



**HAL**  
open science

## Exponential bounds for regularized Hotelling's T<sub>2</sub> statistic in high dimension

El Mehdi Issouani, Patrice Bertail, Emmanuelle Gautherat

► **To cite this version:**

El Mehdi Issouani, Patrice Bertail, Emmanuelle Gautherat. Exponential bounds for regularized Hotelling's T<sub>2</sub> statistic in high dimension. *Journal of Multivariate Analysis*, 2024, 203, pp.105342. 10.1016/j.jmva.2024.105342 . hal-04696035

**HAL Id: hal-04696035**

**<https://hal.science/hal-04696035>**

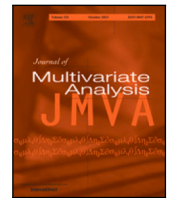
Submitted on 12 Sep 2024

**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 4.0 International License



# Exponential bounds for regularized Hotelling's T2 statistic in high dimension

El Mehdi Issouani <sup>a,\*</sup>, Patrice Bertail <sup>a,1</sup>, Emmanuelle Gautherat <sup>b,1</sup>

<sup>a</sup> MODAL'X, Univ. Paris Nanterre, CNRS, F92000 Nanterre, France

<sup>b</sup> Université de Reims Champagne-Ardenne, CRIEG, REGARD'S, Reims, France

## ARTICLE INFO

### AMS 202 subject classifications:

primary 62G05  
secondary 62G10  
62H10  
62G99

### Keywords:

Eigenvalues  
Exponential bounds  
High dimension  
Hotelling statistics  
Penalized covariance  
Regularized covariance matrix  
Self-normalization

## ABSTRACT

We obtain exponential inequalities for regularized Hotelling's  $T_n^2$  statistics, that take into account the potential high dimensional aspects of the problem. We explore the finite sample properties of the tail of these statistics by deriving exponential bounds for symmetric distributions and also for general distributions under weak moment assumptions (we never assume exponential moments). For this, we use a penalized estimator of the covariance matrix and propose an optimal choice for the penalty coefficient.

## 1. Introduction

In many applications (for instance in genomics or natural language processing), the dimension of the parameter of interest  $p$  is large in comparison to the sample size  $n$  and sometimes increases with  $n$ . Consider for instance the problem of estimating or testing a mean of variables in  $\mathbb{R}^p$ , with  $p > n$ ; in that case, the empirical covariance matrix is not full rank and does not approximate the true one (for the Frobenius or spectral norms) when  $n$  tends to infinity (see Theorem 2 in Mestre [24]. It is also ill-conditioned, see Johnstone [16]). As a consequence, the usual Hotelling's  $T_n^2$  tests in a large dimension framework are no longer valid (see Chen et al. [9], Li et al. [22]). It is thus important to construct estimators and testing procedures that take into account the high dimensional aspects of the problem (as done for instance in Ledoit and Wolf [19,20], Bodnar et al. [7], Yao et al. [35], see also the references therein). Many papers have considered this problem of  $p$  larger than  $n$  in the last decades starting from the works of Dempster [10] (see also the references and discussions in Bai and Saranadasa [4]). More recently, Srivastava and Du [30] proposed to use the diagonal matrix of the sample variances instead of the covariance matrix (see also some variations in Dong et al. [11]).

One relevant proposition for our framework, which has been developed in the statistical literature is to use a penalized estimator of the covariance matrix which is non-singular, and to use this matrix in tests. In that spirit, Chen et al. [9] have obtained asymptotically valid regularized Hotelling's  $T_n^2$  tests for the mean in the Gaussian case in a high dimensional framework when  $n$  and  $p \equiv p(n)$  tend to infinity at some specific rate. Li et al. [22] have extended these results to some sub-gaussian distributions. Dong et al. [11] propose some related penalized tests for comparing two high-dimensional means. The purpose of this paper is to further

\* Corresponding author.

E-mail address: [elmehdi.issouani@gmail.com](mailto:elmehdi.issouani@gmail.com) (E.M. Issouani).

<sup>1</sup> The three authors contributed equally. This paper is an extension of a chapter of the Ph.D. thesis of El Mehdi Issouani.

<https://doi.org/10.1016/j.jmva.2024.105342>

Received 22 March 2023; Received in revised form 10 June 2024; Accepted 10 June 2024

Available online 12 June 2024

0047-259X/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

explore the finite sample properties of Hotelling tests by deriving exponential bounds of some correctly regularized Hotelling’s  $T_n^2$  for general distributions, including ones with very few moments.

Exponential bounds allow for building conservative confidence regions for the parameter of interest or testing the value of a parameter in a large-dimension framework as in Chen et al. [9] and Li et al. [22]. We will develop a testing procedure exploiting our inequalities in the paper. The main advantage of these inequalities is that they are valid under very few moments assumption whereas most authors assume normality or subgaussian assumptions. They are also of interest in statistical learning to control risk even with unbounded loss functions.

For this, we derive exponential bounds for some regularized Hotelling’s  $T_n^2$  statistics in the spirit of Bertail et al. [6], who obtained bounds for self-normalized quadratic forms or the Hotelling’s  $T_n^2$  statistic when  $p < n$ . We show that for symmetric distributions, only moments of order 2 are needed and we only assume the existence of moments of order 8 for general distributions.

Let  $Z, Z_1, \dots, Z_n$  be i.i.d. centered random vectors with probability distribution  $P$ , defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with values in  $(\mathbb{R}^{p(n)}, \mathcal{B}, P)$  endowed with the  $L_2$  norm  $\|\cdot\|_2$ . We denote  $\mathbb{E}$  the expectation under  $P$ . Put  $Z^{(n)} = (Z_i)_{1 \leq i \leq n}$ . As  $n$  and  $p(n)$  go to infinity, notice that  $(Z^{(n)})_n$  defines a triangular array of random variables with varying dimensions. However, since we are interested in finite sample properties, we will drop the dependence in  $n$ . In particular, we use  $p$  instead of  $p(n)$ . But keep in mind that  $p$  is a function of  $n$  in an asymptotic framework. The covariance matrix of the observation is given by  $\Sigma_n = \mathbb{E}(ZZ^\top)$ , where we denote by  $Z^\top$  the transpose of  $Z$  and  $\Sigma_n^{1/2}$  the square root of  $\Sigma_n$ . Denote by  $\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$  the sample mean. The sample covariance matrix is defined here by

$$S_n = \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top.$$

Notice that we do not center the  $Z_i$ ’s by the empirical mean. By a slight abuse of notation, we call Hotelling’s  $T_n^2$  the statistic defined by

$$T_n^2 = n \bar{Z}_n^\top S_n^{-1} \bar{Z}_n,$$

If we use a recentered version say

$$\mathcal{T}_n^2 = n \bar{Z}_n^\top \left( \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}_n)(Z_i - \bar{Z}_n)^\top \right)^{-1} \bar{Z}_n,$$

it is known that controlling  $T_n$  is equivalent to controlling  $\mathcal{T}_n$  as shown by Eaton and Efron [13] using the relationship

$$T_n^2 = \frac{\mathcal{T}_n^2}{1 + \mathcal{T}_n^2}.$$

For some positive real numbers,  $\rho_1$  and  $\rho_2$ , define  $S_n(\rho_1, \rho_2)$  the linear combination of the identity matrix with the sample covariance matrix

$$S_n(\rho_1, \rho_2) = \rho_1 I_p + \rho_2 S_n,$$

with  $I_p$  the identity matrix of size  $p$ . For  $\rho_1 = 0$  and  $\rho_2 = 1$ ,  $S_n(0, 1) = S_n$  is the empirical covariance matrix, which is singular for  $p > n$ . When  $\rho_2 = 1$  and  $\rho_1 > 0$  (and small),  $S_n(\rho_1, 1)$  corresponds to a Tikhonov regularization of the sample covariance matrix: see Tikhonov [32]. It is precisely this estimator which is used in the tests proposed by Chen et al. [9] or Bai et al. [3] and which is studied in detail in Ledoit and Wolf [19]. In particular, it is shown in Ledoit and Wolf [19] that if one chooses adequately  $\rho_1$  and  $\rho_2$  then one can obtain a well-conditioned estimator of the covariance matrix which is invertible and more accurate than the sample covariance for some  $L_2$ -distance.

We denote by  $\tilde{\Sigma}_n$  the expectation of  $S_n(\rho_1, \rho_2)$ , which is given for any  $\rho_1, \rho_2 > 0$  by

$$\tilde{\Sigma}_n \equiv \tilde{\Sigma}_n(\rho_1, \rho_2) = \rho_1 I_p + \rho_2 \Sigma_n.$$

In the following, we are interested in Hotelling’s  $T_n^2$  statistic with a linear combination of the sample covariance and the identity, that we now call the regularized Hotelling’s  $T_n^2$  statistic defined by

$$T_n^2(\rho_1, \rho_2) = n \bar{Z}_n^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n$$

generalizing the proposal of Chen et al. [9].

In the framework of high dimension, such quantities also appear naturally when studying empirical likelihood under a lot of constraints, penalized in its dual form by an  $L_2$ -norm: see for instance Newey and Smith [25], Lahiri and Mukhopadhyay [17] among others.

When  $p < n$ , exponential bounds for  $T_n^2(0, 1)$  (that is, with the empirical covariance matrix instead of a regularized one) have been obtained by Bertail et al. [6]. Their exponential bound is controlled by two terms: (i) an exponential term corresponding to a ‘‘Hoeffding’’ or Pinelis [27] type of inequality applied to a symmetrized version of the observations and (ii) an exponential bound which essentially controls the minimum eigenvalue of the sample covariance matrix and the proximity of  $S_n$  to  $\Sigma_n$ . However, for  $p \geq n$  such inequality cannot hold since in this case the minimum of the eigenvalues of  $S_n$  is always 0. Moreover, it can easily be seen from the results of Bertail et al. [6] that the bound becomes very bad when  $p > n$  or/and when  $p$  and  $n$  are of the same order.

We obtain in this paper general results for  $T_n^2(\rho_1, \rho_2)$  for general values of  $\rho_1$  and  $\rho_2$  as well as for some adequate estimations of these parameters, when  $p$  is bigger than  $n$  and when  $p$  and  $n$  are such that  $p/n \rightarrow \ell \in ]0, \infty[$ .

Many other estimators of the covariance matrix have been proposed in the literature since the proposal of Ledoit and Wolf [19]. See for instance the book by Pourahmadi [28]. However, note that the purpose of this paper is not to estimate correctly the covariance matrix but rather to have an estimator close to the true one, that keeps at the same time the universal properties of self-normalized sums in a multivariate context. The goal is to obtain exponential inequalities for correctly standardized Hotelling’s  $T_n^2$  statistics (the quadratic version of self-normalized sums) under second-order moment assumptions in the symmetric case and very few assumptions in the general case (covering some heavy-tailed multivariate distributions). For reasons that will appear later, we are successful with the estimator proposed by Ledoit and Wolf [19]. Indeed with this estimator, we are able to keep the self-normalized structure of the sum after a correct rotation of the data. This is not the case for many other more complex estimators of the covariance matrix including the one developed in Ledoit and Wolf [20], Yao et al. [35] or Bodnar et al. [7] and most of the ones presented in Pourahmadi [28]. We can obtain some partial results (under some additional assumptions) for the estimators proposed in Bodnar et al. [7] which are generalizations of the estimator of Ledoit and Wolf [19]. Indeed we also obtain in the appendix some bounds for the Hotelling statistics in the symmetric case, when using the estimator proposed by Bodnar et al. [7]. It would be of great interest to extend our work under some sparsity assumptions since in that case it is known that concentration inequalities for some estimators of the variances may be obtained (see the forthcoming book of Tropp [34]) but this is another topic and it requires further developments.

The paper is divided into four parts including this introduction. In the second part, we recall some known exponential inequalities for  $T_n^2(0, 1)$  when  $p < n$  under weak moments assumptions. Then we obtain an oracle exponential inequality for the regularized Hotelling’s  $T_n^2(\rho_1, \rho_2)$ , assuming that the values  $\rho_1$  and  $\rho_2$  are fixed and known. Some interesting sharp bounds that may be useful in statistical learning assuming symmetry are obtained for any  $n$  and  $p$  large. We then establish a general inequality for  $p = O(n)$  for non-symmetric distributions under weak moment assumptions. In the third part, we estimate the optimal values  $\rho_1^*$  and  $\rho_2^*$  and show that the inequality remains valid up to some additional small terms controlling the concentration of these estimators around their true value. We illustrate our results with some simulations in the supplementary material.

## 2. Oracle exponential bounds for regularized Hotelling’s $T_n^2$

In the following, we define the penalized Hotelling’s  $T_n^2$  as the particular regularized Hotelling’s statistic  $T_n^2(\rho, 1)$  with  $\rho \geq 0$ . The aim of this section is to establish an oracle exponential inequality for the penalized Hotelling’s  $T_n^2$  in the case  $p \geq n$  and when the distribution of the data is symmetric (Theorem 1 and Theorem 2) as well as in the general case, that is when the distribution is not necessarily symmetric (Theorem 3).

### 2.1. Known bounds for Hotelling’s $T_n^2$

Some bounds for  $T_n^2$  or self-normalized sums may be quite easily obtained in the symmetric case (that is for random variables having a symmetric distribution see Pinelis [27]) and are well-known in the unidimensional case  $p = 1$ . In non-symmetric and/or multidimensional cases with  $p < n$ , these bounds are new and not trivial to prove. One of the main tools for obtaining exponential inequalities in various settings is the famous Hoeffding inequality (see Hoeffding [15]). For centered independent real random variables  $Y_1, \dots, Y_n$ , that are bounded, say  $|Y_i| < 1$ , for all  $i \in \{1, \dots, n\}$ , we have, for  $a_i \in [-1, 1]$  such that  $\sum a_i^2 = 1$ ,

$$\forall t > 0, \quad \mathbb{P} \left( \left( \sum_{i=1}^n a_i Y_i \right)^2 \geq t \right) \leq 2 \exp \left( -\frac{t}{2} \right).$$

A direct application of this inequality to self-normalized sums (via a randomization step introducing independent Rademacher r.v.’s  $\epsilon_i$ ) yields that, for independent real ( $p = 1$ ) symmetric random variables  $Z_i, i \in \{1, \dots, n\}$ , and not necessarily bounded (nor identically distributed). Indeed, we have by putting  $Y_i = \epsilon_i$  and  $a_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}}$ ; for  $t > 0$ ,

$$\mathbb{P} (T_n^2 \geq t) = \mathbb{P} \left( \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} \geq t \right) = \mathbb{P} \left( \frac{(\sum_{i=1}^n Z_i \epsilon_i)^2}{\sum_{i=1}^n Z_i^2} \geq t \right) = \mathbb{E} \left[ \mathbb{P} \left( \frac{(\sum_{i=1}^n Z_i \epsilon_i)^2}{\sum_{i=1}^n Z_i^2} \geq t \mid (Z_i)_{i \in \{1, \dots, n\}} \right) \right] \leq 2 \exp \left( -\frac{t}{2} \right).$$

Pinelis [27] has obtained with a different technique, a sharp  $\chi^2$  type of bounds which generalizes this kind of results for multivariate data when  $p < n$ . He proved that, if  $Z$  has a symmetric distribution, without any moment assumption on the variables  $Z_i$ , then one has

$$\forall t > 0, \quad \mathbb{P} (T_n^2 \geq t) \leq \frac{2e^3}{9} \overline{F}_p(t), \tag{1}$$

where  $F_p(t)$  is the cumulative distribution function (cdf) of a  $\chi^2(p)$  distribution. The density is denoted by  $f_p$ . A crude approximation yields that for  $t$  large enough,

$$\mathbb{P} (T_n^2 \geq t) \leq \frac{e^3}{9} \frac{2^{2-\frac{p}{2}}}{\Gamma(\frac{p}{2})} t^{\frac{p}{2}-1} \exp(-t/2),$$

where  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$  is the gamma function.

