpha^2 + \mu^2) + \frac{\nu-2}{\nu-4} p\mu^2 \right). \end{aligned} \quad (134)$$

We can conclude,

$$\beta^2 = \left(\frac{\nu-2}{(\nu-4)n} + \frac{1}{n(n-1)} \right) p\mu^2 + \frac{1}{n} \left(\frac{\nu}{\nu-4} + \frac{1}{n-1} \right) (\alpha^2 + \mu^2). \quad (135)$$

And, of course, $\delta^2 = \alpha^2 + \beta^2$.

References

- [1] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.
- [2] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- [3] Lars Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, 1982.
- [4] Victor Demiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22, 2009.
- [5] V. L. Girko. G-analysis of high dimensional observations. *Journal of Soviet mathematics*, 1985.
- [6] V. L. Girko. *Theory of Random Determinants*. Springer Netherlands, Dordrecht, 1990.
- [7] Vyacheslav L. Girko. *G²-estimator for the Stieltjes Transform of the Normalized Spectral Function of Covariance Matrices*, pages 171–212. Springer Netherlands, Dordrecht, 1995.
- [8] Charles M. Stein. 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*, 1956.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python, version 1.2.2 of april 2023. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [10] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.
- [11] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, pages 5016–5029, 2010.
- [12] Yuki Ikeda and Tatsuya Kubokawa. Linear shrinkage estimation of large covariance matrices using factor models. *Journal of Multivariate Analysis*, 152:61–81, 2016.
- [13] Karina Ashurbekova, Antoine Usseglio-Carleve, Florence Forbes, and Sophie Achard. Optimal shrinkage for robust covariance matrix estimators in a small sample size setting. working paper or preprint, 2021.
- [14] Clifford S. Stein. Estimation of a covariance matrix, 1975.
- [15] Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2), 2012.
- [16] Olivier Ledoit and Michael Wolf. Optimal estimation of a large-dimensional covariance matrix under Stein’s loss. *Bernoulli*, 24(4B):3791 – 3832, 2018.
- [17] Olivier Ledoit and Michael Wolf. Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5):3043 – 3065, 2020.

- [18] Romain Couillet and Matthew McKay. Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131:99–120, 2014.
- [19] Bala Rajaratnam and Dario Vincenzi. A theoretical study of Stein’s covariance estimator. working paper or preprint, 2015.
- [20] Yuki Ikeda, Tatsuya Kubokawa, and Muni S. Srivastava. Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Computational Statistics & Data Analysis*, 95:95–108, 2016.
- [21] Olivier Ledoit and Michael Wolf. The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218, 2020.