



Analysis of a multi-target linear shrinkage covariance estimator

Benoit Oriol

► To cite this version:

| Benoit Oriol. Analysis of a multi-target linear shrinkage covariance estimator. 2024. hal-04596412

HAL Id: hal-04596412

<https://hal.science/hal-04596412v1>

Preprint submitted on 31 May 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Analysis of a multi-target linear shrinkage covariance estimator

Benoît Oriol

CEREMADE, Université Paris-Dauphine, PSL, Paris, France
Société Générale Corporate and Investment Banking, Puteaux, France
benoit.oriol@dauphine.eu

Abstract

Multi-target linear shrinkage is an extension of the standard single-target linear shrinkage for covariance estimation. We combine several constant matrices - the targets - with the sample covariance matrix. We derive the oracle and a *bona fide* multi-target linear shrinkage estimator with exact and empirical mean. In both settings, we proved its convergence towards the oracle under Kolmogorov asymptotics. Finally, we show empirically that it outperforms other standard estimators in various situations.

Keywords: Multi-target, linear shrinkage, covariance estimation, general asymptotics

1 Introduction and related work

The covariance matrix plays a major role in numerous machine learning algorithms and statistics. Just to cite a few, array signal processing [1], [2], generalized method of moments estimators [3], financial portfolio management [4], [5] and forecasting [6], functional genomics [7]. However, those algorithms are designed to use the population covariance matrix, which is often unaccessible. The sample covariance matrix severely fails in many applications: for instance, for Markowitz portfolio management it does often not beat a naive uniform diversification among the assets [8].

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 $\frac{p_n}{n} \rightarrow c > 0$, the sample covariance eigenvalue spectrum is spread: high eigenvalues tends at being too high, and low ones, too low [9], [10], [11].

Influenced by the work of Stein on Gaussian mean estimation in 1956 [12], we focus on a specific type of regularization: shrinkage estimation. The motivation is to design a covariance estimator that behaves well in high dimension, with a lower variance than the sample covariance at the cost of an increased bias. This variance-bias trade-off aims at reducing the overall mean-squared error (MSE) of the covariance estimator with respect to the population covariance Σ . Linear shrinkage estimators linearly combine the sample covariance matrix with a finite set of predefined constant matrices - the targets.

More formally, we denote by S the sample covariance, and we consider $N \in \mathbb{N}$ targets, denoted $(T^{(i)})_{i=1}^N$, which are constant matrices of the same size as S . Usually, when there is only one target, $N = 1$, we chose $T^{(1)} = I$ the identity matrix [13]. With $N = 2$ targets, a standard example is the following. Consider the a diagonal block population covariance $\Sigma = \begin{pmatrix} \Sigma_1 & (0) \\ (0) & \Sigma_2 \end{pmatrix}$. Even if Σ_1 and Σ_2 are unknown, we can leverage our knowledge of the block structure, and chose the following targets:

$$T^{(1)} = \begin{pmatrix} I_1 & (0) \\ (0) & (0) \end{pmatrix} \text{ and } T^{(2)} = \begin{pmatrix} (0) & (0) \\ (0) & I_2 \end{pmatrix}. \quad (1)$$

where I_1 and I_2 are the identity matrix of same size as, respectively, Σ_1 and Σ_2 . A linear shrinkage covariance estimator $\hat{\Sigma}$ is, by definition:

$$\hat{\Sigma} = c_0 S + \sum_{i=1}^N c_i T^{(i)}, \text{ where } (c_0, \dots, c_N) \in \mathbb{R}^N. \quad (2)$$

It is a linear combination of the sample covariance and the targets. The challenge is to estimate (c_0, \dots, c_N) only with the observations in order to minimize the Frobenius norm $\|\hat{\Sigma} - \Sigma\|_F^2$, where Σ is the population covariance.

Many single-target shrinkage estimators (STSE) were proposed, just to name a few: Ledoit-Wolf linear shrinkage [13], its extension for Gaussian distributions named Oracle Approximating Shrinkage (OAS) estimator [14], linear shrinkage with factor models [15], linear shrinkage for elliptical distributions with unknown mean and known radius distribution [16].

However, the performance of those algorithms highly depends on the choice of the target and its similarity to the population covariance Σ . Thus, there is no general ideal choice of target fitting all covariances Σ . The use of several targets instead of one can lead to superior classification performance [17, 18]. The methods to compute multi-target linear shrinkage estimators are less numerous. With only two targets, an analytic estimator is given in [17]. The general case for any number of targets an estimator was proposed for Gaussian distributions in [19]. In the distribution-free case, an implicit analytic estimator was proposed in [18], retrieved from solving a strictly convex problem. Nevertheless, to the best of our knowledge, there is no theoretical analysis of any of those algorithms comparable to the quadratic convergences we have in single-target shrinkage.

We address the lack of explicit estimators and theoretical results for multi-target shrinkage covariance estimation. This paper proposes an explicit analytic distribution-free multi-target shrinkage estimator (MTSE). It provides an asymptotic theoretical analysis, proving its quadratic convergence to the oracle estimator in the scope of Kolmogorov asymptotics. Experimentally, our estimator outperforms single-target estimators in the sense of MSE in various situations.

The paper is organized as follows. In section 2, we introduce the notation, definitions and hypotheses necessary to define and address the MTSE problem. In section 3, we expose the problem, the explicit oracle and proposed estimator, and show its quadratic convergence. Section 4 presents the numerical simulations. Finally, section 5 exposes the principal limitations of the approach.

For easier reading, technical proofs of the theoretical results are derived in the appendix, we will only give the idea of proof for intuition in the main corpus.

2 Notation, definitions and hypotheses

Let us introduce the following notation.

Notation 1 (Observations). *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 (Norm). *Let A_n and B_n be two $p_n \times p_n$ matrices. We consider the Frobenius norm: $\|A_n\|_n = \sqrt{\text{tr}(A_n A_n^T)/p_n}$, 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 (Convergence). *Let $(E_n)_n$ be 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 define the sample covariance in the setting of known and unknown mean, and the set of targets.

Definition 1 (Empirical covariance). *As we are studying both situations of known or unknown mean, we will use a different sample covariance S_n depending on the situation. If we know the mean of the distribution, for an observation matrix X_n of size $p_n \times n$, we define the empirical covariance as:*

$$S_n = (X_n - \mathbb{E}[X])(X_n - \mathbb{E}[X])^T/n. \quad (3)$$

Otherwise, if we do not know the mean of the distribution, we will use the unbiased sample covariance:

$$S_n = \tilde{X}_n \tilde{X}_n^T / (n-1), \text{ with } (\tilde{X}_n)_{ik} = (X_n)_{ik} - \frac{1}{n} \sum_{k'=1}^n (X_n)_{ik'}. \quad (4)$$

Definition 2 (Set of targets). *For each n , we define the finite sequence of $N_n \in \mathbb{N}^*$ targets: $\left(T_n^{(k)}\right)_{k=1}^{N_n}$, where each $T_n^{(k)}$ is a symmetric matrix of size (p_n, p_n) , not necessarily positive. We assume that for each n , $\left(T_n^{(k)}\right)_{k=1}^{N_n}$ form a linearly independent family, orthogonal and normalized.*

The “orthogonal and normalized” condition is not really a restriction in practice. Indeed, we can build any linearly independent family $T_n^{(k)}$ of symmetric matrices. Then, we can use the Gram-Schmidt algorithm [20] on this family to make it ortho-normalized, without modifying $\text{Span} \left(\left(T_n^{(k)}\right)_{k=1}^{N_n} \right)$, so without changing the space of multi-target estimators of the form $\hat{\Sigma} = c_0 S + \sum_{i=1}^N c_i T^{(i)}$, where $(c_0, \dots, c_N) \in \mathbb{R}^N$.

So, we consider without loss of generality that the family $\left(T_n^{(k)}\right)_{k=1}^{N_n}$ is orthogonal and normalized.

We describe now several assumptions, extensions to our framework of the assumptions used in the linear shrinkage of Ledoit and Wolf [13], which will be used in the following. The first one is the core hypothesis in Kolmogorov asymptotics: the dimension can grow along with the number of samples.

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

Bounded 8-th moments are necessary for the proofs, even if experimentally we will see that bounded 4-th moments seem sufficient for the convergence to hold. For the notation, y_{i1}^n designates the i -th dimension of the first sample in Y_n . As all samples are i.i.d., it could be any other arbitrary choice of sample in $\llbracket 1, n \rrbracket$ of course. Similarly, X_1^n is the first sample, of size p_n , from X_n .

Assumption 2 (Moments). *There exists a constant K_2 independent of n such that $\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbb{E} [(y_{i1}^n)^8] \leq K_2$ and for all $k \in \llbracket 1, N_n \rrbracket$, $\mathbb{E} \left[\left\langle X_1^n (X_1^n)^T, T_n^{(k)} \right\rangle_n^4 \right] \leq K_2$.*

The number of target N_n can depend of n , while staying negligible compared to n . Intuitively, each target comes with an additional coefficient to estimate, and the estimation error decreases the performance of the *bona fide* estimator.

Assumption 3 (Number of targets). *The number of targets N_n verifies that: $N_n = o(n)$.*

Already present in [13], this technical assumption, not really restrictive in practice as argued in the same article, is still necessary here.

Assumption 4 (Technical covariance assumption). *With Q_n 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 have:*

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

Finally, for the MTSE problem to have a.s. a unique solution, we need the following characterization.

In practice, if $\|S_n\|_n^2 - \sum_k \langle S_n, T_n^{(k)} \rangle_n^2 = 0$, i.e. if $S_n \in \text{Span} \left(\left(T_n^{(k)}\right)_{k=1}^{N_n} \right)$, simply return S_n .