Notice that, for  $p = 1$  this bound (only valid for large  $t$ ) is better than the crude Hoeffding bound since we recover the missing factor  $\frac{1}{\sqrt{t}}$  in front of the exponential (see Talagrand [31]). When  $p < n$ , using a multidimensional version of Panchenko's symmetrization lemma (Panchenko [26]) Bertail et al. [6] have obtained an exponential inequality for the general distribution of  $Z$  with finite kurtosis  $\gamma_4 = \mathbb{E} \left( \left\| \Sigma_n^{-1/2} Z \right\|_2^4 \right)$ . More precisely, they establish that under  $0 < \gamma_4 < \infty$ ,

- (i) for  $t > np$ ,  $\mathbb{P}(T_n^2 \geq t) = 0$ ,
- (ii) for any  $a > 1$ , and any positive  $t$  such that  $2p(1+a) \leq t \leq np$ , the following bound holds for some explicit constant  $C(p)$ ,

$$\mathbb{P}(T_n^2 \geq t) \leq \frac{2e^3}{9\Gamma(\frac{p}{2} + 1)} \left( \frac{t - p(1+a)}{2(1+a)} \right)^{\frac{p}{2}} \exp \left( -\frac{t - p(1+a)}{2(1+a)} \right) + C(p)n^{3-\frac{6}{p+1}} \exp \left( -\frac{n \left(1 - \frac{1}{a}\right)^2}{\gamma_4(p+1)} \right).$$

The first term is essentially equivalent to the tail of a  $\chi^2(p)$  distribution (up to an explicit constant), while the second term controls the speed of convergence of  $S_n$  to  $\Sigma_n$ , when  $\gamma_4 < \infty$ . The constant  $a$  controls the balance between these two terms on the right-hand side of the inequality and may be optimized. Notice that this second exponential term is small when  $p \ll n$  but explodes in  $n^3$  if  $p/n \rightarrow l > 0$  for large  $n$ , making this bound useless in that case.

In the general multidimensional framework considered in Bertail et al. [6] and in this paper, the main difficulty is to keep the self-normalized structure when symmetrizing the initial sum. In the next sections, the results of Bertail et al. [6] obtained for  $p < n$  are extended to the case  $p \geq n$  by using a regularized version of  $S_n$  proposed in Ledoit and Wolf [19]. This inequality is based on an appropriate diagonalization of the regularized sample covariance matrix which allows applying Pinelis' inequality [27] (see Section 2.2). This crude inequality is refined in Section 2.3. When dealing with the general case (see Section 2.4), we establish first a multivariate symmetrization Lemma 3 in the spirit of Panchenko [26]. This symmetrization partially destroys the self-normalized structure (the normalization is then  $S_n(\rho_1, \rho_2) + \tilde{\Sigma}_n = 2S_n(\rho_1, \rho_2) + (\tilde{\Sigma}_n - S_n(\rho_1, \rho_2))$  instead of the expected normalization  $S_n(\rho_1, \rho_2)$ ), but the right standardization can be recovered (up to the factor 2) by obtaining a lower tail control of the distance between  $S_n(\rho_1, \rho_2)$  and  $\tilde{\Sigma}_n$ . To control this distance and make it as small as possible we will use the results of Ledoit and Wolf [19]. We also discuss what can be obtained with the estimator proposed by Bodnar et al. [7].

### 2.2. Bounds for regularized Hotelling's $T_n^2$ in a symmetric framework

We now obtain a simple inequality for the regularized Hotelling's  $T_n^2$  in the symmetric case, based on previous results by Pinelis [27]. It essentially shows that the tail of the regularized Hotelling's  $T_n^2$  is controlled by the tail of a  $\chi^2(n)$  distribution.

**Theorem 1.** Assume that  $Z$  has a symmetric distribution, then without any moment assumption, we have, for any  $n > 1$ , for  $t > n$ , for any  $\rho_1, \rho_2 > 0$ ,

$$\mathbb{P} \left( T_n^2 \left( \frac{\rho_1}{\rho_2}, 1 \right) \geq t \right) = \mathbb{P} \left( n \bar{Z}_n^T S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n \geq \frac{t}{\rho_2} \right) \leq \frac{2e^3}{9} F_n(t) \leq \frac{2e^3}{9} \exp \left( -\frac{(t-n)^2}{4t} \right), \tag{2}$$

where  $F_n$  is the cdf of a  $\chi^2(n)$  distribution.

Moreover, for any  $\rho > 0$ , we have

$$\mathbb{P} \left( \frac{T_n^2(\rho, 1) - n}{\sqrt{2n}} \geq t \right) = \mathbb{P} \left( \frac{n \bar{Z}_n^T S_n^{-1}(\rho, 1) \bar{Z}_n - n}{\sqrt{2n}} \geq t \right) \leq \frac{2e^3}{9} \exp \left( -\frac{t^2}{2 \left( 1 + \sqrt{2} \frac{t}{\sqrt{n}} \right)} \right). \tag{3}$$

Inequality (2) yields a control of  $T_n^2(\rho_1, \rho_2) = n \bar{Z}_n^T S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n$ , when using a linear shrinkage estimator of the variance. This, in turn, can be simplified in inequality (3), to a truly penalized Hotelling's  $T_n^2$ . Note that for any  $\rho_1, \rho_2 > 0$ ,

$$\frac{S_n(\rho_1, \rho_2)}{\rho_2} = \frac{\rho_2 S_n + \rho_1 I_p}{\rho_2} = S_n + \frac{\rho_1}{\rho_2} I_p$$

and for any  $\rho > 0$ ,

$$S_n(\rho, 1) = S_n + \rho I_p$$

is a penalized estimator of the covariance matrix. Inequality (3) can be interpreted as a Bernstein-type inequality.

**Remark 1.** These inequalities hold for any choice of positive  $\rho_1$  and  $\rho_2$ . However for the inequalities to be sharp,  $\rho_1$  and  $\rho_2$  should be chosen adequately. First from the proof of Theorem 1, we see that the inequality is sharp only when  $\rho_1$  is close to 0, which is in accordance with what we know about Tikhonov regularization [32]. Notice that the proof of Theorem 1 and the inequality remain valid if we use  $n \bar{Z}_n^T (S_n)^- \bar{Z}_n$  rather than  $n \bar{Z}_n^T S_n^{-2}(\rho, 1) \bar{Z}_n$ . In the procedure of Chen et al. [9] this means that asymptotically there is no difference between standardizing by the regularized variance or by the generalized inverse of the covariance matrix. Such a proposal using the Moore–Penrose inverse was also studied in Srivastava [29]. Actually when  $\rho$  tends to 0, then  $n \bar{Z}_n^T S_n^{-1}(\rho, 1) \bar{Z}_n$

converges to  $n\bar{Z}_n^\top(S_n)^-\bar{Z}_n$  where  $(A)^-$  is the Moore–Penrose or generalized inverse of  $A$  which is unique for symmetric matrices (see Theorem 4.3 in the survey of Barata et al. [5]). We give a short proof of this result in Section 6 of the supplementary material. We also refer to the book by Golub and Van Loan [14] for general results about the Moore–Penrose inverse and the link to least-square problems and Tikhonov regularization (see p. 257). As a consequence, the tail probabilities of  $n\bar{Z}_n^\top S_n^{-1}(\rho, 1)\bar{Z}_n$  and  $n\bar{Z}_n^\top(S_n)^-\bar{Z}_n$  are the same by a straightforward continuity argument and dominated convergence theorem. The regularization just serves as a trick to approximate the generalized inverse. However, the finite sample properties of the regularized Hotelling’s  $T_n^2$  will strongly depend on the choice of  $\rho_1$  and  $\rho_2$ .

**Remark 2.** Bodnar et al. [7] have proposed an improved estimator of the variance–covariance matrix in a large dimension setting by replacing the identity in Ledoit and Wolf’s shrinkage estimator by a general matrix, that is an estimator of the form:

$$\rho_1 \Sigma_0 + \rho_2 S_n,$$

for some matrix  $\Sigma_0$  such that for any  $p$  and  $n$ ,  $Tr(\Sigma_0) < \infty$ , where  $Tr(A)$  denotes the trace of the matrix  $A$ . They proved the strong consistency of this estimator for well-chosen values of  $\rho_1$  and  $\rho_2$  for several different matrix norms (including the Frobenius norm), under improved conditions. The purpose of this paper is not to have the best estimator of the true variance–covariance matrix but to use an estimator of this quantity, that keeps the nice self-normalized properties of the penalized Hotelling’s statistics. This is more restrictive and imposes some strong conditions on the type of regularization that can be chosen if one wants to obtain a sharp control. The main reason for this is the fact that eigenvalues are continuous functions of matrices while eigenvectors are not. When using a perturbation  $\Sigma_0$  of the empirical variance–covariance function  $S_n$  (centered at the true mean), the corresponding orthogonal transition matrices of this penalized matrix will not keep the self-normalized structure unless the perturbation matrix  $\Sigma_0$  and the empirical covariance matrix  $S_n$  are commuting which is true for a penalization matrix  $\Sigma_0$  proportional to the identity but more difficult to check for other cases. We give in the supplementary material (see subsection 6.5) an example of a random matrix  $\Sigma_0$  depending on  $S_n$ , which commutes with  $S_n$ . Notice that even if the perturbation matrix  $\Sigma_0$  is diagonal, in general,  $S_n$  and  $\Sigma_0$  will not commute. The particular case considered in Bodnar et al. [7] when  $\Sigma_0 = 1/pI_p$  with the corresponding optimal value given in their paper (see their expressions in Equations (3.8) and (3.9)) is covered by our results and Theorem 1 still holds in that case (by replacing  $\rho_1$  by the term  $\rho_1/p$  in the inequality).

In the following paragraph, we will obtain an improved bound in the symmetric case. Recall that, for symmetric matrices  $A$  and  $B$  of size  $p \times p$ , we have  $A \leq B$  if the matrix  $B - A$  is nonnegative. Note that, if we assume that there exists  $\eta_p > 0$  such that  $\eta_p I_p \leq \Sigma_0$ , then without any further assumptions, we can control the tail distribution of  $n\bar{Z}_n^\top (\rho_1 \Sigma_0 + \rho_2 S_n)^{-1} \bar{Z}_n$  by the tail of  $n\bar{Z}_n^\top (\eta_p \rho_1 I_p + \rho_2 S_n)^{-1} \bar{Z}_n$ . This amounts to replacing  $\rho_1$  by  $\eta_p \rho_1$  in the preceding inequalities. This will induce some additional cost in the inequalities of Theorems 3 and 4. Therefore, for the sake of simplicity, we will only present our results for  $S_n(\rho_1, \rho_2)$ .

Actually, we also obtain in Section 8 of the supplementary material a general result with the estimator proposed in Bodnar et al. [7] in the symmetric case (by working on a previous self-normalization of the data). The result is more difficult to state in comparison to the one obtained with Ledoit and Wolf’s estimator [19]. However, it may be of independent interest.

### 2.3. An improved bound for penalized Hotelling’s $T_n^2$ in the symmetric case

It can be seen from the proof of Theorem 1 that the penalized Hotelling’s  $T_n^2$  statistic essentially behaves like a weighted sum of independent  $\chi^2(1)$  random variables. This also explains the results of Chen et al. [9]. Actually, we can obtain a bound for this quantity relying on the results of Pinelis [27] and Laurent and Massart [18] (see p. 24 of their paper) who control the tail of the weighted sum of independent  $\chi^2(1)$  random variables. Let  $\lambda = (\lambda_j)_{j=1, \dots, p} \in \mathbb{R}_+^p$  be the eigenvalues of  $S_n$  (ordered in a decreasing order). We define for any  $\rho_1, \rho_2 > 0$ , the following effective dimensions (see Chen et al. [9] for other expressions of these quantities):

$$\Theta_1(\lambda, \rho_1, \rho_2) = \sum_{j=1}^{\inf(n,p)} \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j}; \quad \Theta_2(\lambda, \rho_1, \rho_2) = \sqrt{\sum_{j=1}^{\inf(n,p)} \frac{\lambda_j^2}{(\rho_1 + \rho_2 \lambda_j)^2}}; \quad \Theta_\infty(\lambda, \rho_1, \rho_2) = \sup_{1 \leq j \leq \inf(n,p)} \left( \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j} \right).$$

In the next result, we obtain a sharp bound for regularized and penalized Hotelling’s  $T_n^2$ . Notice that, in that case, the recentering factor depends on  $\Theta_1(\lambda, \rho_1, \rho_2)$  and is random. In the proof of Theorem 2, this value is essentially bounded by  $n/\rho_2$ , which is a very bad approximation when  $\rho_2$  is small. Theorem 2 tells that, for  $p \geq n$ , the tail of the regularized Hotelling’s  $T_n^2$  statistic behaves as the weighted sum of  $n$  independent  $\chi^2(1)$  r.v.’s where the weights are given by the random factors  $\frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j}$ . We get some Bernstein bounds for this weighted sum by first randomizing by some independent Gaussian r.v.’s, then conditioning on the data and applying Laurent and Massart’s Bernstein inequality [18]. This inequality in turn can be transformed into some exact bounds for the statistics of interest.

**Theorem 2.** Assume that  $Z$  has a symmetric distribution then, without any moment assumption, we have, for any  $n > 1$  and  $p > 0$ , for any  $t > 0$  and for any  $\rho_1, \rho_2 > 0$  including random values depending on the data we have,

$$\mathbb{P} \left( \frac{T_n^2(\rho_1, \rho_2) - \Theta_1(\lambda, \rho_1, \rho_2)}{\sqrt{2\Theta_2(\lambda, \rho_1, \rho_2)^2}} \geq \sqrt{2} \left( \sqrt{t} + \frac{\Theta_\infty(\lambda, \rho_1, \rho_2)}{\Theta_2(\lambda, \rho_1, \rho_2)} t \right) \right) \leq C \exp(-t), \quad C = \frac{3}{4} e^2.$$

Equivalently, we have for the penalized Hotelling’s statistic, for  $n > 1$  and  $p > 0$ , for any  $t > 0$  and, for any  $\rho > 0$ ,

$$\mathbb{P}\left(\frac{T_n^2(\rho, 1) - \Theta_1(\lambda, \rho, 1)}{\Theta_2(\lambda, \rho, 1)} \geq \sqrt{2t} + \frac{\Theta_\infty(\lambda, \rho, 1)}{\Theta_2(\lambda, \rho, 1)} t\right) \leq C \exp\left(-\frac{t}{2}\right).$$

**Remark 3.** (i) As mentioned before, it is possible to extend this theorem to the case when we use a standardization of the form  $S_n^A(\rho_1, \rho_2) = \rho_1 A + \rho_2 S_n$  where  $A$  is any symmetric (deterministic or not) matrix provided that  $S_n^A(\rho_1, \rho_2)$  is invertible. This covers both Ledoit and Wolf [19] and Bodnar et al. [7] estimators. However, in some cases, the eigenvalues of  $S_n^A$  may explode making the inequality rather useless. We give a general result (with the correct standardization) and discuss the choice for  $A$  in Section 8 of the supplementary material.

(ii) The constant  $C \approx 5.54$  comes from Remark 4.1 of Chasapis et al. [8] who extended a result of Pinelis [27]. Indeed they state that, when symmetrizing, for smooth functions of quadratic forms, Rademacher variables may be replaced by standard normal variables. However, their constant is probably not optimal and we expect the optimal value  $C$  to be  $2e^3/9 \approx 4.46$  as in Pinelis [27]. This is indeed supported by some simulations in the supplementary material.

The bounds in Theorem 2 can be used in practice for testing purposes in particular in anomaly detection in statistical learning. See for instance the literature on intrusion detection systems using multivariate control charts based on Hotelling’s  $T_n^2$  (for instance see Tracy et al. [33] and Ajadi et al. [2] for recent references in quality control). We develop this idea in the following example.

From Theorem 2, we can derive and construct an acceptance region for high-dimensional mean testing. Let  $W, W_1, \dots, W_n$  be i.i.d. random vectors of dimension  $p$ , orthon symmetric around their mean. We aim to test if the true mean vector of these observations is equal to  $\mu_0$  or not, for some fixed risk error  $\delta$ . For this we define  $\forall i \in \{1, \dots, n\} \quad Z_i = W_i - \mu_0$ . Now we want to test:

$$H_0 : \mathbb{E}(Z_i) = \mu = 0_p \quad \text{vs} \quad H_1 : \mu \neq 0_p$$

where  $0_p = (0, \dots, 0)^T$ .

First of all, compute the corresponding penalized Hotelling’s  $T_n^2$  (using any positive value for  $\rho_1$  and  $\rho_2$  including the optimal estimated values proposed in Ledoit and Wolf [19]) defined by

$$T_n^2(\rho_1, \rho_2) = n \bar{Z}_n^\top (S_n(\rho_1, \rho_2))^{-1} \bar{Z}_n$$

where  $S_n = \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top$  and  $S_n(\rho_1, \rho_2) = \rho_1 I_p + \rho_2 S_n$ . Now put  $c(\delta) = \log \frac{C}{\delta}$  with  $C = \frac{3}{4} e^2 \approx 5.54$ .

Then the acceptance region of level  $1 - \delta$  is given by

$$T_n^2(\rho_1, \rho_2) \leq \Theta_1(\lambda, \rho_1, \rho_2) + 2\Theta_2(\lambda, \rho_1, \rho_2) \left( \sqrt{c(\delta)} + \frac{\Theta_\infty(\lambda, \rho_1, \rho_2)}{\Theta_2(\lambda, \rho_1, \rho_2)} c(\delta) \right).$$

The proof of this result is left to the reader. This result holds for any  $n$  and  $p$ . When  $p \leq n$  is large, we can put  $\rho_1 = 0$  and get some Pinelis’ type bounds (when the  $\chi^2$  distribution tail is itself approximated by a Gaussian tail).

#### 2.4. Bounds for regularized Hotelling’s $T_n^2$ for non symmetric distribution

We now consider  $Z$  with a general (not necessarily symmetric) distribution. We will later prove a symmetrization lemma that generalizes the one obtained in Bertail et al. [6]. In the following, we use the results of Ledoit and Wolf [19] to optimally control the distance between  $S_n(\rho_1, \rho_2)$  and  $\Sigma_n$ . For this, consider the modified Frobenius scalar product between matrices and the corresponding norm given by

$$\langle A, B \rangle = \frac{\text{Tr}(AB^\top)}{p}, \quad \|A\|^2 = \langle A, A \rangle = \frac{\text{Tr}(AA^\top)}{p}.$$

Note that dividing the standard Frobenius scalar product by  $p$  enables the norm of the identity  $I_p$  to be equal to 1, which is more convenient. In the following, we extend this modified Frobenius norm to vectors by considering, for any vector  $Z \in \mathbb{R}^p$ ,

$$\|Z\|^2 = \text{Tr}(ZZ^\top) / p.$$

##### 2.4.1. Additional notations and hypotheses

Put  $\Sigma_n = (\sigma_{kj})_{1 \leq k, j \leq p}$  and consider  $\Lambda$  the diagonal matrix of the eigenvalues of  $\Sigma_n$  and  $O$  the matrix of associated eigenvectors. The eigenvalues are denoted  $\mu_1, \dots, \mu_p$  with  $\mu_{\max} = \mu_1 \geq \dots \geq \mu_p = \mu_{\min}$ . We have  $\Sigma_n = O\Lambda O^\top$ . Now, for  $i \in \{1, \dots, n\}$ , we define  $Y_i = O^\top Z_i$  with  $Y_i = (Y_{i,1}, \dots, Y_{i,p})^\top$ .

In order to provide a well-conditioned estimator for large dimensional covariance matrices, Ledoit and Wolf (2004) [19] have studied the minimum of  $\mathbb{E}\left(\|S_n(\rho_1, \rho_2) - \Sigma_n\|^2\right)$ . This minimization can be seen as a projection problem in the Hilbert space of random matrices, equipped with the inner product  $\langle A, B \rangle_H = \mathbb{E}[\langle A, B \rangle]$  with associated norm  $\|\cdot\|_H^2 = \mathbb{E}\|\cdot\|^2$ . We assume the four following assumptions:

- $A_1 \exists K_0, K_1 > 0$  such that, for any  $n$  and any  $p \geq n$ ,  $K_0 \leq \frac{p}{n} \leq K_1$ .
- $A_2 \exists K_2 > 0$  such that, for any  $n$  and any  $p \geq n$ ,  $\frac{1}{p} \sum_{j=1}^p \mathbb{E}\left[Y_{1,j}^8\right] \leq K_2$ .

$A_3 \exists K_3 > 0$  such that for any  $n$  and any  $p \geq n$ ,  $\frac{1}{K_3} < \mu_{\min} \leq \mu_{\max} < K_3$ .

$A_4 \exists K_4 > 0$  such that for any  $n$  and any  $p \geq n$ ,

$$v = \frac{p^2}{n^2} \times \frac{\sum_{(i,j,k,l) \in Q} (\text{Cov}(Y_{1,i}Y_{1,j}, Y_{1,k}Y_{1,l}))^2}{\text{Card}(Q)} \leq \frac{K_4}{n},$$

where  $Q$  denotes the set of all the quadruples that are made of four distinct integers between 1 and  $p$ .

**Remark 4.**  $A_2$  and  $A_4$  are already assumed in Ledoit and Wolf [19]. First assumption  $A_1$  essentially means that  $p = p(n)$  is of the same order as  $n$ .  $A_2$  states that the moment of order 8 is bounded in average: this condition holds if the following moment of order 8,  $\frac{1}{p} \sum_{j=1}^p \mathbb{E} \left[ Z_{1,j}^8 \right]$  is finite (by sub-multiplicative inequality and the fact that  $\|O\| = 1$ ). Notice that we need this condition to have a rate of convergence (and control the variance of the estimator) whereas this condition is used to obtain consistency in Ledoit and Wolf [19]. As noticed in Bodnar et al. [7], it is a strong hypothesis for consistency but we cannot avoid it to obtain a rate. This is still a weak condition since we do not require exponential moments and allow for a fat tail behavior of the sample.  $A_3$  ensures that the largest and the smallest eigenvalue of the true covariance matrix are bounded. This rules out the case when the components of the vector  $Z$  are too correlated: consider for instance the degenerate case where  $\Sigma_n$  is a matrix full of 1, then in that case the smallest eigenvalue is 0 and the largest is  $p$ . The case of a vector with long memory components is studied in Merlevède et al. [23]: they show that the largest eigenvalue is unbounded. Thus, this case does not enter our framework. Assumption  $A_4$  is immediate in the Gaussian case, since  $v = 0$  because of the rotation which makes the  $Y_{1,j}$ 's,  $j \in \{1, \dots, p\}$ , independent. Obviously, for  $(Z_{1,j})_j$  independent,  $v = 0$  as well. More generally if the components of the vector satisfy some adequate  $\alpha$ -mixing conditions, then the sum in the hypothesis  $A_4$  can be seen as a sum of cumulants and may also be controlled using the arguments of Doukhan and León [12].

2.4.2. Inequalities for random variables with a general distribution

The next Theorem 3 extends Theorem 1 to general distributions which are not necessarily symmetric. From now on, following Ledoit and Wolf [19], we denote  $\rho_1^*$  and  $\rho_2^*$  the optimal values defined as the minimum arguments of  $\mathbb{E} \left\| S_n(\rho_1, \rho_2) - \Sigma_n \right\|^2$ . Ledoit and Wolf [19] have obtained

$$\rho_1^* = \frac{\beta^2}{\delta^2} \sigma^2, \quad \rho_2^* = \frac{\alpha^2}{\delta^2}, \quad \sigma^2 = \langle \Sigma_n, I_p \rangle, \quad \alpha^2 = \left\| \Sigma_n - \sigma^2 I_p \right\|^2, \\ \beta^2 = \mathbb{E} \left\| S_n - \Sigma_n \right\|^2, \quad \delta^2 = \alpha^2 + \beta^2 = \mathbb{E} \left\| S_n - \sigma^2 I_p \right\|^2.$$

Now, we define, for  $\alpha^2 \neq 0$ ,  $\rho^* = \frac{\rho_1^*}{\rho_2^*} = \frac{\beta^2}{\alpha^2} \sigma^2$ , which yields the optimal penalized estimator of  $S_n$ :

$$S_n^* = \frac{S_n(\rho_1^*, \rho_2^*)}{\rho_2^*} = S_n + \rho^* I_p.$$

If  $\alpha^2 = 0$ , take  $S_n^* = \sigma^2 I_p$  (in that case we will need to estimate  $\sigma^2$ ).

**Theorem 3.** Assume that  $Z$  has a general distribution with finite variance  $\Sigma_n$ . Assume in addition that assumptions  $A_1$  to  $A_3$  hold. Put  $a^* = 1 + \frac{K_3}{\rho^*}$ . Then we have, for any  $n > 1$ , for any  $p \geq n$ , and for  $t > 2n$ ,

$$\mathbb{P} \left( T_n^2(\rho^*, 1) \geq (1 + a^*)t \right) \leq \frac{2e^3}{9} \left( \frac{t-n}{2} \right)^{\frac{n}{2}} \frac{\exp\left(-\frac{t-n}{2}\right)}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

**Remark 5.** Here the bounding function for large  $t$  behaves like a centered  $\chi^2(n)$  distribution, up to the factor  $\frac{2e^3}{9}$ . The term  $(1 + a^*)$  ensures that the maximal eigenvalue of  $\Sigma_n$  does not contribute to the inequality.

Notice that the inequality is still valid when using  $S_n(\rho_1, \rho_2)$ , the regularized version of  $S_n$  instead of the penalized version  $S_n^*$ , up to a small modification of the bound  $(1 + a^*)t$  by the factor  $1/\rho_2^*$ : for  $n > 1$ ,  $p \geq n$ , for any  $t > 2n$

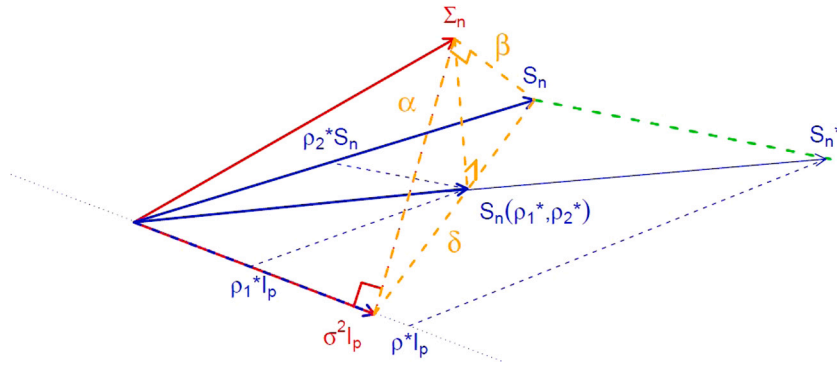
$$\mathbb{P} \left( T_n^2(\rho_1^*, \rho_2^*) \geq \frac{1}{\rho_2^*} (1 + a^*)t \right) \leq \frac{2e^3}{9} \left( \frac{t-n}{2} \right)^{\frac{n}{2}} \frac{\exp\left(-\frac{t-n}{2}\right)}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

Notice that the green dashed line in Fig. 1 represents the set of penalized estimators  $S_n(\rho, 1)$  for which we obtain universal bounds in Theorems 1 and 2.

3. Inequality with estimated parameters

We have proved an exponential inequality for the penalized Hotelling's  $T_n^2$  with theoretical values  $a^*$  and  $\rho^*$ . In practice these values are unknown. In this section, we estimate these quantities and obtain an inequality for the penalized Hotelling's  $T_n^2$  with estimated parameters.