Assumption 5 (Well-posed problem). *For each $n \in \mathbb{N}^*$, using the associated probability \mathbb{P}_n of the distribution indexed by n ,*

$$\mathbb{P}_n \left[\|S_n\|_n^2 - \sum_k \langle S_n, T_n^{(k)} \rangle_n^2 = 0 \right] = 0.$$

3 Oracle and proposed estimator

Remark 1. As all the estimators in the following are invariant by translation in the case where we do not know the mean, we can suppose $\mathbb{E}[X] = 0$ to make the notation lighter. Indeed, we only work with $(\tilde{X}_n)_{ik} = (X_n)_{ik} - \frac{1}{n} \sum_{k'=1}^n (X_n)_{ik'}$, which is not affected by any change of mean.

Given a sample covariance S_n , we are looking for a minimizer of the 2-norm in the scope of linear combinations of S_n and $T_n^{(k)}$.

$$\underset{c \in \mathbb{R}^{N_n+1}}{\text{minimize}} \left\| c_0 S_n + \sum_{k=1}^{N_n} c_k T_n^{(k)} - \Sigma_n \right\|_n^2. \quad (5)$$

The minimizer $c^* \in \mathbb{R}^{N_n+1}$, solve the following equation:

$$A_n c^* = b_n, \text{ with } A_n = \begin{pmatrix} \|S_n\|_n^2 & \langle S_n, T_n^{(1)} \rangle_n & \dots & \langle S_n, T_n^{(N_n)} \rangle_n \\ \langle S_n, T_n^{(1)} \rangle_n & 1 & & (0) \\ \vdots & & \ddots & \\ \langle S_n, T_n^{(N_n)} \rangle_n & (0) & & 1 \end{pmatrix}, \quad (6)$$

$$b_n = [\langle S_n, \Sigma_n \rangle_n, \langle T_n^{(1)}, \Sigma_n \rangle_n, \dots, \langle T_n^{(N_n)}, \Sigma_n \rangle_n]^T.$$

We have:

$$\det(A_n) = \|S_n\|_n^2 - \sum_k \langle S_n, T_n^{(k)} \rangle_n^2,$$

$$\text{com}(A_n)^T = \begin{pmatrix} 1 & -\langle S_n, T_n^{(1)} \rangle_n & \dots & -\langle S_n, T_n^{(N_n)} \rangle_n \\ -\langle S_n, T_n^{(1)} \rangle_n & & & \\ \vdots & & (B_n) & \\ -\langle S_n, T_n^{(N_n)} \rangle_n & & & \end{pmatrix}, \quad (7)$$

$$\text{with } B_n = \left(\|S_n\|_n^2 - \sum_{k=1}^{N_n} \langle S_n, T_n^{(k)} \rangle_n^2 \right) I_{N_n} + Q Q^T, Q = [\langle S_n, T_n^{(i)} \rangle_n]_{i=1, \dots, N_n}^T.$$

Under Assumption 5, A_n is invertible and we have: $A_n^{-1} = \frac{1}{\det(A_n)} \text{com}(A_n)^T$.

Remark 2. In the particular case where A_n is not invertible (not supposing Assumption 5), remove a target $T_n^{(i)}$ such as $\langle S_n, T_n^{(i)} \rangle \neq 0$ and the new built A_n will be invertible.

Previous derivations lead to the following oracle estimators.

Definition 3 (Oracle estimators). The oracle estimators solving the minimization problem 5 are, for $i \in \llbracket 1, N_n \rrbracket$:

$$c_0^* = \frac{1}{\det(A_n)} \left(\langle S_n, \Sigma_n \rangle_n - \sum_{k=1}^{N_n} \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right), c_i^* = \langle \Sigma_n, T_n^{(i)} \rangle_n - c_0^* \langle S_n, T_n^{(i)} \rangle_n. \quad (8)$$

And we define the oracle estimator of covariance $\Sigma_n^* = \mathcal{P}_{S_n^+} \left(c_0^* S_n + \sum_{k=1}^{N_n} c_k^* T_n^{(k)} \right)$, where $\mathcal{P}_{S_n^+}$ is the orthogonal projection on the cone of symmetric positive matrices.

The orthogonal projection $\mathcal{P}_{S_n^+}$ at the end does not strictly follows the definition of linear shrinkage. However, as Σ_n is symmetric positive, the projection step decreases the distance $\|\Sigma_n^* - \Sigma_n\|_n$. This step is not necessary for the convergence results to hold, but it is a free improvement we recommend.

Nevertheless, the oracle estimator still depends on Σ_n , and cannot be computed as it is. We propose a *bona fide* estimator, and show that asymptotically it converges to the oracle one. The strategy is the following, both for known and unknown mean:

- For $i \in \llbracket 1, N_n \rrbracket$, in c_i^* , $\langle \Sigma_n, T_n^{(i)} \rangle_n$ is estimated by $\langle S_n, T_n^{(i)} \rangle_n$,
- $\langle S_n, \Sigma_n \rangle_n$ is approximated by $\|S_n\|_n^2 - \underbrace{\mathbb{E}[\|S_n - \Sigma_n\|_n^2]}_{\mathbb{V}(S_n)}$. Note that in expectation both

objects are equal. $\mathbb{V}(S_n)$ is estimated by $\hat{\mathbb{V}}(S_n)$ defined below,

- Similarly, for each $k \in \llbracket 1, N_n \rrbracket$, $\langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n$ is approximated by $\langle S_n, T_n^{(k)} \rangle_n^2 - \underbrace{\mathbb{E}[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^2]}_{\mathbb{V}(\langle S_n, T_n^{(k)} \rangle_n)}$. $\mathbb{V}(\langle S_n, T_n^{(k)} \rangle_n)$ is estimated by $\hat{\mathbb{V}}(\langle S_n, T_n^{(k)} \rangle_n)$, defined below.

Definition 4 (Known mean - Variance estimators). *We suppose the mean to be **known**. We define:*

$$\hat{\mathbb{V}}(S_n) = \frac{1}{n(n-1)} \sum_{k=1}^n \left\| X_{\cdot k}^n (X_{\cdot k}^n)^T - S_n \right\|_n^2. \quad (9)$$

Let $i \in \llbracket 1, N_n \rrbracket$. We define:

$$\hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) = \frac{1}{n(n-1)} \sum_{k=1}^n \left\langle X_{\cdot k}^n (X_{\cdot k}^n)^T - S_n, T_n^{(i)} \right\rangle_n^2. \quad (10)$$

Definition 5 (Unknown mean - Variance estimators). *We suppose the mean to be **unknown**. We define:*

$$\begin{aligned} \bar{b}_n^2 &= \frac{1}{n^2} \sum_{k=1}^n \left\| \frac{n}{n-1} \tilde{X}_{\cdot k}^n \left(\tilde{X}_{\cdot k}^n \right)^T - S_n \right\|_n^2, \\ \hat{\mathbb{V}}(S_n) &= \frac{(n-1)^2}{(n-2)(n-3)} \bar{b}_n^2 - \frac{1}{n(n-2)} \|S_n\|_n^2 - \frac{n-1}{n(n-2)(n-3)} p_n \langle S_n, I_{p_n} \rangle_n^2. \end{aligned} \quad (11)$$

Let $i \in \llbracket 1, N_n \rrbracket$. We define:

$$\begin{aligned} \bar{b}_{T^{(i)},n}^2 &= \frac{p_n}{(n-1)^2} \sum_{k=1}^n \left\langle \tilde{X}_{\cdot k}^n \left(\tilde{X}_{\cdot k}^n \right)^T, T_n^{(i)} \right\rangle_n^2 - \frac{1}{n} \left\langle S_n T_n^{(i)}, T_n^{(i)} S_n \right\rangle_n, \\ \hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) &= q_n^{(0)} \bar{b}_{T^{(i)},n}^2 + q_n^{(1)} \left\langle S_n T_n^{(i)}, T_n^{(i)} S_n \right\rangle_n - q_n^{(2)} \langle S_n, T_n^{(i)} \rangle_n^2, \\ q_n^{(0)} &= \frac{(n-1)^2}{p_n(n-2)(n-3)}, q_n^{(1)} = \frac{n-1}{p_n n(n-2)}, q_n^{(2)} = \frac{n^2 - 2n - 1}{n(n-2)(n-3)}. \end{aligned} \quad (12)$$

Remark 3 (Implementation trick). $\bar{b}_{T^{(i)},n}^2$ can be made fast to compute in Python with Numpy. We use a complex matrix square root of $T_n^{(i)}$, denoted $R_n^{(i)}$. We have a $O(p_n^3 + np_n)$ -complexity formula:

$$\sum_{k=1}^n \left\langle \tilde{X}_{\cdot k}^n \left(\tilde{X}_{\cdot k}^n \right)^T, T_n^{(i)} \right\rangle_n^2 = \frac{1}{p_n^2} \sum_{k=1}^n \left(\sum_{i=1}^{p_n} (R_n^{(i)} \tilde{X})_{i,k}^2 \right)^2. \quad (13)$$

Lemma 1 (Coordinate estimators). *Under Assumptions 1 to 3, for all $k \in \llbracket 1, N_n \rrbracket$, $\langle S_n, T_n^{(k)} \rangle_n$ is an unbiased estimator of $\langle \Sigma_n, T_n^{(k)} \rangle_n$. Moreover, it exists $M \in \mathbb{R}_+$, independent of n such as $\sum_{k=1}^{N_n} \mathbb{E}[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^4] \leq \frac{M}{n} K_1 K_2$, where K_1 and K_2 are defined in Assumptions 1 and 2.*

The proof works around the fact that $\left(T_n^{(k)} \right)_k$ is an orthonormal family to use Assumption 2 then.

Theorem 1 (Variance estimators). *Denoting $\mathbb{V}(S_n) = \mathbb{E}[\|S_n - \Sigma_n\|_n^2]$ and $\mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n) = \mathbb{E}[\langle S_n - \Sigma_n, T_n^{(i)} \rangle_n^2]$, under assumptions 1 to 4, in both situations where the mean is known or unknown for all $i \in \llbracket 1, N_n \rrbracket$, $\hat{\mathbb{V}}(S_n)$ and $\hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n)$ are unbiased estimators of $\mathbb{V}(S_n)$ and $\mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n)$ respectively. Moreover, $\mathbb{E}[\left(\hat{\mathbb{V}}(S_n) - \mathbb{V}(S_n) \right)^2] \rightarrow 0$ and $\mathbb{E}[\left(\sum_{i=1}^{N_n} \hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) - \mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n) \right)^2] \rightarrow 0$.*

The proof of this theorem resides in carefully developing for the first part. The convergences are proven by making appear the assumptions 3 and 4 by Cauchy-Schwarz and Jensen inequalities.