**Fig. 1.** True covariance  $\Sigma_n$ , sample covariance  $S_n$ , and  $S_n(\rho_1^*, \rho_2^*), S_n^*$  respectively regularized and penalized sample covariance. The scalar product is  $\langle \cdot, \cdot \rangle_H$  with its associated norm. We represent  $S_n(\rho_1^*, \rho_2^*)$ , the optimal combination of  $S_n$  and  $I_p$  defined by orthogonal projection of the true covariance matrix  $\Sigma_n$  on the random vector-space generated by  $S_n$  and  $I_p$ . Thus  $S_n^* = S_n(\rho^*, 1)$  is the penalization of  $S_n$  by  $I_p$  with  $\rho^* = \frac{\hat{\alpha}_n^2}{\hat{\alpha}_n^2 + \hat{\beta}_n^2}$  for which we obtain a bound in [Theorem 3](#).

We first recall several results of Ledoit and Wolf [19] on the asymptotic behavior of regularized empirical covariance estimator  $S_n(\rho_1, \rho_2)$ . [Lemma 1](#) and [Proposition 1](#) below summarize these results with our notations and are proved by Ledoit and Wolf [19] in different lemmas and a theorem of their paper.

We use the same assumptions as in Ledoit and Wolf [19]:  $\xrightarrow{L_4}$  denotes the fourth-moment convergence as  $n$  goes to infinity, i.e.

$$U_n \xrightarrow{L_4} U \iff \mathbb{E} \left[ (U_n - U)^4 \right] \xrightarrow{n \rightarrow \infty} 0.$$

Ledoit and Wolf [19] essentially have shown that  $L_4$ -consistent estimators for  $\sigma^2, \alpha^2, \beta^2$  and  $\delta^2$  are simply given by their empirical counterparts that is

$$\hat{\sigma}_n^2 = \langle S_n, I_p \rangle, \quad \hat{\delta}_n^2 = \|S_n - \hat{\sigma}_n^2 I_p\|^2, \quad \hat{\alpha}_n^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2, \quad \hat{\beta}_n^2 = \frac{1}{n^2} \sum_{i=1}^n \|Z_i(Z_i)^\top - S_n\|^2, \quad \hat{\beta}_n^2 = \min(\hat{\beta}_n^2, \hat{\delta}_n^2).$$

**Lemma 1** (Ledoit and Wolf [19] Lemma 3.2, Lemma 3.3, Lemma 3.4, Lemma 3.5). *Under Assumptions  $A_1$  to  $A_4$ , we have*

- (i)  $\sigma^2, \alpha^2$  and  $\beta^2$  remain bounded (as  $n$  and  $p$  tend to  $\infty$ ).
- (ii) For all  $n, \mathbb{E}[\hat{\sigma}_n^2] = \sigma^2, \hat{\sigma}_n^2 - \sigma^2 \xrightarrow{L_4} 0, \hat{\sigma}_n^4 - \sigma^4 \xrightarrow{L_4} 0.$
- (iii)  $\hat{\delta}_n^2 - \delta^2 \xrightarrow{L_4} 0.$
- (iv)  $\hat{\beta}_n^2 - \beta^2 \xrightarrow{L_4} 0$  and  $\hat{\beta}_n^2 - \beta^2 \xrightarrow{L_4} 0.$
- (v)  $\hat{\alpha}_n^2 - \alpha^2 \xrightarrow{L_4} 0.$

After replacing the unobservable scalars  $\sigma^2, \alpha^2, \beta^2$  and  $\delta^2$  by their sample counterparts in the formula of  $S_n(\rho_1, \rho_2)$ , Ledoit and Wolf [19] obtained an estimation of the regularized empirical covariance matrix say

$$\hat{S}_n = \frac{\hat{\beta}_n^2}{\hat{\delta}_n^2} \hat{\sigma}_n^2 I_p + \frac{\hat{\alpha}_n^2}{\hat{\delta}_n^2} S_n.$$

Ledoit and Wolf [19] have shown that  $\hat{S}_n$  and  $S_n(\rho_1, \rho_2)$  are asymptotically equivalent in the modified Frobenius norm.

**Proposition 1** (Ledoit and Wolf [19], Theorem 3.2). *Under Assumptions  $A_1$ – $A_4$ , we have*

- (i)  $\lim_{n \rightarrow \infty} \mathbb{E} \left\| \hat{S}_n - S_n(\rho_1, \rho_2) \right\|^2 = 0.$
- (ii) Moreover,  $\hat{S}_n$  has the same asymptotic expected loss (or risk) as  $S_n(\rho_1, \rho_2)$ , i.e.,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\| \hat{S}_n - \Sigma_n \right\|^2 - \mathbb{E} \left\| S_n(\rho_1, \rho_2) - \Sigma_n \right\|^2 = 0.$$

In the same way as Ledoit and Wolf [19] we define the optimal coefficients  $\rho_1^*$  and  $\rho_2^*$ . They are estimated respectively by  $\hat{\rho}_1^*$  and  $\hat{\rho}_2^*$ , where  $\hat{\rho}_1^* = \frac{\hat{\beta}_n^2}{\hat{\delta}_n^2} \hat{\sigma}_n^2$  and  $\hat{\rho}_2^* = \frac{\hat{\alpha}_n^2}{\hat{\delta}_n^2}$ . Now, if  $\hat{\alpha}_n^2 \neq 0$ , we introduce  $\hat{S}_n^*$  the “estimated optimal” penalized version of  $S_n$  given by

$$\hat{S}_n^* = S_n \left( \frac{\hat{\rho}_1^*}{\hat{\rho}_2^*}, 1 \right) = S_n + \hat{\rho}_n^* I_p, \quad \hat{\rho}_n^* = \frac{\hat{\beta}_n^2 \hat{\sigma}_n^2}{\hat{\alpha}_n^2}.$$

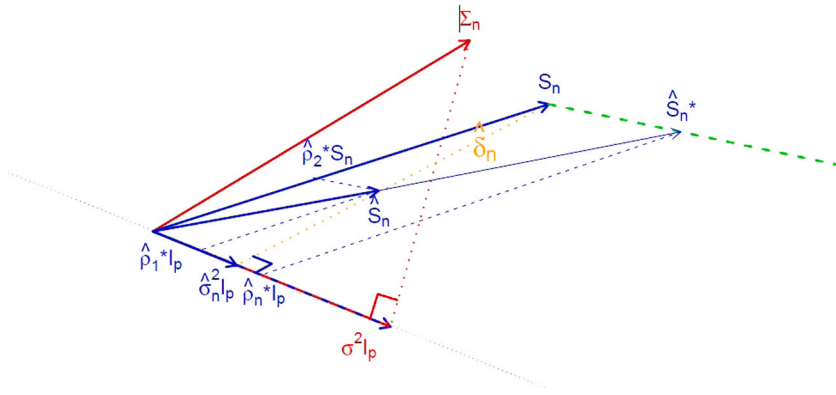


Fig. 2. We denote the true covariance  $\Sigma_n$ , the sample covariance  $S_n$ , the regularized and the penalized estimators of  $S_n$ , respectively  $\hat{S}_n$  and  $\hat{S}_n^*$ . The principle is similar to Fig. 1 except that  $\hat{S}_n$  is determined first so that the regularized estimator belongs to the yellow line and the penalized optimal estimator  $\hat{S}_n^* = S_n \left( \frac{\hat{\rho}_n^*}{\rho_n^*}, 1 \right)$  is the closest value to  $\Sigma_n$  on this line. This difference induces an additional error term in our inequalities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Similarly the unobservable threshold constant  $a^*$  introduced in Theorem 3 is estimated by  $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$ . Theorem 4 establishes an exponential bound for the penalized self-normalized sums, when  $S_n^*$  is replaced by the estimator  $\hat{S}_n^*$  and  $a^*$  by  $\hat{a}_n^*$ , up to a small error term that we control explicitly (see Fig. 2).

**Theorem 4.** Under Assumptions  $A_1$  to  $A_4$ , we have, for any  $n > 1$ , for any  $p > n$ , for any  $t > 2n$  and for any small value of  $\epsilon > 0$ ,

$$\mathbb{P} \left( T_n^2(\hat{\rho}_n^*, 1) \geq t \left( 1 + \hat{a}_n^* + 2\epsilon \right) \right) \leq \frac{2e^3}{9} \left( \frac{t-n}{2} \right)^{\frac{n}{2}} \frac{e^{-\frac{t-n}{2}}}{\Gamma \left( \frac{n}{2} + 1 \right)} + \frac{C(\epsilon)}{n\epsilon}, \tag{4}$$

where  $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$ , and  $C(\epsilon)$  is a real positive function, independent of  $n$ , defined by

$$C(\epsilon) = 4K_1 \sqrt{K_2} \left( 2 + \frac{1}{p} + K_1 \right) + 2K_1 G \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + \frac{4K_1^2 \sigma^4}{\epsilon} G \left( \frac{\epsilon}{2\sigma^2 K_1} \right) + \frac{K_3}{\epsilon} G \left( \frac{\epsilon}{K_3} \right).$$

The function  $G$  is defined explicitly in Lemma 7. Notice that  $C(\epsilon)/\epsilon$  explodes when  $\epsilon$  goes to 0.

These results essentially show that we have a  $\chi^2(n)$  control in the tail of the distribution, for a threshold larger than  $2n(1 + \hat{a}_n^*)$  (recall that  $2n$  is the variance of a  $\chi^2(n)$  distribution). The loss  $(1 + \hat{a}_n^*)$  is essentially due to the correlation between the components of  $Z$  and the deviation from homoscedasticity. The value of  $\epsilon$  cannot be too small but can be optimized by balancing the two terms in the inequality. For a given  $\epsilon$  and a given level  $\delta$  it is possible to solve numerically the second term of inequality (4) equal to  $\delta$  to get a valid bound for the Hotelling’s  $T_n^2$  for any  $n$  and  $p$ .

#### 4. Simulations and conclusion

We defer the intensive simulation study to Section 2 of the supplementary material. For different multidimensional distributions (respectively, Gaussian, Log-Normal, and multivariate Pareto type II distributions, with different dependency structures, see details in the supplementary material) we give the distribution of the estimated optimal penalty as well as the distribution of the penalized Hotelling  $T^2$  (with estimated penalty) compared to our exponential bounds. From these simulations, we draw the following conclusions:

- (i) When the components of the vector are independent, the penalty coefficient  $\hat{\rho}_n^*$  can be very large, especially for very asymmetric and fat tail distributions (see Remarks 1 and 2 of the supplementary material). It means that the covariance matrix does not play any role in the standardization. Conversely, when the data coordinates are dependent, the penalty coefficient  $\hat{\rho}_n^*$  tends to become relatively small. This is in accordance with the simulation results obtained by Dong et al. [11], in a slightly different framework (comparison of two samples) showing that it may be useless to use the covariance matrix component in the independent case.
- (ii) The values of  $\hat{\rho}_n^*$  depend on the level of dependence but also on the dispersion of eigenvalues in the case of independence. Specifically, if  $\alpha^2$  is small (indicating low dispersion of eigenvalues), the observed mean value of  $\hat{\rho}_n^*$  is smaller compared to the case with a large  $\alpha^2$  (indicating high dispersion of eigenvalues). This phenomenon is more pronounced for asymmetric and for heavy-tailed distributions. In that case, it is clear that the estimator of the covariance matrix is closer to the identity (up to a scaling factor) than to the empirical covariance matrix (see also the conclusions of Dong et al. [11]).
- (iii) On the contrary, in the case of dependence, the estimated value  $\hat{\rho}_n^*$  remains relatively small (does not explode as in independent scenarios). As the level of dependence increases, the  $\hat{\rho}_n^*$  value decreases.

(iv) For the normal distribution, a strong linear relationship is observed between the optimal estimated penalty coefficient  $\hat{\rho}_n^*$  and the ratio  $p/n$ . It seems that the penalty should be of order  $2p/n$  in the independent case but more of order  $p/n$  in the dependent case (for our specific values of the parameters).

(v) The distribution of the penalized Hotelling’s  $T_n^2$  statistics clearly depends on the ratio  $p/n$  but exhibits a stable behavior even when the penalty coefficient becomes exceedingly large.

Since our bounds are universally applicable and valid regardless of the data distribution, they inherently tend to be conservative (especially in the Gaussian case) but can be used safely when the asymptotic normal approximation is quite inappropriate (see the log-normal and Pareto cases in particular). From this simulation study, we conclude that our work gives some interesting information both on the optimal penalty that one may choose and on the order of the bounds. However, there is still room for improvement.

**Acknowledgments**

The authors sincerely thank Prof. Dietrich von Rosen, the Editor-in-Chief, the Associate Editor, and the anonymous reviewers for their valuable feedback and constructive comments and suggestions, which have greatly contributed to enhancing the quality and scope of the results. We would also like to thank the algebraist Xavier Mary, MODAL’X, Univ. Paris Nanterre, for his comments and his suggestions. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

**Appendix A. Proofs**

In the first part of this section, we provide all the proofs of [Theorems 1, 2, 3](#) and [4](#) given in Sections 2, 3 and 4. In the second part of the appendix, we detail all the calculations to obtain an explicit constant  $C(\epsilon)$  appearing in [Theorem 4](#) when replacing the true quantities by their empirical estimators.

We set some notations that we will consider in the following proofs.  $S_n$  is a symmetric and nonnegative matrix. Denote by  $O_n$  an orthogonal matrix in  $\mathcal{M}_p(\mathbb{R})$  such that  $S_n = O_n \Lambda_n O_n^T$  where  $\Lambda_n$  is a diagonal matrix and

$$A_n = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_n & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \text{ for any } p > n.$$

Let  $\lambda_1 \geq \dots \geq \lambda_p$  denote eigenvalues of  $S_n$ . Put  $\hat{Y}_i = O_n^T Z_i$  with  $\hat{Y}_i = (\hat{Y}_{i,1}, \dots, \hat{Y}_{i,p})^T$ .

**Proof of Theorems 1 and 2.** We first establish a simple inequality for the penalized Hotelling’s  $T_n^2$  in the symmetric case, based on previous results by Pinelis [27]. The idea of the theorem is to use a rotation trick of the  $Z_i$  that allows us to return to the “small” dimension case given by Pinelis [27]. This yields a bound given by the survival function of a  $\chi^2$  with  $n$  degrees of freedom.  $\square$

**Proof of Theorem 1.** Note that Vectors  $\hat{Y}_i$  remain symmetric in distribution and uncorrelated. It is easy to see that, by construction, the empirical covariance matrix of the  $\hat{Y}_1, \dots, \hat{Y}_n$  is

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_i \hat{Y}_i^T = \frac{1}{n} \sum_{i=1}^n O_n^T Z_i Z_i^T O_n = O_n^T S_n O_n = \Lambda_n.$$

This implies that for any vector  $\hat{Y}_i$ , their coordinates for  $j \geq n + 1$  are zero. Indeed, for  $j \geq n + 1$ ,  $n^{-1} \sum_{i=1}^n \hat{Y}_{i,j}^2 = 0$ , implies in turn that each  $\hat{Y}_{i,j} = 0$ ,  $j \in \{n + 1, \dots, p\}$ ,  $i \in \{1, \dots, n\}$ . Define  $\tilde{Y}_i$  the  $n$ -dimensional vector version of  $\hat{Y}_i$  with these non-zero components, that is to say  $\forall j \leq n$ ,  $\tilde{Y}_{i,j} = \hat{Y}_{i,j}$  and their corresponding empirical mean  $\tilde{Y}_n$  on the collection  $\tilde{Y}^{(n)} = (\tilde{Y}_i)_{1 \leq i \leq n}$ . Thus, for all  $\rho_1, \rho_2 > 0$ , we have

$$\begin{aligned} n \tilde{Z}_n^T S_n^{-1}(\rho_1, \rho_2) \tilde{Z}_n &= n \left( \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i^T \right) (\rho_1 I_p + \rho_2 \Lambda_n)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i \right) \\ &= n \sum_{j=1}^n \frac{(n^{-1} \sum_{i=1}^n \tilde{Y}_{i,j})^2}{\rho_1 + \rho_2 \lambda_j} \leq n \sum_{j=1}^n \frac{(n^{-1} \sum_{i=1}^n \tilde{Y}_{i,j})^2}{\rho_2 \lambda_j} \leq \frac{1}{\rho_2} \sum_{j=1}^n \frac{(n^{-1/2} \sum_{i=1}^n \tilde{Y}_{i,j})^2}{\lambda_j}. \end{aligned}$$

As  $\lambda_j = n^{-1} \sum_{i=1}^n \hat{Y}_{i,j}^2$ , we have reduced the problem to the sum of  $n$  self normalized sums, which can be seen as Hotelling’s  $T_n^2$  of symmetric random variables in  $\mathbb{R}^n$ . In other words,  $n \tilde{Z}_n^T S_n^{-1}(\rho_1, \rho_2) \tilde{Z}_n \leq \frac{1}{\rho_2} n \tilde{Y}_n^T S_n^{-1}(\tilde{Y}^{(n)}) \tilde{Y}_n$ . Thus, by applying Pinelis [27] (1), we have

$$\forall t > 0, \quad \mathbb{P} \left( n \tilde{Z}_n^T S_n^{-1}(\rho_1, \rho_2) \tilde{Z}_n \geq \frac{t}{\rho_2} \right) \leq \frac{2e^3}{9} \bar{F}_n(t).$$

Recall that, if  $N_1, \dots, N_n$  are independent  $\mathcal{N}(0, 1)$  r.v.s, then by Lemma 1 of Laurent and Massart [18], one has,

$$\forall u > 0, \quad \mathbb{P} \left( \frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \geq \sqrt{2}(\sqrt{u} + \frac{u}{\sqrt{n}}) \right) \leq e^{-u}.$$

By inverting the polynomial in  $\sqrt{u}$ , this is a Bernstein type inequality for i.i.d random variables

$$\forall v > 0, \quad \mathbb{P} \left( \frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \geq v \right) \leq \exp \left[ - \frac{2v^2}{\left(1 + \sqrt{1 + 2\sqrt{2}\frac{v}{\sqrt{n}}}\right)^2} \right] \leq \exp \left[ - \frac{v^2}{2(1 + \sqrt{2}\frac{v}{\sqrt{n}})} \right].$$

It follows that, for  $t > n$ ,

$$\bar{F}_n(t) = \mathbb{P} \left( \frac{\sum_{i=1}^n N_i^2 - n}{\sqrt{2n}} \geq \frac{t - n}{\sqrt{2n}} \right) \leq \exp \left( - \frac{(t - n)^2}{4t} \right). \quad \square$$

**Proof of Theorem 2.** Recall that:  $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$  with  $Z_i \in \mathbb{R}^p$ . Introduce independent Rademacher r.v.'s  $\epsilon_i$  taking the values  $\pm 1$  with probability  $1/2$ . Define  $\bar{Z}_n^\epsilon = \frac{1}{n} \sum_{i=1}^n \epsilon_i Z_i$ . Then, in the symmetric case considered here,  $\bar{Z}_n$  and  $\bar{Z}_n^\epsilon$  have the same distribution. Now write

$$n(\bar{Z}_n^\epsilon)^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n^\epsilon = n \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i \hat{Y}_i^\top \right) (\rho_1 I_p + \rho_2 \Lambda_n)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i \hat{Y}_i \right) = \epsilon^\top V V^\top \epsilon, \tag{5}$$

where  $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)^\top$ ,  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\top$ , and  $V = \frac{1}{\sqrt{n}} \hat{Y} (\rho_1 I_p + \rho_2 \Lambda_n)^{-1/2}$ .

Chasapis et al. [8] obtain an extension of Pinelis [27] result, stating that for smooth functions of quadratic forms, Rademacher variables may be replaced by standard normal variables. More precisely, define the Euclidian norm  $\|x\|_2 = \sqrt{\langle x, x \rangle}$  and consider  $\xi_1, \dots, \xi_n$  independent standard Gaussian random variables. Then, for any  $t \geq 0$ , for any vectors  $v_1, \dots, v_n$  in  $\mathbb{R}^p$ , we have

$$\mathbb{P} \left[ \|\epsilon_1 v_1 + \dots + \epsilon_n v_n\|_2 \geq t \right] \leq C \mathbb{P} \left[ \|\xi_1 v_1 + \dots + \xi_n v_n\|_2 \geq t \right], \quad C = \frac{3}{4} e^2.$$

Note that the value  $C = \frac{3}{4} e^2 \approx 5.54$  (see Chasapis et al. [8]) is probably not optimal. See the supplementary material for numerical experiments supporting the fact that the optimal value is the one appearing in Pinelis equal to  $2e^3/9$  for smooth function. Since we have

$$\epsilon_1 v_1 + \dots + \epsilon_n v_n = \epsilon^\top V$$

where  $V$  is the matrix of vectors  $v_i = (v_{i1}, \dots, v_{ip})$  corresponding to the rows, we can rewrite

$$\|\epsilon_1 v_1 + \dots + \epsilon_n v_n\|_2^2 = \|\epsilon^\top V\|_2^2 = \epsilon^\top V V^\top \epsilon.$$

It follows that, for any  $u > 0$ ,

$$\mathbb{P} \left[ \epsilon^\top V V^\top \epsilon \geq u \right] \leq C \mathbb{P} \left[ \xi^\top V V^\top \xi \geq u \right]$$

By conditioning according to  $\hat{Y}_i$ 's and using Eq. (5), we have, for any  $n > 1$ , for any  $p > 1$ , for any  $u > 0$  and, for any  $\rho_1, \rho_2 > 0$ ,

$$\mathbb{P} \left[ n(\bar{Z}_n^\epsilon)^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n^\epsilon \geq u \right] = \mathbb{E} \left[ \mathbb{P} \left( \epsilon^\top V V^\top \epsilon \geq u \mid \hat{Y}_1, \dots, \hat{Y}_n \right) \right] \leq C \mathbb{E} \left[ \mathbb{P} \left( \xi^\top V V^\top \xi \geq u \mid \hat{Y}_1, \dots, \hat{Y}_n \right) \right].$$

Moreover, recall from the preceding proof that we have

$$\begin{aligned} n(\bar{Z}_n^\epsilon)^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n^\epsilon &= n \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i \hat{Y}_i^\top \right) (\rho_1 I_p + \rho_2 \Lambda_n)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i \hat{Y}_i \right) = n \sum_{j=1}^{\inf(p,n)} \frac{(n^{-1} \sum_{i=1}^n \epsilon_i \hat{Y}_{i,j})^2}{\rho_1 + \rho_2 \lambda_j} \\ &= n \sum_{j=1}^{\inf(p,n)} \frac{(n^{-1} \sum_{i=1}^n \epsilon_i \hat{Y}_{i,j})^2}{\lambda_j} \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j} \end{aligned}$$

We obtain

$$\mathbb{P} \left[ n(\bar{Z}_n^\epsilon)^\top (S_n(\rho_1, \rho_2))^{-1} \bar{Z}_n^\epsilon > u \right] \leq C \mathbb{E} \left[ \mathbb{P} \left[ n \sum_{j=1}^{\inf(p,n)} \frac{(n^{-1} \sum_{i=1}^n \xi_i \hat{Y}_{i,j})^2}{\lambda_j} \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j} > u \mid \hat{Y}_1, \dots, \hat{Y}_n \right) \right]. \tag{6}$$

Let us now work conditionally to  $\hat{Y}_1, \dots, \hat{Y}_n$ . Put  $K_j = \sqrt{n} (n^{-1} \sum_{i=1}^n \xi_i \hat{Y}_{i,j}) / \sqrt{\lambda_j}$ ,  $j \in \{1, \dots, \inf(p,n)\}$ . Thus for any  $j \neq k$

$$\text{Cov} \left( K_j, K_k \mid \hat{Y}_1, \dots, \hat{Y}_n \right) = \text{Cov} \left( \sqrt{n} \frac{n^{-1} \sum_{i=1}^n \xi_i \hat{Y}_{i,j}}{\sqrt{\lambda_j}}, \sqrt{n} \frac{n^{-1} \sum_{i=1}^n \xi_i \hat{Y}_{i,k}}{\sqrt{\lambda_k}} \mid \hat{Y}_1, \dots, \hat{Y}_n \right) = \frac{1}{n} \sum_{i=1}^n \frac{\hat{Y}_{i,j} \hat{Y}_{i,k}}{\sqrt{\lambda_j \lambda_k}} = 0.$$

Since  $K = (K_1, \dots, K_{\inf(p,n)})$  is a Gaussian vector (as a linear combination of independent variables), it follows that  $K_1^2, \dots, K_{\inf(p,n)}^2$  are iid  $\chi^2(1)$ .

Now, consider the vector  $b = (b_1, \dots, b_p)$  with nonnegative components (conditionally to  $\hat{Y}_{i,j}$ 's) defined by

$$b_j = \frac{\lambda_j}{\rho_1 + \rho_2 \lambda_j}.$$

A direct application of Laurent and Massart [18] Lemma 1 to  $\sum_{j=1}^{\inf(p,n)} b_j (K_j^2 - 1)$  gives for any  $u > 0$

$$\mathbb{P} \left( \sum_{j=1}^{\inf(p,n)} b_j (K_j^2 - 1) > 2\|b\|_2 \sqrt{u} + 2\|b\|_\infty u \right) \leq \exp(-u).$$

In other words, for any  $u > 0$ , we have

$$\mathbb{P} \left( \frac{\sum_{j=1}^{\inf(p,n)} b_j K_j^2 - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2} \frac{\|b\|_\infty}{\|b\|_2} u \right) \leq \exp(-u). \tag{7}$$

Now by combining (6) and (7) we obtain the following result for the recentered version of our quantity of interest,

$$\begin{aligned} & \mathbb{P} \left( \frac{n(\bar{Z}_n^\xi)^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n^\xi - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2} \frac{\|b\|_\infty}{\|b\|_2} u \right) \\ & \leq C \mathbb{E} \left[ \mathbb{P} \left( \frac{\sum_{j=1}^{\inf(p,n)} b_j K_j^2 - \|b\|_1}{\sqrt{2\|b\|_2^2}} > \sqrt{2}\sqrt{u} + \sqrt{2} \frac{\|b\|_\infty}{\|b\|_2} u \mid (\hat{Y}_{1,j}, \dots, \hat{Y}_{n,j})_{j \in \{1, \dots, \inf(p,n)\}} \right) \right] \leq C \exp(-u). \end{aligned}$$

The result of the theorem follows by noticing that  $\|b\|_k = \Theta_k(\lambda, \rho_1, \rho_2)$ ,  $k \in \{1, 2\}$  and  $\|b\|_\infty = \Theta_\infty(\lambda, \rho_1, \rho_2)$ .  $\square$

**Proof of Theorem 3.** In the following, we adapt Panchenko’s symmetrization lemma to a  $\chi^2$  distribution. This ensures that, if we have a  $\chi^2(k)$  type of control for the tail of a random variable  $v$ , which stochastically dominates some random variable  $\xi$ , then we are also able to control the tail of  $\xi$ . For large values, this tail is essentially the same as the one of a  $\chi^2(k)$  distribution. We use exactly the same ideas as in Panchenko’s lemma 1 and corollary 1 (which assumes an exponential control of the tail of the distribution of the variable  $v$ ).  $\square$

**Lemma 2.** Let  $v$  and  $\xi$  be two real r.v.’s. For  $a \in \mathbb{R}$ , put  $\Phi_a(x) = \max(x - a; 0)$ . Assume that:

(i) for any  $a \in \mathbb{R}$ ,

$$\mathbb{E}\Phi_a(\xi) \leq \mathbb{E}\Phi_a(v)$$

(ii) there exists  $k$  and constants  $C_1 > 0, c_1 > 0$ , such that for any  $t > 0$

$$\mathbb{P}(v \geq t) \leq C_1 \bar{F}_k(c_1 t)$$

then, for  $t > 2k/c_1$ , we have

$$\mathbb{P}(\xi \geq t) \leq C_1 \left( \frac{c_1 t - k}{2} \right)^{\frac{k}{2}} \frac{e^{-\frac{c_1 t - k}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)}$$

and, for  $t > k/c_1$ , we also get

$$\mathbb{P}(\xi \geq t) \leq C_1 \bar{F}_{k+2}(c_1 t - k).$$

**Proof of Lemma 2.** We follow the lines of the proof of Panchenko’s lemma, with a function  $\Phi_a$  with  $a = t - \frac{k}{c_1}$  given by  $\Phi(x) = \max(x - t + k/c_1; 0)$ , for  $t > k/c_1$ . Remark that  $\Phi(x)$  is convex, nondecreasing and that  $\Phi(0) = 0$  and  $\Phi(t) = k/c_1$ . We thus have by Markov’s inequality

$$\mathbb{P}(\xi \geq t) \leq \frac{E\Phi(\xi)}{\Phi(t)} \leq \frac{E\Phi(v)}{\Phi(t)} \leq \frac{1}{\Phi(t)} \left( \Phi(0) + \int_{t-k/c_1}^{+\infty} \Phi^\top(x) \mathbb{P}(v \geq x) dx \right) \leq C_1 \frac{c_1}{k} \int_{t-k/c_1}^{+\infty} \bar{F}_k(c_1 x) dx.$$