Definition 6 (Bona fide estimator). *The bona fide estimators c for the minimization problem 5 are defined, for $i \in \llbracket 1, N_n \rrbracket$, as:*

$$c_0 = \min((\hat{c}_0)_+, 1), \text{ with } \hat{c}_0 = \frac{1}{\det(A_n)} \left(\|S_n\|_n^2 - \hat{\mathbb{V}}(S_n) - \sum_{k=1}^{N_n} \left(\langle S_n, T_n^{(k)} \rangle_n^2 - \hat{\mathbb{V}}(\langle S_n, T_n^{(k)} \rangle_n) \right) \right),$$

$$c_i = (1 - c_0) \langle S_n, T_n^{(i)} \rangle_n, \quad (14)$$

The MTSE is defined as: $S_n^* = \mathcal{P}_{S_n^+} \left(c_0 S_n + (1 - c_0) \sum_{i=1}^{N_n} c_i T_n^{(i)} \right)$.

The truncation of c_0 between 0 and 1 theoretically does not change the asymptotic behavior of the estimator. Geometrically, it ensures that S_n^* is a convex combination between S_n and the matrix $\sum_{i=1}^{N_n} c_i T_n^{(i)}$. Empirically, it improves the robustness of the estimator, particularly in small sample size.

Immediately from Theorem 1, we have the following result.

Corollary 1. *Under assumptions 1 to 4, we have $(\hat{c}_1 - c_1^*) \det(A_n) \xrightarrow{q.m} 0$.*

Theorem 2 (Loss convergence). *Under assumptions 1 to 5, S_n^* quadratically converges to the optimal estimator Σ_n^* , i.e. $S_n^* - \Sigma_n^* \xrightarrow{q.m} 0$. Consequently, S_n^* has asymptotically the same loss as Σ_n^* , i.e.*

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

The idea of the proof is to prove the convergence on the estimators without the projection step onto the cone of positive matrices, and use the Lipschitz property of projection to conclude. The initial convergence relies on the Lemma A.1 in [13]. The check of the hypotheses for the use of the Lemma comes essentially from the previous results we proved.

4 Experimental results

This section illustrates the practical consequences of the choice of targets on the estimation, the robustness of the estimator to heavy tails, and an application to real non-stationary time series. The experiments were computed on a Lenovo laptop, Intel i5, 32Go Ram. The performance is measured with the Percentage Relative Improvement in Average Loss (PRIAL):

$$\text{PRIAL}(\hat{S}_n) = \frac{\mathbb{E}[\|S_n - \Sigma_n\|_n^2] - \mathbb{E}[\|\hat{S}_n - \Sigma_n\|_n^2]}{\mathbb{E}[\|S_n - \Sigma_n\|_n^2]}. \quad (15)$$

4.1 Target set impact: Usefulness of a target

We fix the dimension at $p = 50$, and $n = 25$, and we study the impact of the number of targets N_n . The population covariance Σ is designed to clearly exhibit a choice of useful targets, and compare the role of targets that give information on the population covariance - i.e. that improve significantly the oracle PRIAL-, to those which give less. Σ is built as follows: this is a block diagonal matrix composed of 10 block matrices $(B_i)_{i=1}^{10}$ of shape $(5, 5)$. For each $i \in \llbracket 1, 10 \rrbracket$, B_i is drawn from a Wishart distribution with 5 degrees of freedom and $\sqrt{11 - i} I_5$ as scale matrix.

$$\Sigma = \begin{pmatrix} B_1 & & (0) \\ & \ddots & \\ (0) & & B_{10} \end{pmatrix} \quad (16)$$

The underlying distribution of X is a multivariate t-distribution with $\nu = 9$.

We consider two sets of targets: one set of “aligned” targets $(T_1)^{(i)}$, and one set of “misaligned” ones $(T_2)^{(i)}$. The first one is built as following: the first target $T_1^{(1)} = I$, to match Ledoit-Wolf STSE [13],

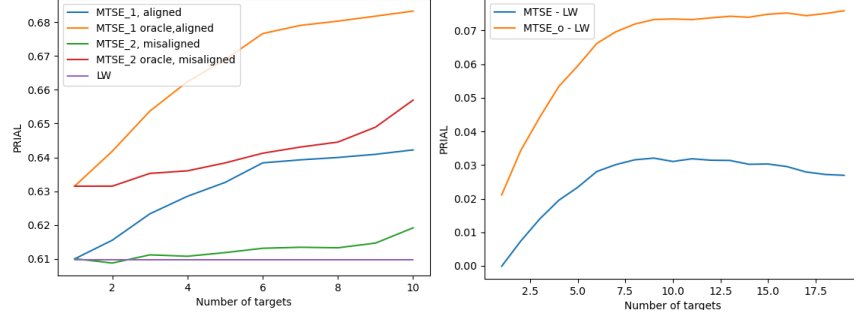


Figure 1: (left) PRIAL of the MTSE, and the oracle, with target sets 1 and 2, and the Ledoit-Wolf STSE as reference, in function of the number of targets in the experiment of target usefulness.

Figure 2: (right) PRIAL difference between the oracle MTSE and Ledoit Wolf linear shrinkage one (MTSE_o - LW) in orange, and estimator MTSE and Ledoit Wolf linear shrinkage one (MTSE - LW) in blue, in function of the number of targets in the setup of the experiment of target set impact.

and for $i \in \llbracket 2, 10 \rrbracket$, the target $T_1^{(i)}$ is block-diagonal; where Σ has the block B_i , $T_1^{(i)}$ has an Identity block; elsewhere $T_1^{(i)}$ has null values.

$$T_1^{(i)} = \begin{pmatrix} (0) & \dots & (0) \\ \vdots & (\text{block } i :)I & \vdots \\ (0) & \dots & (0) \end{pmatrix} \quad (17)$$

The set of target is orthogonalized with the Gram-Schmidt algorithm.

For the second set of targets $(T_2)^{(i)}$, we break the perfect alignment between $T_1^{(i)}$ and the block B_i . For that, we apply a permutation σ on both rows and columns of $T_1^{(i)}$, where σ is defined as:

$$\sigma : j \in \llbracket 1, p \rrbracket \longrightarrow (j + 2) \bmod p. \quad (18)$$

For the target $i \in \llbracket 1, 10 \rrbracket$, at index $(k, l) \in \llbracket 1, p \rrbracket^2$, we define $(T_2^{(i)})_{\sigma(k), \sigma(l)} = (T_1^{(i)})_{k, l}$.

We consider the PRIAL of the MTSE (MTSE_1 and MTSE_2 respectively for target set 1 and 2), the oracle ones, and the Ledoit-Wolf STSE (LW), in function of the number k of targets used, between $k = 1$, which is equivalent to the STSE, and $k = 10$ where the sets of targets $(T_1^{(i)})_{i=1}^k$ and $(T_2^{(i)})_{i=1}^k$ are used completely. The results, with 50000 draws, are shown in Figure 1. The maximum standard deviation is 4×10^{-4} , error bars were removed.

It confirms the intuition about what is a useful target: the “aligned” set $(T_1)^{(i)}$ presents a significant better PRIAL than the second misaligned one $(T_2)^{(i)}$. This comprehension of what makes a target useful is of utter importance to design the target set in a practical use case. The second conclusion is that even if the misaligned targets are not optimal, they still present a better PRIAL than the STSE using only one target.

4.2 Target set impact: Adding useless targets to a good target set

We keep the previous setting for p, n, Σ and the samples X but we change the set of targets.

The first target is the identity matrix I , to match with Ledoit-Wolf STSE. For $i \in \llbracket 2, 10 \rrbracket$, $T^{(i)}$ is designed according to Equation 17. For target $i \geq 11$, the target is drawn from a Wishart distribution with p degrees of freedom and I_p as scale matrix. The latter are supposed to be “useless” regarding the structure of Σ . The set of targets is then orthogonalized with the Gram-Schmidt algorithm.

We ran the experiment for 5000 times, the maximum standard deviation on any point of the figure is 7×10^{-4} , we did not plot the boxes for readability. We plot the **differences** of PRIAL between the MTSE and the Ledoit-Wolf STSE, which has only one target. We do the same for the MTSE oracle. While the difference of PRIAL is positive, it means that adding additional targets improved the PRIAL. The result can be seen in figure 2.

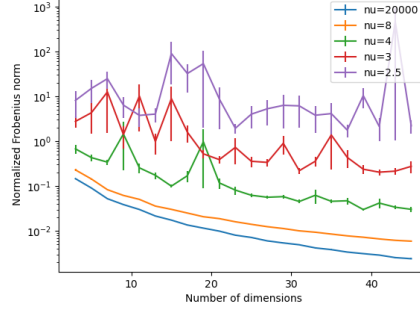


Figure 3: (right) Convergence of $\|S_n^* - \Sigma_n^*\|_F^2$ for different values of ν , where ν is the degree of freedom of a multivariate t-Student distribution.

Two distinct behaviors emerge from this experiment. The first 10 useful targets results in a significant improvement of the PRIAL. For the random targets, they can be tagged as useless as they do not improve significantly the oracle PRIAL. The interesting conclusion is that the estimator PRIAL is only slightly negatively affected but stays above the STSE PRIAL. Consequently, adding good targets gives a high reward, and adding useless targets comes at a low price.

4.3 Heavy-tails robustness

In the case of Ledoit-Wolf STSE, Ledoit and Wolf empirically argue that Assumption 2 about finite eighth moments is in practice not necessary, only finite fourth moments would suffice [13]. Thus, we study the convergence of S_n^* towards Σ_n^* in a setting of heavy tails. For that, we consider multivariate t-Student distributions with $\nu > 2$ degrees of freedom, for different values of ν .

We consider Σ block diagonal with 5 blocks of same size drawn independently from Wishart distributions of $p/5$ degrees of freedom, and the set of targets is made of 5 block diagonal matrices, target i with the Identity as block i and 0 elsewhere. We compare, for 20000 independent experiments using t-Students with $\nu \in \{20000, 8, 4, 3, 2.5\}$ degrees of freedom. $\nu = 20000$, verifying all Assumptions, is a reference; $\nu = 8$ is at the edge of Assumption 2, $\nu = 4$ is at the edge of the empirical conjecture of Ledoit and Wolf, $\nu = 3$ is what we found to be the frontier for the convergence to hold, and $\nu = 2.5$ is an extreme case.

This experiment, which results are in Figure 3, strengthens the hypothesis of Ledoit and Wolf [13] saying that bounded fourth moments is experimentally enough.

4.4 Application to non-stationary time series

In order to give the practitioner a real example on an application of the MTSE where a “structure” prior is known and can be used, we consider a classical portfolio management problem in finance.

We use daily log-returns of 323 assets that stayed in the S&P500 between 2010 and 2022¹. We recall that the log-return r_t at day t of a price s_t is defined as the log-increment: $r_t = \log(s_t) - \log(s_{t-1})$. So we have a 300-dimensional time series, with one vector per day.

Those 323 assets, are classified in 11 industrial sectors, known as the GICS sectors. The general structural idea we want to use is that log-returns of different GICS sectors are less correlated than within the same sector. We expect to have a covariance structure similar to a stochastic block matrix, not as perfect as in Equation 16 due to other mechanisms, but we expect to take advantage of it.

To evaluate the performance of different covariance estimators, we used the General Minimum Variance - GMV - portfolio problem, a standard method to compare covariance estimators. The idea of GMV is to determine at the beginning of each month T a weight vector $w_T \in \mathbb{R}^{323}$, as there are 323 assets here, and invest $w_{i,T}$ on the asset i for one month: this is our portfolio.

At each day t of the month T , around 20 business days, we collect the log-returns of our portfolio $R_t = w_T^T r_t$, and at the end of the month we store the empirical variance of our portfolio during the

¹Data were retrieved from YahooFinance using yfinance package.

Estimator	3 months	4 months	6 months	9 months	12 months	15 months
MTSE	7.40×10^{-3}	7.13×10^{-3}	7.30×10^{-3}	7.49×10^{-3}	8.12×10^{-3}	8.03×10^{-3}
LWO	7.60×10^{-3}	7.34×10^{-3}	7.51×10^{-3}	7.70×10^{-3}	8.57×10^{-3}	8.27×10^{-3}
OAS	7.63×10^{-3}	7.42×10^{-3}	7.74×10^{-3}	8.24×10^{-3}	9.63×10^{-3}	9.52×10^{-3}
ANS	7.62×10^{-3}	7.41×10^{-3}	7.56×10^{-3}	7.91×10^{-3}	1.15×10^{-2}	2.14×10^{-2}
QIS	3.73×10^{-2}	7.83×10^{-3}	8.56×10^{-3}	1.04×10^{-2}	1.68×10^{-2}	4.90×10^{-2}
S	5.55×10^{-2}	7.28×10^{-2}	5.84×10^{-2}	6.89×10^{-2}	8.95×10^{-2}	4.04×10^{-1}

Table 1: GMV Cumulative Variance for different estimators and number of months to fit them.

month: $v_T = \mathbb{V}[\{R_t\}_{t=T}^{T+20}]$. The weight vector w_T is designed to minimize this variance v_T , under constraint that $w_T^T \mathbf{1} = 1$. There is closed-form formula, where Σ is the covariance matrix of the random vector r_t : $w_T = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$. The pseudo-inverse is used when not invertible.

This problem depends only on the covariance and is perfect to compare different estimators. The experiment goes as following: at the beginning of each month T , for each estimator $\hat{\Sigma}$, we estimate it on the last K months, *i.e* roughly $20K$ data points, and we compute the weight vector $\hat{w}_T = \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}}$. At the end of the month we store the empirical variance \hat{v}_T . We do this procedure over 135 months, between 2010 and 2022, and we compare the Cumulative Variance over every months $\sum_{T=1}^{135} v_T$, the results are shown in the Table 1. The better the estimator $\hat{\Sigma}$, the lower is the Cumulative Variance.

The number K of months used to fit the covariance estimators varies between 3 and 15. The estimators we compare, along with sample covariance S , are:

- MTSE, where the 11 targets are chosen to be the Identity matrix on the sub-selection of assets belonging to the same GICS sector, and 0 on the other indexes of the matrix,
- two STSE: LWO, the corrected Ledoit-Wolf estimator in the setting of unknown mean [21], and OAS, the Oracle Approximated Shrinkage estimator [14],
- two non-linear shrinkage: ANS, the Analytical Nonlinear Shrinkage estimator [22], and QIS, the Quadratic Inverse Shrinkage estimator [23].

The first remark is that increasing the number of samples, from 3 to 15 months, does not improve the numerical performance of the estimators we consider. That comes from the non-stationarity of financial time series, one of its stylized facts [24]. It virtually makes the number of samples low.

In this setting, the additional information we plug in MTSE makes the difference in the experiment, as it consistently has a lower Cumulative Variance than the other considered algorithms.

5 Limitations

The strongest assumptions made in the theoretical analysis are that we have n i.i.d. samples and bounded eighth moments. In practice, the i.i.d. assumption is hard to assess, particularly in time series analysis, as we saw in the financial dataset. Eighth moments can be a limitations, even if we saw that experimentally the convergence seems to hold for lower moments.

As in single-target estimation, the central point for performance in a task is a meticulous choice of targets, and general methods to design them still lack in the literature to the best of our knowledge.

6 Conclusion

We derived the oracle estimator solving the multi-target linear shrinkage covariance estimator and proposed an explicit analytic estimator of it. The main result of the contribution is the proof of the convergence of the estimator under Kolmogorov asymptotics, while the number of targets remains negligible compared to the number of samples.

The numerical simulations highlight the gain in performance to add useful targets to the estimation. Experimentally, we argue that the estimator convergence holds with only bounded fourth moments. The experiment in a non-stationary setup underlines the potential of adaptability of multi-target estimation in those situations, compared to other linear or non-linear shrinkage methods.

References

- [1] Richard Abrahamsson, Yngve Selén, and Peter Stoica. Enhanced covariance matrix estimators in adaptive beamforming. In *2007 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol II, Pts 1-3*, number 2 in International Conference on Acoustics Speech and Signal Processing (ICASSP), pages 969–972, 2007.
- [2] J.R. Guerci. Theory and application of covariance matrix tapers for robust adaptive beamforming. *IEEE Transactions on Signal Processing*, 47(4):977–985, 1999.
- [3] Lars Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54, 1982.
- [4] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- [5] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30, 07 2003.
- [6] Ruey S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 2. ed. edition, 2005.
- [7] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4:Article32, 02 2005.
- [8] 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.
- [9] V. L. Girko. G-analysis of high dimensional observations. *Journal of Soviet mathematics*, 1985.
- [10] V. L. Girko. *Theory of Random Determinants*. Springer Netherlands, Dordrecht, 1990.
- [11] Vyacheslav L. Girko. *G2-estimator for the Stieltjes Transform of the Normalized Spectral Function of Covariance Matrices*, pages 171–212. Springer Netherlands, Dordrecht, 1995.
- [12] 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.
- [13] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.
- [14] 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.
- [15] Yuki Ikeda and Tatsuya Kubokawa. Linear shrinkage estimation of large covariance matrices using factor models. *Journal of Multivariate Analysis*, 152:61–81, 2016.
- [16] 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.
- [17] Zohar Halbe, Maria Bortman, and Mayer Aladjem. Regularized mixture density estimation with an analytical setting of shrinkage intensities. *Neural Networks and Learning Systems, IEEE Transactions on*, 24:460–470, 03 2013.
- [18] Tomer Lancelwiczki and Mayer Aladjem. Multi-target shrinkage estimation for covariance matrices. *IEEE Transactions on Signal Processing*, 62(24):6380–6390, 2014.
- [19] Harry Gray, Gwenaél G. R. Leday, Catalina A. Vallejos, and Sylvia Richardson. Shrinkage estimation of large covariance matrices using multiple shrinkage targets, 2018.
- [20] J. P. Gram. Ueber die entwicklung reeller functionen in reihen mittels der methode der kleinsten quadrate. *Journal für die Reine und Angewandte Mathematik*, 94:41–73, 1883.
- [21] Benoit Oriol and Alexandre Miot. Ledoit-wolf linear shrinkage with unknown mean, 2023.
- [22] Olivier Ledoit and Michael Wolf. Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5):3043 – 3065, 2020.
- [23] Olivier Ledoit and Michael Wolf. Quadratic shrinkage for large covariance matrices. *SSRN Electronic Journal*, 28, 11 2019.
- [24] Vaibhav Sherkar and Rituparna Sen. Study of stylized facts in stock market data, 2023.

7 Theoretical proofs

In this appendix, we develop the proofs of the theoretical results of the work.

7.1 Proof of Lemma 1

By linearity of the expectation, for all $k \in \llbracket 1, N_n \rrbracket$, $\langle S_n, T_n^{(k)} \rangle_n$ is an unbiased estimator of $\langle \Sigma_n, T_n^{(k)} \rangle_n$.

For the second part of the lemma, let $k \in \llbracket 1, N_n \rrbracket$ and let denote $T = T_n^{(k)}$. The following proof relies upon the Proposition 1 and its method of graph construction from indices in the Preliminary combinatorial result in [21]. We recall the method and the proposition here.