By integration by parts, we get

$$\int_{t-k/c_1}^{+\infty} \bar{F}_k(c_1 x) dx = \int_{t-k/c_1}^{+\infty} c_1 x f_k(c_1 x) dx - (t - k/c_1) \int_{t-k/c_1}^{+\infty} c_1 f_k(c_1 x) dx.$$

Recall that

$$f_k(u) = \frac{1}{2^{k/2} \Gamma(\frac{k}{2})} u^{\frac{k}{2}-1} \exp(-\frac{u}{2}),$$

we thus have

$$\frac{c_1}{k} \int_{t-k/c_1}^{+\infty} c_1 x f_k(c_1 x) dx = \frac{c_1}{2^{k/2+1} \frac{k}{2} \Gamma(\frac{k}{2})} \int_{t-k/c_1}^{+\infty} (c_1 x)^{\frac{k+2}{2}-1} \exp(-\frac{c_1 x}{2}) dx = \bar{F}_{k+2}(c_1 t - k).$$

It follows by straightforward calculations that, for  $t > k/c_1$ ,

$$\mathbb{P}(\xi \geq t) \leq C_1 \left( \bar{F}_{k+2}(c_1 t - k) - \frac{c_1 t - k}{k} \bar{F}_k(c_1 t - k) \right).$$

Using the recurrence relation 26.4.8 of Abramovitch and Stegun [1], (page 941), for  $u \geq 2k$ ,

$$C_1 \left( \bar{F}_{k+2}(u - k) - \frac{u - k}{k} \bar{F}_k(u - k) \right) \leq C_1 \left( \bar{F}_{k+2}(u - k) - \bar{F}_k(u - k) \right) \leq \left( \frac{u - k}{2} \right)^{k/2} \frac{C_1 e^{-\frac{(u-k)}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)}.$$

We get with  $u = c_1 t$ , for  $t \geq 2k/c_1$ ,

$$\mathbb{P}(\xi \geq t) \leq \left( \frac{(c_1 t - k)}{2} \right)^{k/2} \frac{C_1 e^{-\frac{(c_1 t - k)}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)}.$$

Moreover, for  $t > k/c_1$  we have  $\mathbb{P}(\xi \geq t) \leq C_1 \left( \bar{F}_{k+2}(c_1 t - k) \right)$ . Notice that we only lose 2 degrees of freedom in this case. It will not be important if  $k$  is large, typically of the order of  $n$  in our case.  $\square$

Now, we extend Panchenko symmetrization lemma (see Panchenko [26] Corollary 1, p. 2069) to the multidimensional framework. Let  $\mathcal{J}_p = \{u \in \mathbb{R}^p, \|u\|_2 = 1\}$  be the unit circle of  $\mathbb{R}^p$ . Let  $X^{(n)} = (X_i)_{1 \leq i \leq n}$  be an independent copy of  $Z^{(n)} = (Z_i)_{1 \leq i \leq n}$ .

Since  $p > n$ , the matrix  $S_n(Z^{(n)} - X^{(n)}) = \frac{1}{n} \sum_{i=1}^n (Z_i - X_i)(Z_i - X_i)^\top$  is not invertible. We derive from  $S_n(Z^{(n)} - X^{(n)})$  the corresponding penalized empirical covariance matrix

$$\tilde{S}_n \equiv \tilde{S}_n(\rho_1, \rho_2) = 2\rho_1 I_p + \rho_2 S_n(Z^{(n)} - X^{(n)}).$$

It is easy to see that

$$\mathbb{E}(S_n(Z^{(n)} - X^{(n)})) = 2\Sigma_n \text{ and } \mathbb{E}(S_n(Z^{(n)} - X^{(n)}) | Z^{(n)}) = S_n + \Sigma_n.$$

Since  $\tilde{S}_n(\tilde{\rho}_1, \tilde{\rho}_2) = \tilde{\rho}_1 I_p + \tilde{\rho}_2 S_n(Z^{(n)} - X^{(n)})$ , we get that

$$\mathbb{E}(\tilde{S}_n | Z^{(n)}) = \tilde{\rho}_1 I_p + \tilde{\rho}_2 (S_n + \Sigma_n) = 2\rho_1 I_p + \rho_2 (S_n + \Sigma_n).$$

As a consequence, define

$$\tilde{\rho}^2 = \mathbb{E} \left( \left\| S_n(Z^{(n)} - X^{(n)}) - 2\Sigma_n \right\|^2 \right) = \mathbb{E} \left( \left\| S_n(Z^{(n)}) - \Sigma_n \right\|^2 \right) + \mathbb{E} \left( \left\| S_n(X^{(n)}) - \Sigma_n \right\|^2 \right) = 2\rho^2.$$

Similarly, put

$$\tilde{\alpha}^2 = 2\alpha^2, \quad \tilde{\delta} = 2\delta^2, \quad \tilde{\sigma}^2 = \langle 2\Sigma_n, I_n \rangle = 2\sigma^2,$$

then we have

$$\tilde{\rho}_1 = \frac{\tilde{\alpha}^2}{\tilde{\delta}^2} \tilde{\sigma}^2 = 2 \frac{\alpha^2}{\delta^2} \sigma^2 = 2\rho_1, \quad \tilde{\rho}_2 = \frac{\tilde{\rho}^2}{\tilde{\delta}^2} = \frac{\rho^2}{\delta^2} = \rho_2.$$

It thus follows with this natural choice of  $\tilde{\rho}_1$  and  $\tilde{\rho}_2$  that we have

$$\mathbb{E}(\tilde{S}_n(\tilde{\rho}_1, \tilde{\rho}_2) | Z^{(n)}) = S_n(\rho_1, \rho_2) + \tilde{\Sigma}_n, \quad \mathbb{E}(\tilde{S}_n(\tilde{\rho}_1, \tilde{\rho}_2)) = 2(\rho_1 I_p + \rho_2 \Sigma_n) = 2\tilde{\Sigma}_n.$$

These quantities will appear in the following symmetrization lemma. Its proof is an extension of Corollary 1 of Panchenko (see [26]) with some adaptations to the multidimensional  $\chi^2$  case. See also Bertail et al. [6] for the non-penalized version of this result for  $p < n$ .

**Lemma 3.** Choose some fixed positive value of  $\rho_1$  and  $\rho_2$ . If there exists  $k \in \mathbb{N}^*$ ,  $C_2 > 0$  and  $c_2 > 0$  such that, for all  $t \geq 0$ ,

$$\mathbb{P} \left( \sup_{u \in \mathcal{J}_p} \left( \frac{\sqrt{nu}^\top (\bar{Z}_n - \bar{X}_n)}{\sqrt{u^\top \tilde{S}_n u}} \right) \geq \sqrt{t} \right) \leq C_2 \bar{F}_k(c_2 t),$$

then, for all  $t \geq 2k/c_2$ ,

$$\mathbb{P} \left( \sup_{u \in J_p} \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top (S_n(\rho_1, \rho_2) + \tilde{\Sigma}_n) u}} \right) \geq \sqrt{t} \right) \leq C_2 \left( \frac{(c_2 t - k)}{2} \right)^{k/2} \frac{e^{-\frac{(c_2 t - k)}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)}$$

and, for all  $t \geq k/c_2$ ,

$$\mathbb{P} \left( \sup_{u \in J_p} \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top (S_n(\rho_1, \rho_2) + \tilde{\Sigma}_n) u}} \right) \geq \sqrt{t} \right) \leq C_2 \bar{F}_{k+2}(c_2 t - k).$$

**Proof of Lemma 3.** Denote

$$A_n(Z^{(n)}) = n \sup_{u \in J_p} \sup_{b > 0} \{ \mathbb{E} [4b(u^\top (\bar{Z}_n - \bar{X}_n) - bu^\top \bar{S}_n u) \mid Z^{(n)}] \}$$

and

$$C_n(Z^{(n)}, X^{(n)}) = n \sup_{u \in J_p} \sup_{b > 0} \{4b(u^\top (\bar{Z}_n - \bar{X}_n) - bu^\top \bar{S}_n u)\}.$$

We have by Jensen’s inequality, that for any convex function  $\phi$

$$\phi(A_n(Z^{(n)})) \leq \mathbb{E} [\phi(C_n(Z^{(n)}, X^{(n)})) \mid Z^{(n)}]. \tag{8}$$

Finally, we can rewrite  $A_n(Z^{(n)})$  and  $C_n(Z^{(n)}, X^{(n)})$  in an explicit form of self-normalized sums by maximizing according to  $b$ , the two expressions above, which leads to

$$A_n(Z^{(n)}) = \sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{\hat{\rho}_1 + \hat{\rho}_2 u^\top (S_n + \Sigma_n) u}} \right)^2 \right\} = \sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u + u^\top \tilde{\Sigma}_n u}} \right)^2 \right\}.$$

Similarly, we have

$$C_n(Z^{(n)}, X^{(n)}) = \sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top (\bar{Z}_n - \bar{X}_n)}{\sqrt{u^\top \bar{S}_n u}} \right)^2 \right\}.$$

Now we conclude by applying Lemma 2 to inequality (8) with these expressions of  $A_n(Z^{(n)})$  and  $C_n(Z^{(n)}, X^{(n)})$  with  $C_2 = C_1$  and  $c_2 = c_1$ .  $\square$

**Proof of Theorem 3.** We now control the Hotelling’s  $T_n^2$  in the general case, by cutting its distribution tail into two parts. The first part allows us to get back to the expression above  $\sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u + u^\top \tilde{\Sigma}_n u}} \right)^2 \right\}$  controlled by Lemma 2. The second term is controlled by the largest eigenvalue of  $\Sigma_n$ .

Let

$$B_n = \sup_{u \in J_p} \left\{ \frac{u^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u}} \right\}.$$

Notice that by construction we have, for any  $t > 0$ , (and particularly for any  $t > 2n$ )

$$\{n \bar{Z}_n^\top S_n(\rho_1, \rho_2)^{-1} \bar{Z}_n \geq t\} = \{n^{1/2} B_n \geq \sqrt{t}\}.$$

To transform the penalized self-normalized sum from the expression  $n \bar{Z}_n^\top (S_n(\rho_1, \rho_2))^{-1} \bar{Z}_n$  to its “pseudo” version with the wrong

normalization,  $\sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u + u^\top \tilde{\Sigma}_n u}} \right)^2 \right\}$ , let us introduce  $D_n$  defined by

$$D_n = \sup_{u \in J_p} \left\{ \sqrt{1 + \frac{u^\top \tilde{\Sigma}_n u}{u^\top S_n(\rho_1, \rho_2) u}} \right\} = \sup_{u \in J_p} \left\{ \sqrt{1 + \frac{u^\top (\rho_1 I_p + \rho_2 \Sigma_n) u}{u^\top (\rho_1 I_p + \rho_2 S_n) u}} \right\}.$$

First, notice that we have

$$\begin{aligned} \sqrt{n} \frac{B_n}{D_n} &= \sup_{u \in J_p} \left\{ \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u}} \right\} \inf_{u \in J_p^*} \left\{ \left( \sqrt{1 + \frac{u^\top \tilde{\Sigma}_n u}{u^\top S_n(\rho_1, \rho_2) u}} \right)^{-1} \right\} \leq \sup_{u \in J_p} \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u}} \left( \sqrt{1 + \frac{u^\top \tilde{\Sigma}_n u}{u^\top S_n(\rho_1, \rho_2) u}} \right)^{-1} \right) \\ &\leq \sqrt{\sup_{u \in J_p} \left\{ \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top S_n(\rho_1, \rho_2) u + u^\top \tilde{\Sigma}_n u}} \right)^2 \right\}}, \end{aligned} \tag{9}$$

for which we have an exponential bound by Lemma 3 and Theorem 1.

Thus by splitting the probability according to the event  $\{D_n^2 \geq 1 + a\}$ , for  $a > 1$  and, for any  $t \geq 2n$ , we have

$$\mathbb{P}(n \bar{Z}_n^\top S_n^{-1}(\rho_1, \rho_2) \bar{Z}_n \geq t) \leq \mathbb{P}\left(B_n \geq \sqrt{\frac{t}{n}}, D_n \leq \sqrt{1+a}\right) + \mathbb{P}(D_n \geq \sqrt{1+a}) \leq \mathbb{P}\left(\frac{B_n}{D_n} \geq \sqrt{\frac{t}{n(1+a)}}\right) + \mathbb{P}(D_n \geq \sqrt{1+a}). \tag{10}$$

So now, it remains to treat the second term in the right-hand side of inequality (10). Notice that we have, for  $a > 1$ ,

$$\{D_n \geq \sqrt{1+a}\} = \left\{ \sup_{u \in J_p} \left( \frac{u^\top \tilde{\Sigma}_n u}{u^\top S_n(\rho_1, \rho_2) u} \right) \geq a \right\} = \left\{ \inf_{u \in J_p^*} \left( \frac{u^\top S_n(\rho_1, \rho_2) u}{u^\top \tilde{\Sigma}_n u} \right) \leq \frac{1}{a} \right\}.$$

First, if  $\Sigma_n = \sigma^2 I_p$  is diagonal, then we have

$$u^\top \tilde{\Sigma}_n u = u^\top (\rho_1 I_p + \rho_2 \sigma^2 I_p) u = \rho_1 + \rho_2 \sigma^2.$$

Since

$$\inf_{u \in J_p} (u^\top S_n(\rho_1, \rho_2) u) = \inf_{u \in J_p} (u^\top (\rho_1 I_p + \rho_2 S_n) u) = \rho_1,$$

if we choose  $a$  such that  $a > (\rho_1 + \rho_2 \sigma^2) / \rho_1$ , then we have

$$\mathbb{P}[D_n \geq \sqrt{1+a}] \leq \mathbb{P}\left(\frac{\inf_{u \in J_p} (u^\top S_n(\rho_1, \rho_2) u)}{\rho_1 + \rho_2 \sigma^2} \leq \frac{1}{a}\right) = 0.$$

Remark that, in this case, we have  $\rho_1^* = \sigma^2$  and  $\rho_2^* = 0$  and it follows that the inequality is true for any  $a > 1$ . Notice that the proximity between  $\Sigma_n$  and  $\sigma^2 I_p$  is precisely controlled by the term  $\alpha^2 = \|\Sigma_n - \sigma^2 I_p\|$ .