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 (Proposition 1 from [21]). *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} .*

Known mean When the mean is known, choosing $M' \in \mathbb{R}_+$ the number of graphs with 4 nodes and 2 connected components, we have:

$$\begin{aligned} \mathbb{E} [\langle S_n - \Sigma_n, T \rangle_n^4] &= \frac{1}{n^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} \left[\prod_{m \in \{k_1, k_2, k_3, k_4\}} \langle X_m^n (X_m^n)^T - \Sigma_n, T \rangle_n \right] \\ \mathbb{E} [\langle S_n - \Sigma_n, T \rangle_n^4] &\leq \frac{M' n(n-1)}{n^4} \mathbb{E} [\langle X_1^n (X_1^n)^T - \Sigma_n, T \rangle_n^2]^2 + \frac{n}{n^4} \mathbb{E} [\langle X_1^n (X_1^n)^T - \Sigma_n, T \rangle_n^4]. \end{aligned} \quad (19)$$

Moreover, we have:

$$\begin{aligned} \sum_{k=1}^{N_n} \mathbb{E} [\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(k)} \rangle_n^4] &\leq \mathbb{E} \left[\left(\sum_k \langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(k)} \rangle_n^2 \right)^2 \right] \\ &\leq \mathbb{E} \left[\left\| X_1^n (X_1^n)^T - \Sigma_n \right\|_n^4 \right] \\ &\leq \mathbb{E} \left[\left\| X_1^n (X_1^n)^T \right\|_n^4 \right] \\ &= \frac{1}{p_n^2} \mathbb{E} \left[\left(\sum_i y_{i1}^2 \right)^4 \right] \\ &= p_n^2 \mathbb{E} \left[\left(\frac{1}{p_n} \sum_i y_{i1}^2 \right)^4 \right] \\ &\leq p_n^2 \frac{1}{p_n} \sum_i \mathbb{E} [y_{i1}^8] \\ \sum_{k=1}^{N_n} \mathbb{E} [\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(k)} \rangle_n^4] &\leq p_n^2 K_2. \end{aligned} \quad (20)$$

And,

$$\begin{aligned}
\sum_{k=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(k)} \rangle_n^2 \right] &\leq \max_i \left(\mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^2 \right] \right) \sum_{k=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(k)} \rangle_n^2 \right] \\
&\leq \max_i \left(\mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^2 \right] \right) \mathbb{E} \left[\|X_1^n (X_1^n)^T\|_n^2 \right] \\
\sum_{k=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(k)} \rangle_n^2 \right] &\leq \max_i \left(\mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^2 \right] \right) p_n \sqrt{K_2}.
\end{aligned} \tag{21}$$

For any of the targets T , we have with Assumption 2: $\mathbb{E} \left[\langle X_1^n (X_1^n)^T, T \rangle_n^2 \right] \leq \sqrt{K_2}$. Finally, using $M = M' + K_1$, we have:

$$\sum_{k=1}^{N_n} \mathbb{E} \left[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^4 \right] \leq \frac{M' + K_1}{n} K_1 K_2 = \frac{M}{n} K_1 K_2. \tag{22}$$

Unknown mean When the mean is unknown, we have a larger decomposition. We use the same technics as in the known mean setting, with more terms and cross-products to handle:

$$\begin{aligned}
\mathbb{E} \left[\langle S_n - \Sigma_n, T \rangle_n^4 \right] &= \\
\frac{1}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \mathbb{E} \left[\prod_{s=1}^4 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \right] &\quad \text{Term (1)} \\
- 4 \mathbb{E} \left[\prod_{s=1}^3 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_{k_4}^n (X_{k'_4}^n)^T - \delta_{k_4=k'_4} \Sigma_n, T \right\rangle_n \right] &\quad \text{Term (2)} \\
+ 6 \mathbb{E} \left[\prod_{s=1}^2 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \prod_{t=3}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] &\quad \text{Term (3)} \\
- 4 \mathbb{E} \left[\left\langle X_{k_1}^n (X_{k_1}^n)^T - \Sigma_n, T \right\rangle_n \prod_{t=2}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] &\quad \text{Term (4)} \\
+ \mathbb{E} \left[\prod_{t=1}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] &\quad \text{Term (5)}
\end{aligned} \tag{23}$$

We prove the higher bound of the form $M'' K_1 K_2 / n$ term by term, in absolute value, summing over all targets $(T_n^{(k)})_{k=1}^{N_n}$. For that, we decompose each term and bound them by expectations we can control.

Term (1) This term is equal to $\frac{n^4}{(n-1)^4} \mathbb{E} \left[\langle S_n^{(\text{known mean})} - \Sigma_n, T \rangle_n^4 \right]$. Supposing $n \geq 4$, we have $\frac{n^4}{(n-1)^4} \leq 4$, so, with $M_1'' = 4M$, we have:

$$\sum_{k=1}^{N_n} \frac{1}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{s=1}^4 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \right] \right| \leq \frac{M_1''}{n} K_1 K_2. \tag{24}$$

Term (2) For this term, we have the following development, when discarding 0-expectation terms:

$$\begin{aligned}
& \frac{4}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{s=1}^3 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_{k_4}^n (X_{k'_4}^n)^T - \delta_{k_4=k'_4} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{4}{n} \times |\text{Term(1)}| + \frac{24}{(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^2 \left\langle X_2^n (X_2^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_2^n)^T, T \right\rangle_n \right] \right| \\
& \leq \frac{4}{n} \times |\text{Term(1)}| + \frac{24}{(n-1)^3} \sqrt{\mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^4 \right]^{3/2} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right]} \\
& \leq \frac{4}{n} \times |\text{Term(1)}| + \frac{6}{(n-1)^3} \left(3 \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^4 \right] + \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right]^2 \right). \tag{25}
\end{aligned}$$

At this point, we recall that we have to sum over each target $(T_n^{(k)})_{k=1}^{N_n}$ and then bound by a term of the form $M_2'' K_1 K_2 / n$ for some $M_2'' \in \mathbb{R}$. Term(1) is obviously controlled by the previous part. For the rest, we use the two following inequalities:

- from the proof in the known mean setting, we have that:

$$\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(k)} \right\rangle_n^4 \right] \leq p_n^2 K_2, \tag{26}$$

- and:

$$\begin{aligned}
\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n^2 \right] & \leq \max_i \left(\mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(i)} \right\rangle_n^2 \right] \right) \sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n \right] \\
& \leq \max_i \left(\mathbb{E} \left[\frac{1}{p_n^2} (X_2^n)^T T_n^{(i)} X_1^n (X_1^n)^T T_n^{(i)} X_2^n \right] \right) \mathbb{E} \left[\|X_1^n (X_2^n)^T\|_n^2 \right] \\
& \leq \max_i \left(\frac{1}{p_n} \left\langle \Sigma_n T_n^{(i)}, T_n^{(i)} \Sigma_n \right\rangle_n \right) p_n \sqrt{K_2} \\
& \leq \max_i \left(\frac{1}{p_n} \left\| \Sigma_n T_n^{(i)} \right\|_n^2 \right) p_n \sqrt{K_2} \\
& \leq \|\Sigma_n\|_n^2 p_n \sqrt{K_2} \\
\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n^2 \right] & \leq p_n K_2. \tag{27}
\end{aligned}$$

Supposing $n \geq 4$, with $M_2'' = M_1'' + 54K_1$, we then have:

$$\begin{aligned}
& \sum_{k=1}^{N_n} \frac{4}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{s=1}^3 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_{k_4}^n (X_{k'_4}^n)^T - \delta_{k_4=k'_4} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{M_2''}{n} K_1 K_2. \tag{28}
\end{aligned}$$

Term (3) Similarly for this term, we obtain:

$$\begin{aligned}
& \frac{6}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{s=1}^2 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \prod_{t=3}^4 \left\langle X_{k_t}^n (X_{k_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{6(2n-1)}{n^2} \times |\text{Term(2)}| \\
& \quad + \frac{12(n-2)}{n(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^2 \left\langle X_2^n (X_3^n)^T - \Sigma_n, T \right\rangle_n^2 \right] \right| \\
& \quad + \frac{48(n-2)}{n(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_2^n (X_2^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_3^n)^T, T \right\rangle_n \left\langle X_2^n (X_3^n)^T, T \right\rangle_n \right] \right| \\
& \quad + \frac{24}{n(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_2^n (X_2^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right] \right| \\
& \quad + \frac{24}{n(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^2 \left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right] \right| \\
& \leq \frac{6(2n-1)}{n^2} \times |\text{Term(2)}| + \frac{24}{n(n-1)^2} \left(\mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^4 \right] + \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^4 \right] \right) \\
& \quad + \frac{6(n-2)}{n(n-1)^3} \left(\mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^2 \right]^2 + \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right]^2 \right). \tag{29}
\end{aligned}$$

We use the two following inequalities to finish the work, combined with the two ones we used for Term (2):

- from the proof in the known mean setting, we have that:

$$\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(k)} \right\rangle_n^2 \right]^2 \leq p_n K_2, \tag{30}$$

- and:

$$\begin{aligned}
\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n^4 \right] & \leq \mathbb{E} \left[\left(\sum_{k=1}^{N_n} \left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n^2 \right)^2 \right] \\
& \leq \mathbb{E} \left[\left\| X_1^n (X_2^n)^T \right\|_n^4 \right] \\
& = \frac{1}{p_n^2} \mathbb{E} \left[\left(\sum_i y_{i1} y_{i2} \right)^4 \right] \\
& \leq p_n^2 \frac{1}{p_n} \sum_i \mathbb{E} [y_{i1}^4 y_{i2}^4] \\
& \leq p_n^2 \frac{1}{p_n} \sum_i \mathbb{E} [y_{i1}^8]
\end{aligned} \tag{31}$$

$$\sum_{k=1}^{N_n} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T_n^{(k)} \right\rangle_n^4 \right] \leq p_n^2 K_2.$$

Supposing $n \geq 4$, with $M_3'' = 3M_2'' + 86K_1 + 1$, we then have:

$$\begin{aligned}
& \sum_{k=1}^{N_n} \frac{6}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{s=1}^2 \left\langle X_{k_s}^n (X_{k_s}^n)^T - \Sigma_n, T \right\rangle_n \prod_{t=3}^4 \left\langle X_{k_t}^n (X_{k_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{M_3''}{n} K_1 K_2. \tag{32}
\end{aligned}$$

Term (4) For that one, we have:

$$\begin{aligned}
& \frac{4}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\left\langle X_{k_1}^n (X_{k_1}^n)^T - \Sigma_n, T \right\rangle_n \prod_{t=2}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{4(n-1)(n+3)}{n^3} \times |\text{Term(3)}| \\
& \quad + \frac{96(n-2)}{n^2(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_2^n)^T, T \right\rangle_n \left\langle X_2^n (X_3^n)^T, T \right\rangle_n \left\langle X_3^n (X_1^n)^T, T \right\rangle_n \right] \right| \\
& \quad + \frac{96(n-2)}{n^2(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_2^n)^T, T \right\rangle_n \left\langle X_2^n (X_3^n)^T, T \right\rangle_n^2 \right] \right| \\
& \quad + \frac{24}{n^2(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n \left\langle X_1^n (X_2^n)^T, T \right\rangle_n^3 \right] \right| \\
& \leq \frac{4(n-1)(n+3)}{n^3} \times |\text{Term(3)}| + \\
& \quad \frac{48n-88}{n^2(n-1)^3} \left(3\mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^4 \right] + \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T \right\rangle_n^2 \right]^2 \right)
\end{aligned} \tag{33}$$

Supposing $n \geq 4$, with $M_4'' = M_3'' + 47K_1 + 4$, we then have:

$$\begin{aligned}
& \sum_{k=1}^{N_n} \frac{4}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\left\langle X_{k_1}^n (X_{k_1}^n)^T - \Sigma_n, T_k^{(n)} \right\rangle_n \prod_{t=2}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T_k^{(n)} \right\rangle_n \right] \right| \\
& \leq \frac{M_4''}{n} K_1 K_2.
\end{aligned} \tag{34}$$

Term (5) Following the same idea, we have for the last term:

$$\begin{aligned}
& \frac{1}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{t=1}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T \right\rangle_n \right] \right| \\
& \leq \frac{n^3 + n^2 - 4}{n^4} \times |\text{Term(4)}| + \frac{8}{n^3(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^4 \right] \right| \\
& \quad + \frac{24(n-2)}{n^3(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \left\langle X_1^n (X_3^n)^T, T \right\rangle_n^2 \right] \right| \\
& \quad + \frac{12(n-2)(n-3)}{n^3(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \left\langle X_3^n (X_4^n)^T, T \right\rangle_n^2 \right] \right| \\
& \quad + \frac{48(n-2)(n-3)}{n^3(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \left\langle X_2^n (X_3^n)^T, T \right\rangle_n \left\langle X_3^n (X_1^n)^T, T \right\rangle_n \right] \right| \\
& \quad + \frac{48(n-2)(n-3)}{n^3(n-1)^3} \left| \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n \left\langle X_2^n (X_3^n)^T, T \right\rangle_n \left\langle X_3^n (X_4^n)^T, T \right\rangle_n \left\langle X_4^n (X_1^n)^T, T \right\rangle_n \right] \right| \\
& \leq \frac{n^3 + n^2 - 4}{n^4} \times |\text{Term(4)}| + \frac{72n+8}{n^3(n-1)^3} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^4 \right] \\
& \quad + \frac{60(n-2)(n-3)}{n^3(n-1)^3} \mathbb{E} \left[\left\langle X_1^n (X_2^n)^T, T \right\rangle_n^2 \right]^2.
\end{aligned} \tag{35}$$

Supposing $n \geq 4$, with $M_5'' = M_4'' + 11K_1 + 2$, we then have:

$$\sum_{k=1}^{N_n} \frac{1}{n^4(n-1)^4} \sum_{\substack{k_1, k_2, k_3, k_4 \\ k'_1, k'_2, k'_3, k'_4}} \left| \mathbb{E} \left[\prod_{t=1}^4 \left\langle X_{k_t}^n (X_{k'_t}^n)^T - \delta_{k_t=k'_t} \Sigma_n, T_k^{(n)} \right\rangle_n \right] \right| \leq \frac{M_5''}{n} K_1 K_2. \quad (36)$$

Finally, combining the 5 terms, we have with $M'' = \sum_{i=1}^5 M_i''$:

$$\sum_{k=1}^{N_n} \mathbb{E} \left[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^4 \right] \leq \frac{M''}{n} K_1 K_2, \quad (37)$$

which concludes the proof for the unknown mean.

7.2 Proof of Theorem 1

Known mean When the mean is known, $\hat{\mathbb{V}}(S_n)$ is trivially unbiased when developing. $\mathbb{E} \left[\left(\hat{\mathbb{V}}(S_n) - \mathbb{V}(S_n) \right)^2 \right] \rightarrow 0$ is a consequence of Lemma 3.4 in [13] about the quadratic mean convergence of $\bar{b}_n^2 - \beta_n^2 = \frac{n-1}{n} \hat{\mathbb{V}}(S_n) - \mathbb{V}(S_n)$.

$\hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n)$ is trivially unbiased when developing. Let's prove that $\mathbb{E} \left[\left(\sum_i \hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) - \mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n) \right)^2 \right] \rightarrow 0$.

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_i \frac{1}{n(n-1)} \sum_k \left\langle X_k^n (X_k^n)^T - \Sigma_n, T_n^{(i)} \right\rangle_n^2 - \mathbb{E} \left[\left\langle X_k^n (X_k^n)^T - \Sigma_n, T_n^{(i)} \right\rangle_n^2 \right] \right)^2 \right] \\ &= \frac{1}{n(n-1)^2} \mathbb{E} \left[\left(\sum_i \left\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(i)} \right\rangle_n^2 - \mathbb{E} \left[\left\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(i)} \right\rangle_n^2 \right] \right)^2 \right] \\ &\leq \frac{N_n}{n(n-1)^2} \mathbb{E} \left[\sum_i \left\langle X_1^n (X_1^n)^T - \Sigma_n, T_n^{(i)} \right\rangle_n^4 \right] \\ &\leq \frac{N_n}{n(n-1)^2} \times 7 \mathbb{E} \left[\sum_i \left\langle X_1^n (X_1^n)^T, T_n^{(i)} \right\rangle_n^4 \right] \\ &\leq 7 \frac{nN_n}{(n-1)^2} K_1^2 K_2. \end{aligned} \quad (38)$$

And,

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_i \frac{1}{n-1} \left\langle S_n - \Sigma_n, T_n^{(i)} \right\rangle_n^2 - \mathbb{E} \left[\frac{1}{n-1} \left\langle S_n - \Sigma_n, T_n^{(i)} \right\rangle_n^2 \right] \right)^2 \right] \\ &\leq \frac{N_n}{(n-1)^2} \sum_i \mathbb{E} \left[\left\langle S_n - \Sigma_n, T_n^{(i)} \right\rangle_n^4 \right] \\ &\leq \frac{N_n M K_1 K_2}{n(n-1)^2}. \end{aligned} \quad (39)$$

(Lemma 1)

Which concludes the proof when the mean is known.

Unknown mean When the mean is unknown, $\hat{\mathbb{V}}(S_n)$ is unbiased thanks to Lemma 9 in [21]. $\mathbb{E} \left[\left(\hat{\mathbb{V}}(S_n) - \mathbb{V}(S_n) \right)^2 \right] \rightarrow 0$ is a consequence of Lemma 9 in [21] about the quadratic mean convergence of $b_n^2 - \beta_n^2 = \hat{\mathbb{V}}(S_n) - \mathbb{V}(S_n)$.

$\hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n)$ is unbiased thanks to Lemma 8 in [21] and the development of the variance. Let's prove that $\mathbb{E} \left[\left(\sum_i \hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) - \mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n) \right)^2 \right] \rightarrow 0$. For that, we need to prove the convergence of three terms:

$$(I.) \quad \mathbb{V} \left[\frac{1}{(n-1)^2} \sum_{i=1}^{N_n} \sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right] \rightarrow 0,$$

$$(II.) \quad \mathbb{V} \left[\frac{1}{pn} \sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right] \rightarrow 0,$$

$$(III.) \quad \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^{N_n} \langle S_n, T_n^{(i)} \rangle_n^2 \right] \rightarrow 0.$$

Term (I) We first show that $\mathbb{V} \left[\frac{1}{(n-1)^2} \sum_{i=1}^{N_n} \sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right] \rightarrow 0$.

$$\begin{aligned} & \mathbb{V} \left[\frac{1}{(n-1)^2} \sum_{i=1}^{N_n} \sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right] \\ & \leq \mathbb{E} \left[\left(\frac{1}{(n-1)^2} \sum_{i=1}^{N_n} \sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \\ & \leq \frac{N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \\ & = \frac{N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n^2 - 4 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right. \right. \\ & \quad \left. \left. + 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 + \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 - 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right. \right. \\ & \quad \left. \left. + 2 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \\ & \leq \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \quad \text{Term (I.1)} \\ & \quad + \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \quad \text{Term (I.2)} \\ & \quad + \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \quad \text{Term (I.3)} \\ & \quad + \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \quad \text{Term (I.4)} \\ & \quad + \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \quad \text{Term (I.5)} \\ & \quad + \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 2 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \quad \text{Term (I.6).} \\ & \quad (40) \end{aligned}$$

With this decomposition, we have 6 distinct terms, and we prove the convergence to 0 of each one.