Now consider the general case. For sake of clarity, we will use the notation  $\mu_{\min}(A)$  for the minimum eigenvalue of a matrix  $A$ , and respectively  $\mu_{\max}(A)$  for the maximum. First, notice that

$$\begin{aligned} \inf_{u \in J_p} \left( \frac{u^\top S_n(\rho_1, \rho_2) u}{u^\top \tilde{\Sigma}_n u} \right) &= \inf_{u \in J_p} \left( u^\top \tilde{\Sigma}_n^{-1} S_n(\rho_1, \rho_2) \tilde{\Sigma}_n^{-1} u \right) = \inf_{u \in J_p} \left( \frac{\overbrace{u^\top \tilde{\Sigma}_n^{-1}}^{=v^\top} S_n(\rho_1, \rho_2) \overbrace{\tilde{\Sigma}_n^{-1} u}^{=v}}{\|\tilde{\Sigma}_n^{-1} u\|_2} \frac{\|\tilde{\Sigma}_n^{-1} u\|_2}{\|\tilde{\Sigma}_n^{-1} u\|_2} \|\tilde{\Sigma}_n^{-1} u\|_2^2 \right) \\ &\geq \inf_{v \in J_p} (v^\top S_n(\rho_1, \rho_2) v) \times \inf_{u \in J_p} (u^\top \tilde{\Sigma}_n^{-2} u) \geq \rho_1 \mu_{\min}(\tilde{\Sigma}_n^{-1}) = \frac{\rho_1}{\mu_{\max}(\tilde{\Sigma}_n)}. \end{aligned}$$

Now, using the optimal values  $\rho_1^*$  and  $\rho_2^*$ , we have the decomposition

$$\tilde{\Sigma}_n(\rho_1^*, \rho_2^*) = \rho_1^* I_p + \rho_2^* \Sigma_n,$$

it follows that we get

$$\mu_{\max}(\tilde{\Sigma}_n(\rho_1^*, \rho_2^*)) = \rho_1^* + \rho_2^* \mu_{\max}(\Sigma_n), \quad \inf_{u \in J_p} \left( \frac{u^\top S_n(\rho_1^*, \rho_2^*) u}{u^\top \tilde{\Sigma}_n(\rho_1^*, \rho_2^*) u} \right) \geq \frac{\rho_1^*}{\rho_1^* + \rho_2^* \mu_{\max}(\Sigma_n)}.$$

It follows that if we choose  $a$  such that

$$\frac{1}{a} < \frac{1}{1 + \frac{\mu_{\max}(\Sigma_n)}{\rho^*}}$$

and, since  $a^* = 1 + \frac{K_3}{\rho^*} > 1 + \frac{\mu_{\max}(\Sigma_n)}{\rho^*}$  by Assumption  $A_3$ , then, if  $a \geq a^*$ , we get

$$\mathbb{P}(D_n \geq \sqrt{1+a}) = 0. \tag{11}$$



As a consequence, we obtain an exponential inequality for any value  $a \geq a^*$ . Combining inequalities (10) and (11), we get, for any  $a \geq a^*$ ,

$$\forall t \geq 2n, \quad \mathbb{P} \left( n \bar{Z}_n^\top S_n^{-1} (\rho_1, \rho_2) \bar{Z}_n \geq t(1+a) \right) \leq \mathbb{P} \left( \sqrt{n} \frac{B_n}{D_n} \geq \sqrt{t} \right). \tag{12}$$

Let  $X^{(n)} = (X_i)_{1 \leq i \leq n}$  be an independent copy of  $Z^{(n)} = (Z_i)_{1 \leq i \leq n}$ . Applying Theorem 1 to  $(Z_i - X_i)_{1 \leq i \leq n}$  which is symmetric, we obtain

$$\mathbb{P} \left( \sup_{u \in \mathcal{J}_p} \left( \frac{\sqrt{nu}^\top (\bar{Z}_n - \bar{X}_n)}{\sqrt{u^\top \tilde{S}_n u}} \right) \geq \sqrt{t} \right) \leq \frac{2e^3}{9} \bar{F}_n(t).$$

Thus, applying Lemma 3 to the inequality above implies that, for all  $t \geq 2n$ ,

$$\mathbb{P} \left( \sup_{u \in \mathcal{J}_p} \left( \frac{\sqrt{nu}^\top \bar{Z}_n}{\sqrt{u^\top (S_n(\rho_1, \rho_2) + \tilde{S}_n) u}} \right) \geq \sqrt{t} \right) \leq \frac{2e^3}{9} \left( \frac{(t-n)}{2} \right)^{n/2} \frac{e^{-\frac{(t-n)}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)}. \tag{13}$$

Finally by combining (9), (12) and (13), the result holds.  $\square$

**Proof of Theorem 4.** The following lemmas will allow us to control explicitly the deviation  $\mathbb{P} \left[ \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| > \epsilon \right]$  for small positive values of  $\epsilon$ .  $\square$

**Lemma 4 (Inversion).** Let  $w > 0$ , and consider  $(W_n)_{n \geq 1}$  a sequence of nonnegative random variables. Assume that there exists a positive constant  $C_3$ , such that  $\forall \epsilon > 0, \exists N > 0, \forall n > N$ ,

$$\mathbb{P} (|W_n - w| > \epsilon) \leq \frac{C_3}{n} \frac{1}{\epsilon^2}.$$

Then there exists a function  $C_{3;1/w}$  positive, such that  $\forall \epsilon > 0, \forall n > N$

$$\mathbb{P} \left( \left| \frac{1}{W_n} - \frac{1}{w} \right| > \epsilon \right) \leq \frac{C_{3;1/w}(\epsilon)}{n\epsilon^2},$$

where  $C_{3;1/w}(\epsilon) = \frac{C_3}{w^4} (1 + (w\epsilon)^{2/5})^5$ .

**Proof of Lemma 4.** Since  $w > 0$ , we have

$$\mathbb{P} \left( \left| \frac{1}{W_n} - \frac{1}{w} \right| > \frac{\epsilon}{w} \right) = \mathbb{P} \left( \left| \frac{w}{W_n} - 1 \right| > \epsilon \right).$$

Now,  $\forall \eta \in ]0, w[$  we get

$$\mathbb{P} \left( \left| \frac{w}{W_n} - 1 \right| > \frac{\epsilon}{w} \right) \leq \overbrace{\mathbb{P} \left( \left| \frac{w}{W_n} - 1 \right| > \epsilon, |W_n - w| \leq \eta \right)}^{(I)} + \overbrace{\mathbb{P} (|W_n - w| > \eta)}^{(II)}.$$

On the interval  $[w - \eta; w + \eta]$ ,  $f : x \mapsto \frac{w}{x}$  is Lipschitz with  $|f'(x)| \leq \frac{w}{(w-\eta)^2}$ , thus we obtain

$$\forall W_n \in [w - \eta; w + \eta], \quad \left| \frac{w}{W_n} - 1 \right| \leq \frac{w}{(w-\eta)^2} |W_n - w|.$$

$$\forall \eta \in ]0; w[, \quad (I) \leq \mathbb{P} \left( \frac{w}{(w-\eta)^2} |W_n - w| > \epsilon \right) \leq \frac{C_3}{n} \frac{w^2}{\epsilon^2 (w-\eta)^4}$$

and since

$$\forall \eta \in ]0; w[, \quad (II) \leq \frac{C_3}{n} \frac{1}{\eta^2},$$

it follows that

$$\begin{aligned} \mathbb{P} \left( \left| \frac{1}{W_n} - \frac{1}{w} \right| > \frac{\epsilon}{w} \right) &\leq \frac{C_3}{n} \times \frac{w^2}{\epsilon^2 (w-\eta)^4} + \frac{C_3}{n} \times \frac{1}{\eta^2} \leq \frac{C_3}{n} \min_{\eta \in ]0; w[} \left\{ \frac{w^2}{\epsilon^2 \left(1 - \frac{\eta}{w}\right)^4 w^4} + \frac{1}{w^2 \left(\frac{\eta}{w}\right)^2} \right\} \\ &\stackrel{\alpha = \frac{\eta}{w}}{\leq} \frac{C_3}{nw^2} \min_{\alpha \in ]0; 1[} \left\{ \frac{1}{\epsilon^2 (1-\alpha)^4} + \frac{1}{\alpha^2} \right\} \leq \frac{C_3}{nw^2} \min_{\alpha \in ]0; 1[} \left\{ \frac{1}{\epsilon^2 (1-\alpha)^4} + \frac{1}{\alpha^4} \right\} \leq \frac{C_3}{nw^2} (1 + \epsilon^{-2/5})^5. \end{aligned}$$

Setting  $\epsilon^\top = \frac{\epsilon}{w}$  and  $C_{3;1/w}(\epsilon^\top) = \epsilon^{\top 2} \times \frac{C_3}{w^2} (1 + (w\epsilon^\top)^{-2/5})^5 = \frac{C_3}{w^4} (1 + (w\epsilon^\top)^{-2/5})^5$ , the result holds.  $\square$

**Lemma 5.** Consider  $u, v$  two positive scalars, and  $(U_n), (V_n)$  some random sequences. Assume that there exists positive constants  $\tilde{C}_4$  and  $\check{C}_4$  such that  $\forall \epsilon > 0, \forall n \geq 1$ :

$$\mathbb{P}(|U_n - u| > \epsilon) \leq \frac{\tilde{C}_4}{n} \frac{1}{\epsilon^2}, \quad \mathbb{P}(|V_n - v| > \epsilon) \leq \frac{\check{C}_4}{n} \frac{1}{\epsilon^2}.$$

Then there exists a function  $C_{4;uv}$  such that  $\forall \epsilon > 0$ ,

$$\mathbb{P}(|U_n V_n - uv| > \epsilon) \leq \frac{C_{4;uv}(\epsilon)}{n} \frac{1}{\epsilon^2},$$

where  $C_{4;uv}(\epsilon) = \tilde{C}_4 \left(\frac{2uv+\epsilon}{u}\right)^2 + \check{C}_4 (2u)^2$  is a positive function of  $\epsilon$  depending on  $u, v, \tilde{C}_4$  and  $\check{C}_4$ .

**Proof of Lemma 5.** By straightforward inequalities, we get for any  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{P}(|U_n V_n - uv| > \epsilon) &= \mathbb{P}(|U_n V_n - uV_n + uV_n - uv| > \epsilon) \leq \mathbb{P}\left(V_n |U_n - u| > \frac{\epsilon}{2}, u |V_n - v| \leq \frac{\epsilon}{2}\right) + \mathbb{P}\left(u |V_n - v| > \frac{\epsilon}{2}\right) \\ &\leq \mathbb{P}\left(\left(v + \frac{\epsilon}{2u}\right) |U_n - u| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|V_n - v| > \frac{\epsilon}{2u}\right) \leq \mathbb{P}\left(|U_n - u| > \frac{\epsilon u}{2uv + \epsilon}\right) + \mathbb{P}\left(|V_n - v| > \frac{\epsilon}{2u}\right) \\ &\leq \frac{\tilde{C}_4}{n} \left(\frac{2uv + \epsilon}{\epsilon u}\right)^2 + \frac{\check{C}_4}{n} \left(\frac{2u}{\epsilon}\right)^2 \leq \frac{C_{4;uv}(\epsilon)}{n} \frac{1}{\epsilon^2}. \quad \square \end{aligned}$$

**Lemma 6 (Proximity Between  $\sigma^2, \alpha^2, \beta^2, \delta^2$  and Their Estimators).** Let  $u^2 \in \{\sigma^2, \alpha^2, \beta^2, \delta^2\}$  be one of these quantities of interest and  $\hat{u}_n^2$  its corresponding estimator. Then  $\forall n \geq 1$  and  $\forall \epsilon > 0$ , we have:

$$\mathbb{P}\left(|\hat{u}_n^2 - u^2| > \epsilon\right) \leq \frac{C_{u^2}(\epsilon)}{n\epsilon^2},$$

with

- $C_{\sigma^2} = \sqrt{K_2}$  for the case where  $u^2 = \sigma^2$  and  $\hat{u}_n^2 = \sigma_n^2$ ,
- $C_{\delta^2} = 2K_4 + (100 + K_1^2)K_2 + 2^4\sqrt{6}K_2^{5/4} + 4K_2^{3/2} + 2^2 3K_2^2 + 4K_2^{1/2} \left(K_2^{1/4} + 2\sqrt{6}\right) \sqrt{K_1^2 K_2 + 4K_2(1 + 3K_2) + 2K_4}$  for the case where  $u^2 = \delta^2$  and  $\hat{u}_n^2 = \delta_n^2$ ,
- $C_{\beta^2}(\epsilon) = 4K_1^2\sqrt{K_2} + C_{\delta^2} + 2K_1\sqrt{K_2} \epsilon$  for the case where  $u^2 = \beta^2$  and  $\hat{u}_n^2 = \beta_n^2$ ,
- $C_{\alpha^2}(\epsilon) = 2^3 C_{\delta^2} + 2^4 K_1^2 \sqrt{K_2} + 2^2 K_1 \sqrt{K_2} \epsilon$  for the case where  $u^2 = \alpha^2$  and  $\hat{u}_n^2 = \alpha_n^2$ .

**Proof of Lemma 6.** Consider  $\hat{\sigma}_n^2$  and  $\sigma^2$ . Recall that  $\hat{\sigma}_n^2 = \frac{1}{p} \sum_{j=1}^p \left(\frac{1}{n} \sum_{i=1}^n y_{ij}^2\right)$  and  $\sigma^2 = \frac{1}{p} \sum_{j=1}^p \mathbb{E} [y_{1j}^2] = \frac{1}{p} \sum_{j=1}^p \mu_j$ .