Term (I.1)

$$\begin{aligned} \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] &\leq \frac{6nN_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^4 \right] \\ &\leq \frac{6nN_n p_n^2 K_2}{(n-1)^2} \end{aligned} \quad (41)$$

$$\frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \longrightarrow 0.$$

Term (I.2)

$$\begin{aligned} &\frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \\ &= \frac{6 \times 16N_n}{n(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^4 \right] + \frac{6 \times 16N_n}{n(n-1)^3} \sum_{i=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^2 \langle X_1^n (X_2^n)^T, T_n^{(i)} \rangle_n^2 \right] \\ &\leq \frac{96nN_n}{(n-1)^4} K_1^2 K_2 + \frac{96nN_n}{(n-1)^3} K_1^2 K_2 \longrightarrow 0. \end{aligned} \quad (42)$$

Term (I.3)

$$\begin{aligned} &\frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \\ &\leq \frac{96n^2 N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\langle X_1^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^4 \right] \\ &= \frac{96N_n}{n^2(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^4 + 6(n-1) \langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n^2 \langle X_1^n (X_2^n)^T, T_n^{(i)} \rangle_n^2 \right. \\ &\quad \left. + 3(n-1)(n-2) \langle X_1^n (X_2^n)^T, T_n^{(i)} \rangle_n^2 \langle X_1^n (X_3^n)^T, T_n^{(i)} \rangle_n^2 \right. \\ &\quad \left. + 4(n-1) \langle X_1^n (X_1^n)^T, T_n^{(i)} \rangle_n \langle X_1^n (X_2^n)^T, T_n^{(i)} \rangle_n^3 \right] \\ &\leq 96 \frac{(3n+1)nN_n}{(n-1)^4} K_1^2 K_2 \longrightarrow 0. \end{aligned} \quad (43)$$

Term (I.4)

$$\begin{aligned} &\frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \\ &= \frac{6N_n n^2}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \\ &\leq \frac{6N_n n^2}{(n-1)^4} 8 \underbrace{\left(\sum_{i=1}^{N_n} \mathbb{E} \left[\left(\langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 - \frac{1}{n} \Sigma_n, T_n^{(i)} \rangle_n^2 \right)^2 \right] \right)}_{= \frac{(n-1)^4}{n^4} \times (\text{term 5, Lemma 1})} + \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\langle \frac{1}{n} \Sigma_n, T_n^{(i)} \rangle_n^2 \right)^2 \right] \quad (44) \\ &\leq \frac{6N_n n^2}{(n-1)^4} \times 8 \left(\frac{(n-1)^4}{n^4} \frac{M_5''}{n} K_1 K_2 + \frac{K_2}{n^4} \right) \longrightarrow 0, \end{aligned}$$

using M_5'' from Equation 36.

Term (I.5)

$$\begin{aligned} & \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 4 \langle X_k^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \\ &= \frac{6 \times 16N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \longrightarrow 0. \end{aligned} \quad (45)$$

Term (I.6)

$$\begin{aligned} & \frac{6N_n}{(n-1)^4} \sum_{i=1}^{N_n} \mathbb{E} \left[\left(\sum_{k=1}^n 2 \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n \right)^2 \right] \\ &\leq \frac{12N_n}{(n-1)^4} \sum_{i=1}^{N_n} \left(\mathbb{E} \left[\left(\sum_{k=1}^n \langle X_k^n (X_k^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] + \mathbb{E} \left[\left(\sum_{k=1}^n \langle \bar{X}^n (\bar{X}^n)^T, T_n^{(i)} \rangle_n^2 \right)^2 \right] \right) \longrightarrow 0. \end{aligned} \quad (46)$$

The convergence to 0 of each one of the 6 terms leads to the convergence of **term (I)**:

$$\mathbb{V} \left[\frac{1}{(n-1)^2} \sum_{i=1}^{N_n} \sum_{k=1}^n \langle \tilde{X}_k^n (\tilde{X}_k^n)^T, T_n^{(i)} \rangle_n^2 \right] \longrightarrow 0. \quad (47)$$

Term (II) We then show that $\mathbb{V} \left[\frac{1}{np_n} \sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right] \longrightarrow 0$.

$$\begin{aligned} \mathbb{V} \left[\frac{1}{np_n} \sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right] &\leq \mathbb{E} \left[\frac{1}{n^2 p_n^2} \left(\sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right)^2 \right] \\ &\leq \frac{N_n}{n^2 p_n^2} \sum_{i=1}^{N_n} \mathbb{E} \left[\sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n^2 \right] \\ &\leq \frac{N_n^2}{n^2 p_n} \mathbb{E} [\|S_n\|_n^2] \\ \mathbb{V} \left[\frac{1}{np_n} \sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right] &\leq \frac{N_n^2}{n^2 p_n} \mathbb{E} [\|S_n - \Sigma_n\|_n^2 + \|\Sigma_n\|_n^2]. \end{aligned} \quad (48)$$

Thanks to Lemma 1 [21], particularly in the proof of the Lemma 1 equations (16) and (17) about $\beta_n^2 = \mathbb{E} [\|S_n - \Sigma_n\|_n^2]$, we have that $\frac{n}{p_n} \mathbb{E} [\|S_n - \Sigma_n\|_n^2]$ is bounded. Additionnally, $\|S_n\|_n^2 \leq \frac{p_n}{n} \sqrt{K_n}$. So:

$$\mathbb{V} \left[\frac{1}{np_n} \sum_{i=1}^{N_n} \langle S_n T_n^{(i)}, T_n^{(i)} S_n \rangle_n \right] \longrightarrow 0. \quad (49)$$

Term (III) We finally show that $\mathbb{V} \left[\frac{1}{n} \sum_{i=1}^{N_n} \langle S_n, T_n^{(i)} \rangle_n^2 \right] \longrightarrow 0$.

$$\begin{aligned} \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^{N_n} \langle S_n, T_n^{(i)} \rangle_n^2 \right] &\leq \frac{N_n}{n^2} \sum_{i=1}^{N_n} \mathbb{E} [\langle S_n, T_n^{(i)} \rangle_n^4] \\ \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^{N_n} \langle S_n, T_n^{(i)} \rangle_n^2 \right] &\leq \frac{N_n}{n^2} \mathbb{E} [\|S_n\|_n^4] \end{aligned} \quad (50)$$

Thanks to the proof of Lemma 3 [21], equation (54), we have that $\mathbb{V} [\|S_n\|_n^2]$ is bounded. Moreover, from Lemma 1 [21], we have that $\mathbb{E} [\|S_n\|_n^2]$ is bounded too. So $\mathbb{E} [\|S_n\|_n^4]$ is bounded. Consequently,

$$\mathbb{V} \left[\frac{1}{n} \sum_{i=1}^{N_n} \langle S_n, T_n^{(i)} \rangle_n^2 \right] \longrightarrow 0. \quad (51)$$

Backing up, the convergences of **Term (I)-(III)** prove that:

$$\mathbb{E} \left[\left(\sum_i \hat{\mathbb{V}}(\langle S_n, T_n^{(i)} \rangle_n) - \mathbb{V}(\langle S_n, T_n^{(i)} \rangle_n) \right)^2 \right] \rightarrow 0. \quad (52)$$

7.3 Proof of Theorem 2

We denote:

$$\begin{aligned} a_n^2 &= c_1 \det(A_n), \alpha_n = c_1^* \det(A_n), d_n^2 = \det(A_n), \delta_n^2 = \mathbb{E} \left[\left\| S_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n^2 \right], \\ u_n^2 &= (a_n^2 - \alpha_n)^2. \end{aligned} \quad (53)$$

We will first prove that $\mathbb{E} \left[\left\| \tilde{S}_n^* - \tilde{\Sigma}_n^* \right\|_n^2 \right] \rightarrow 0$ and deduce that $\mathbb{E} \left[\left\| S_n^* - \Sigma_n^* \right\|_n^2 \right] \rightarrow 0$ by a property of the orthogonal projection.

We have:

$$\mathbb{E} \left[\left\| \tilde{S}_n^* - \tilde{\Sigma}_n^* \right\|_n^2 \right] = \mathbb{E} \left[\frac{u_n^2}{d_n^2} \right] + \sum_k \mathbb{V} \left[\langle S_n, T_n^{(k)} \rangle_n \right]. \quad (54)$$

Firstly, we have thanks to Lemma 1:

$$\begin{aligned} \sum_k \mathbb{V} \left[\langle S_n, T_n^{(k)} \rangle_n \right] &= \sum_k \mathbb{E} \left[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^2 \right] \\ &\leq \sum_k \sqrt{\mathbb{E} \left[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^4 \right]} \\ &\leq \sqrt{N_n \sum_k \mathbb{E} \left[\langle S_n - \Sigma_n, T_n^{(k)} \rangle_n^4 \right]} \\ &\leq \sqrt{\frac{N_n M K_1 K_2}{n}} \\ \sum_k \mathbb{V} \left[\langle S_n, T_n^{(k)} \rangle_n \right] &\rightarrow 0. \end{aligned} \quad (55)$$

Then, we have that $d_n^2 - \delta_n^2 \xrightarrow{q.m} 0$. Indeed:

$$d_n^2 - \delta_n^2 = \|S\|_n^2 - \mathbb{E}[\|S\|_n^2] - \sum_k \left(\langle \Sigma_n, T_n^{(k)} \rangle_n^2 - \langle S_n, T_n^{(k)} \rangle_n^2 \right) \quad (56)$$

So, we have:

$$\begin{aligned} \mathbb{E} \left[(d_n^2 - \delta_n^2)^2 \right] &\leq 2\mathbb{V} [\|S\|_n^2] + 2\mathbb{V} \left[\sum_k \langle \Sigma_n - S_n, T_n^{(k)} \rangle_n^2 \right] \\ &\leq 2\mathbb{V} [\|S\|_n^2] + 2\mathbb{E} \left[\left(\sum_k \langle \Sigma_n - S_n, T_n^{(k)} \rangle_n^2 \right)^2 \right] \\ &\leq 2\mathbb{V} [\|S\|_n^2] + 2N_n \sum_k \mathbb{E} \left[\langle \Sigma_n - S_n, T_n^{(k)} \rangle_n^4 \right] \\ \mathbb{E} \left[(d_n^2 - \delta_n^2)^2 \right] &\leq 2\mathbb{V} [\|S\|_n^2] + 2N_n \frac{M K_1 K_2}{n}. \end{aligned} \quad (\text{Lemma 1}) \quad (57)$$

$\mathbb{V} [\|S\|_n^2] \rightarrow 0$ from the proof of Lemma 3 in [21] and from the proof of Lemma 3.3 in [13]. So we have $d_n^2 - \delta_n^2 \xrightarrow{q.m} 0$.

Now, let's prove $a_n^2 - \mathbb{E}[\alpha_n] \xrightarrow{q.m.} 0$.

Firstly, let's remark that $0 \leq \mathbb{E}[\alpha_n] \leq \delta_n^2$. Indeed:

$$\begin{aligned}\mathbb{E}[\alpha_n] &= \mathbb{E} \left[\langle S_n, \Sigma_n \rangle_n - \sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right] \\ &= \|\Sigma_n\|_n^2 - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \\ &= \left\| \Sigma_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n^2 \\ \mathbb{E}[\alpha_n] &\geq 0\end{aligned}\tag{58}$$

And,

$$\begin{aligned}\delta_n^2 &= \mathbb{E} \left[\left\| S_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n^2 \right] \\ &= \mathbb{E} [\|S_n\|_n^2] - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \\ &= \mathbb{E}[\alpha_n] + \mathbb{E} [\|S_n - \Sigma_n\|_n^2] \\ \delta_n^2 &\geq \mathbb{E}[\alpha_n]\end{aligned}\tag{59}$$

So, we have $0 \leq \mathbb{E}[\alpha_n] \leq \delta_n^2$.

Let's find a lower and an upper bound for $a_n^2 - \mathbb{E}[\alpha_n]$. For the upper bound:

$$\begin{aligned}a_n^2 - \mathbb{E}[\alpha_n] &= \min((\hat{c}_1 d_n^2)_+, d_n^2) - \mathbb{E}[\alpha_n] \\ &\leq (\hat{c}_1 d_n^2)_+ - \mathbb{E}[\alpha_n] \\ &\leq |\hat{c}_1 d_n^2 - \mathbb{E}[\alpha_n]| \quad (\text{as } \mathbb{E}[\alpha_n] \geq 0) \\ a_n^2 - \mathbb{E}[\alpha_n] &\leq \max(|\hat{c}_1 d_n^2 - \mathbb{E}[\alpha_n]|, |d_n^2 - \delta_n^2|)\end{aligned}\tag{60}$$

For the lower bound:

$$\begin{aligned}a_n^2 - \mathbb{E}[\alpha_n] &= \min((\hat{c}_1 d_n^2)_+ - \mathbb{E}[\alpha_n], d_n^2 - \mathbb{E}[\alpha_n]) \\ &\geq \min((\hat{c}_1 d_n^2)_+ - \mathbb{E}[\alpha_n], d_n^2 - \delta_n^2) \quad (\text{as } \mathbb{E}[\alpha_n] \leq \delta_n^2) \\ &\geq \min(-|\hat{c}_1 d_n^2 - \mathbb{E}[\alpha_n]|, -|d_n^2 - \delta_n^2|) \\ a_n^2 - \mathbb{E}[\alpha_n] &\geq -\max(|\hat{c}_1 d_n^2 - \alpha_n|, |d_n^2 - \delta_n^2|)\end{aligned}\tag{61}$$

So,

$$\begin{aligned}\mathbb{E} \left[(a_n^2 - \mathbb{E}[\alpha_n])^2 \right] &\leq \mathbb{E} \left[\max(|\hat{c}_1 d_n^2 - \alpha_n|, |d_n^2 - \delta_n^2|)^2 \right] \\ &\leq \mathbb{E} \left[(\hat{c}_1 d_n^2 - \alpha_n)^2 \right] + \mathbb{E} \left[(d_n^2 - \delta_n^2)^2 \right]\end{aligned}\tag{62}$$

From Corollary 1, $\mathbb{E} \left[(\hat{c}_1 d_n^2 - \alpha_n)^2 \right] \rightarrow 0$, and we proved above that $\mathbb{E} \left[(d_n^2 - \delta_n^2)^2 \right] \rightarrow 0$. So, $a_n^2 - \mathbb{E}[\alpha_n] \xrightarrow{q.m.} 0$. Then, let's prove that $\mathbb{V}[\alpha_n] \rightarrow 0$.

$$\begin{aligned}\mathbb{V}[\alpha_n] &= \mathbb{V}[c_1^* \det(A_n)] \\ &= \mathbb{V} \left[\langle S_n, \Sigma_n \rangle_n - \sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right] \\ \mathbb{V}[\alpha_n] &\leq 2\mathbb{V}[\langle S_n, \Sigma_n \rangle_n] + 2\mathbb{V} \left[\sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right].\end{aligned}\tag{63}$$

$\mathbb{V}[\langle S_n, \Sigma_n \rangle_n] \rightarrow 0$ from the proof of Theorem 3 [21] for the unknown mean and the proof of Theorem 3.3 [13] for the known mean.

For the second term, we have:

$$\begin{aligned}
\mathbb{V} \left[\sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right] &= \mathbb{E} \left[\left(\sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right)^2 \right] - \left(\sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \right)^2 \\
&\leq \mathbb{E} \left[\sum_k \langle S_n, T_n^{(k)} \rangle_n^2 \right] \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 - \left(\sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \right)^2 \\
&= \left(\sum_k \mathbb{V} \left[\langle S_n, T_n^{(k)} \rangle_n \right] \right) \left(\sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \right) \\
&\leq \sqrt{\frac{N_n M K_1 K_2}{n}} \left(\sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n^2 \right) \text{ (Eq 55)} \\
&\leq \sqrt{\frac{N_n M K_1 K_2}{n}} \|\Sigma_n\|^2 \\
\mathbb{V} \left[\sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right] &\longrightarrow 0.
\end{aligned} \tag{64}$$

So, as both terms converges to 0, we have that $\mathbb{V}[\alpha_n] \rightarrow 0$.

Easily, we have then that $\mathbb{E}[u_n^2] \rightarrow 0$ as $\mathbb{E}[u_n^2] \leq 2 (\mathbb{E}[(d_n^2 - \mathbb{E}[\alpha_n])^2] + \mathbb{V}[\alpha_n])$.

Moreover, let's check that $0 \leq \frac{u_n^2}{d_n^2} \leq 2(d_n^2 + \delta_n^2)$.

$$\begin{aligned}
|\alpha_n| &= \left| \langle S_n, \Sigma_n \rangle_n - \sum_k \langle S_n, T_n^{(k)} \rangle_n \langle \Sigma_n, T_n^{(k)} \rangle_n \right| \\
&= \left| \left\langle S_n - \sum_k \langle S_n, T_n^{(k)} \rangle_n T_n^{(k)}, \Sigma_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\rangle_n \right| \\
&\leq \left\| S_n - \sum_k \langle S_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n \left\| \Sigma_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n \\
|\alpha_n| &\leq d_n \delta_n.
\end{aligned} \tag{65}$$

And, by construction, $0 \leq a_n^2 \leq d_n^2$, so $0 \leq \frac{u_n^2}{d_n^2} \leq 2(d_n^2 + \delta_n^2)$.

Backing up, we proved that $d_n^2 - \delta_n^2 \xrightarrow{q.m} 0$, $\mathbb{E}[u_n^2] \rightarrow 0$, and $0 \leq \frac{u_n^2}{d_n^2} \leq 2(d_n^2 + \delta_n^2)$. We then can

apply Lemma A.1 from [13] with $\tau_1 = 2, \tau_2 = 0$, which proves that $\mathbb{E} \left[\frac{u_n^2}{d_n^2} \right] \rightarrow 0$.

Backing up again, it proves that $\tilde{S}_n^* - \tilde{\Sigma}_n^* \xrightarrow{q.m} 0$.

Finally, $\mathcal{P}_{S_n^+}$ is 1-Lipschitz for the norm $\|\cdot\|_n$, so $S_n^* - \Sigma_n^* \xrightarrow{q.m} 0$.

The second part of the theorem follows:

$$\begin{aligned}
\mathbb{E} \left[\left| \|S_n^* - \Sigma_n\|_n^2 - \|\Sigma_n^* - \Sigma_n\|_n^2 \right| \right] &= \mathbb{E} [|\langle S_n^* - \Sigma_n^*, S_n^* + \Sigma_n^* - 2\Sigma_n \rangle_n|] \\
&\leq \sqrt{\mathbb{E} [\|S_n^* - \Sigma_n^*\|_n^2]} \sqrt{\mathbb{E} [\|S_n^* + \Sigma_n^* - 2\Sigma_n\|_n^2]}.
\end{aligned} \tag{66}$$

Moreover, $\mathbb{E} [\|S_n^* - \Sigma_n\|_n^2]$ is bounded. In fact:

$$\begin{aligned}
\mathbb{E} [\|S_n^* - \Sigma_n\|_n^2] &\leq \mathbb{E} [\|\tilde{S}_n^* - \Sigma_n\|_n^2] \\
&\leq \mathbb{E} \left[2\|S_n - \Sigma_n\|_n^2 + 2 \left\| \sum_k \langle S_n, T_n^{(k)} \rangle_n T_n^{(k)} - \Sigma_n \right\|_n^2 \right] \\
&= 2\mathbb{E} [\|S_n - \Sigma_n\|_n^2] + 2 \left\| \Sigma_n - \sum_k \langle \Sigma_n, T_n^{(k)} \rangle_n T_n^{(k)} \right\|_n^2 + 2 \sum_k \mathbb{V} [\langle \Sigma_n, T_n^{(k)} \rangle_n] \\
\mathbb{E} [\|S_n^* - \Sigma_n\|_n^2] &\leq 2\mathbb{E} [\|S_n - \Sigma_n\|_n^2] + 2\|\Sigma_n\|_n^2 + 2 \sum_k \mathbb{V} [\langle \Sigma_n, T_n^{(k)} \rangle_n]
\end{aligned} \tag{67}$$

$\mathbb{E} [\|S_n - \Sigma_n\|_n^2]$ and $\|\Sigma_n\|_n^2$ are bounded from Lemma 3.1 in [13] and Lemma 1 in [21], and $\sum_k \mathbb{V} [\langle \Sigma_n, T_n^{(k)} \rangle_n] \rightarrow 0$, so $\mathbb{E} [\|S_n^* - \Sigma_n\|_n^2]$ is bounded. So, using the first part of the Theorem 2 we've just proved, we have that: $\mathbb{E} [\left| \|S_n^* - \Sigma_n\|_n^2 - \|\Sigma_n^* - \Sigma_n\|_n^2 \right|] \rightarrow 0$.