Following the ideas of Ledoit and Wolf [19] who obtain the convergence of the fourth order moment, we rather control the second order moment as follows:

$$\begin{aligned} \mathbb{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2\right] &= \mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (y_{ij}^2 - \mu_j)\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{p} \sum_{j=1}^p (y_{ij}^2 - \mu_j)\right)^2\right] \\ &= \frac{1}{n^2} \sum_{i_1=1}^n \sum_{i_2=1}^n \mathbb{E}\left[\frac{1}{p} \sum_{j=1}^p (y_{i_1 j}^2 - \mu_j) \times \frac{1}{p} \sum_{j=1}^p (y_{i_2 j}^2 - \mu_j)\right]. \end{aligned}$$

This last expression is equal to zero for any  $i_1 \neq i_2$  because of the independence between observations. Thus we get

$$\begin{aligned} \mathbb{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2\right] &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p (y_{ij}^2 - \mu_j)\right)^2\right] = \frac{1}{n} \mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p (y_{1j}^2 - \mu_j)\right)^2\right] = \frac{1}{n} \left(\mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p y_{1j}^2\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{p} \sum_{j=1}^p y_{1j}^2\right]\right)^2\right) \\ &\leq \frac{1}{n} \mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p y_{1j}^2\right)^2\right] \leq \frac{1}{n} \left(\mathbb{E}\left[\left(\frac{1}{p} \sum_{j=1}^p y_{1j}^2\right)^4\right]\right)^{1/2} \leq \frac{1}{n} \left(\frac{1}{p} \sum_{j=1}^p \mathbb{E}[y_{1j}^8]\right)^{1/2}. \end{aligned}$$

Therefore, using the second assumption  $A_2$ , one gets

$$\mathbb{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2\right] \leq \frac{\sqrt{K_2}}{n}. \tag{14}$$

Finally, we have by Markov inequality the bound

$$\forall \epsilon > 0, \quad \mathbb{P}\left[|\hat{\sigma}_n^2 - \sigma^2| > \epsilon\right] \leq \frac{\mathbb{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2\right]}{\epsilon^2} \leq \frac{\sqrt{K_2}}{n\epsilon^2}.$$

Consider  $\hat{\delta}_n^2$  and  $\delta^2$ . Combining the expressions (A.2) and (A.3) on page 394 in Ledoit and Wolf [19], we get

$$\hat{\delta}_n^2 - \delta^2 = (\hat{\sigma}_n^2 - \sigma^2)^2 - 2\sigma^2 (\hat{\sigma}_n^2 - \sigma^2) + \|S_n\|^2 - \mathbb{E}\left(\|S_n\|^2\right).$$

Similarly using their expressions, (A.4) from page 394 to page 399, and (A.1) on page 390, we have respectively the inequalities

$$\text{Var} \left( \|S_n\|^2 \right) \leq \frac{1}{n} \left( K_1^2 K_2 + 4K_2 (1 + 3K_2) + 2K_4 \right), \quad \sigma^2 \leq \sqrt{K_2}.$$

Combining these expressions with Bienaymé-Tchebychev, Markov and Cauchy-Schwartz inequalities, we obtain a control of  $\mathbb{P} \left( \left| \hat{\delta}_n^2 - \delta^2 \right| > \epsilon \right)$  by a function of  $n, \epsilon, A_2, A_4$  and  $\text{Var}(\|S_n\|^2)$  where  $A_k = \mathbb{E} \left( \left| \hat{\sigma}_n^2 - \sigma^2 \right|^k \right)$ . Indeed we have, by Markov inequality, for all  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \left| \hat{\delta}_n^2 - \delta^2 \right| > \epsilon \right) &\leq \frac{1}{\epsilon^2} \left\{ \mathbb{E} \left[ \left( \hat{\sigma}_n^2 - \sigma^2 \right)^4 \right] + 4\sigma^4 \mathbb{E} \left[ \left( \hat{\sigma}_n^2 - \sigma^2 \right)^2 \right] + \mathbb{E} \left[ \left( \|S_n\|^2 - \mathbb{E} \|S_n\|^2 \right)^2 \right] + 4\sigma^2 \mathbb{E} \left[ \left| \hat{\sigma}_n^2 - \sigma^2 \right|^3 \right] \right. \\ &\quad \left. + 4\sigma^2 \mathbb{E} \left[ \left| \hat{\sigma}_n^2 - \sigma^2 \right| \left( \|S_n\|^2 - \mathbb{E} \|S_n\|^2 \right) \right] + 2\mathbb{E} \left[ \left( \hat{\sigma}_n^2 - \sigma^2 \right)^2 \left| \|S_n\|^2 - \mathbb{E} \|S_n\|^2 \right| \right] \right\} \\ &\leq \frac{1}{\epsilon^2} \left\{ A_4 + 4\sigma^4 A_2 + \text{Var} \left( \|S_n\|^2 \right) + 4\sigma^2 \sqrt{A_2 A_4} + 4\sigma^2 \sqrt{A_2 \text{Var} \left( \|S_n\|^2 \right)} + 2\sqrt{A_4 \text{Var} \left( \|S_n\|^2 \right)} \right\}. \end{aligned}$$

Now by some previous controls established by Ledoit and Wolf (2004) [19] (page 394), we have

$$A_4 \leq \frac{96K_2}{n}; \quad \text{Var} \left( \|S_n\|^2 \right) \leq \frac{1}{n} \left( K_1^2 K_2 + 4K_2 (1 + 3K_2) + 2K_4 \right) = \frac{1}{n} K \quad \text{and} \quad \sigma^2 \leq \sqrt{K_2}.$$

Using the control stated in inequality (14),  $A_2 \leq \sqrt{K_2}/n$ , we can easily get the explicit constant  $C_{\delta^2}$  as a function of  $K_1, K_2$ , and  $K_4$ . For all  $\epsilon > 0$ , for all  $n \in \mathbb{N}^*$ , we have

$$\begin{aligned} \mathbb{P} \left( \left| \hat{\delta}_n^2 - \delta^2 \right| > \epsilon \right) &\leq \frac{1}{n\epsilon^2} \left[ 96K_2 + 4K_2 K_2^{1/2} + K + 4K_2^{1/2} \sqrt{96K_2^{1/2} K_2} + 4K_2^{1/2} \sqrt{K_2^{1/2} K} + 2\sqrt{96K_2 K} \right] \\ &\leq \frac{1}{n\epsilon^2} \left\{ 2K_4 + (100 + K_1^2)K_2 + 2^4 \sqrt{6} K_2^{5/4} + 4K_2^{3/2} + 2^2 3K_2^2 + 4K_2^{1/2} \left( K_2^{1/4} + 2\sqrt{6} \right) \sqrt{K_1^2 K_2 + 4K_2 (1 + 3K_2) + 2K_4} \right\} \leq \frac{C_{\delta^2}}{n\epsilon^2}. \end{aligned}$$

Consider  $\hat{\beta}_n^2$  and  $\beta^2$ . Since  $\delta^2 = \alpha^2 + \beta^2$  yielding  $\delta^2 \geq \beta^2$ , Ledoit and Wolf [19] showed (proof of Lemma 3.4 page 401, lines from -12 to -6) that

$$-\max \left( \left| \hat{\beta}_n^2 - \beta^2 \right|, \left| \hat{\delta}_n^2 - \delta^2 \right| \right) \leq \hat{\beta}_n^2 - \beta^2 \leq \left| \hat{\beta}_n^2 - \beta^2 \right|.$$

From this we deduce

$$\left| \hat{\beta}_n^2 - \beta^2 \right| \leq \max \left\{ \max \left( \left| \hat{\beta}_n^2 - \beta^2 \right|, \left| \hat{\delta}_n^2 - \delta^2 \right| \right), \left| \hat{\beta}_n^2 - \beta^2 \right| \right\} \leq \max \left( \left| \hat{\beta}_n^2 - \beta^2 \right|, \left| \hat{\delta}_n^2 - \delta^2 \right| \right).$$

Controlling  $\left| \hat{\beta}_n^2 - \beta^2 \right|$  leads to a control for  $\left| \hat{\delta}_n^2 - \delta^2 \right|$  and  $\left| \hat{\beta}_n^2 - \beta^2 \right|$ . By the same arguments as in Ledoit and Wolf [19] (proof of Lemma 3.4, page 399, equation (A.7)), we have the following expression

$$\hat{\beta}_n^2 - \beta^2 = \frac{1}{n} \|S_n - \Sigma_n\|^2 + \left( \frac{1}{n^2} \sum_{i=1}^n \|Z_i Z_i^\top - \Sigma_n\|^2 - \mathbb{E} \left[ \frac{1}{n^2} \sum_{i=1}^n \|Z_i Z_i^\top - \Sigma_n\|^2 \right] \right).$$

Now, splitting the probability into two terms, on the one hand, using Markov inequality on the first term and applying Bienaymé-Tchebychev inequality to the second term, we get for all  $\epsilon > 0$ ,

$$\mathbb{P} \left( \left| \hat{\beta}_n^2 - \beta^2 \right| > \epsilon \right) \leq \frac{2}{\epsilon} \mathbb{E} \left( \frac{1}{n} \|S_n - \Sigma_n\|^2 \right) + \frac{4}{\epsilon^2} \text{Var} \left( \frac{1}{n^2} \sum_{i=1}^n \|Z_i Z_i^\top - \Sigma_n\|^2 \right).$$

Following Ledoit and Wolf [19] (proof of Lemma 3.1 page 391 line +5), we have

$$\mathbb{E} \left( \|S_n - \Sigma_n\|^2 \right) \leq K_1 \sqrt{K_2}.$$

Moreover, we have by Ledoit and Wolf [19] (in the proof of Lemma 3.4, page 401 line +3)

$$\text{Var} \left( \frac{1}{n^2} \sum_{i=1}^n \|Z_i Z_i^\top - \Sigma_n\|^2 \right) \leq K_1^2 \sqrt{K_2}/n.$$

We obtain for any  $\epsilon > 0$ ,

$$\mathbb{P} \left( \left| \hat{\beta}_n^2 - \beta^2 \right| > \epsilon \right) \leq \frac{2}{\epsilon} \frac{K_1 \sqrt{K_2}}{n} + \frac{4}{\epsilon^2} \frac{K_1^2 \sqrt{K_2}}{n}.$$

Finally, with  $\mathbb{P} \left( \left| \hat{\delta}_n^2 - \delta^2 \right| > \epsilon \right) \leq \frac{C_{\delta^2}}{n\epsilon^2}$  and

$$\mathbb{P} \left( \left| \hat{\beta}_n^2 - \beta^2 \right| > \epsilon \right) \leq \mathbb{P} \left( \left| \hat{\beta}_n^2 - \beta^2 \right| > \epsilon \right) + \mathbb{P} \left( \left| \hat{\delta}_n^2 - \delta^2 \right| > \epsilon \right),$$

we obtain for any  $\epsilon > 0$ ,

$$\mathbb{P} \left( \left| \hat{\beta}_n^2 - \beta^2 \right| > \epsilon \right) \leq \frac{1}{n\epsilon^2} \left( 4K_1^2 \sqrt{K_2} + C_{\delta^2} + 2K_1 \sqrt{K_2} \epsilon \right) \leq \frac{C_{\beta^2}(\epsilon)}{n\epsilon^2}.$$

Remark that  $C_{\beta^2}(\epsilon)$  tends to  $4K_1^2\sqrt{K_2} + C_{\delta^2}$  when  $\epsilon$  tends to 0. Consider  $\hat{\alpha}_n^2$  and  $\alpha^2$ . Since we have  $\hat{\alpha}_n^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2$  and  $\alpha^2 + \beta^2 = \delta^2$ , one can easily see that  $\hat{\alpha}_n^2 - \alpha^2 = \hat{\delta}_n^2 - \hat{\beta}_n^2 - \delta^2 + \beta^2$ . For all  $\epsilon > 0$ , we get

$$\begin{aligned} \mathbb{P}\left(\left|\hat{\alpha}_n^2 - \alpha^2\right| > \epsilon\right) &\leq \mathbb{P}\left(\left|\hat{\delta}_n^2 - \delta^2\right| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left|\hat{\beta}_n^2 - \beta^2\right| > \frac{\epsilon}{2}\right) \\ &\leq \frac{2^2C_{\delta^2}}{n\epsilon^2} + \frac{2^2C_{\beta^2}(\epsilon/2)}{n\epsilon^2} \leq \frac{1}{n\epsilon^2} \left(2^3C_{\delta^2} + 2^4K_1^2\sqrt{K_2} + 2^2K_1\sqrt{K_2}\epsilon\right) \leq \frac{C_{\alpha^2}(\epsilon)}{n\epsilon^2}. \end{aligned}$$

Remark that  $C_{\alpha^2}(\epsilon)$  tends to  $2^3C_{\delta^2} + 2^4K_1^2\sqrt{K_2}$  when  $\epsilon$  tends to 0.  $\square$

In the next lemma, we control the proximity between  $1/\hat{\rho}_n^*$  and  $1/\rho^*$ , that we denote  $g_n(\epsilon)$  and show that it is of order  $O(1/n)$ . For this, we first apply product Lemma 5 to  $\hat{\beta}_n^2$  and  $\hat{\sigma}_n^2$ . Then, we apply the inverse Lemma 4 to  $\hat{\beta}_n^2\hat{\sigma}_n^2$ . Finally, we use another time product Lemma 5 applied to  $\hat{\alpha}_n^2$  and  $1/\hat{\beta}_n^2\hat{\sigma}_n^2$ .

**Lemma 7 (Proximity Between  $1/\rho^*$  and  $1/\hat{\rho}_n^*$ ).**

For any  $\epsilon > 0$ , we have

$$g_n(\epsilon) = \mathbb{P}\left(\left|\frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*}\right| > \epsilon\right) \leq \frac{G(\epsilon)}{n\epsilon^2}$$

with  $C_{\beta^2}$  and  $C_{\alpha^2}$  defined in Lemma 6 and

$$\begin{aligned} G(\epsilon) &= C_{3;1/\beta^2\sigma^2}(\epsilon) (2\alpha^2 + \epsilon\beta^2\sigma^2)^2 + \frac{2^2C_{\alpha^2}(\epsilon)}{\beta^4\sigma^4} \\ C_{3;1/\beta^2\sigma^2}(\epsilon) &= \left[ K_2^{1/2} \frac{(2\sigma^2\beta^2 + \epsilon)^2}{\beta^8\sigma^{12}} + \frac{2^2C_{\beta^2}(\epsilon)}{\beta^8\sigma^4} \right] \left(1 + (\beta^2\sigma^2\epsilon)^{2/5}\right)^5. \end{aligned}$$

**Remark 6.** the function  $C_{3;1/\beta^2\sigma^2}(\epsilon)$  may be clearly bounded by a polynomial of degree 4 in  $\epsilon$ . As a consequence, the function  $G(\epsilon)$  may be bounded by a polynomial of degree 6.

**Proof of Lemma 7.** We apply the product Lemma 5 to obtain a control for  $\hat{\beta}_n^2\hat{\sigma}_n^2$  thanks to Lemma 6 which gives us some control of  $\hat{\sigma}_n^2$  and  $\hat{\beta}_n^2$ . For all  $\epsilon > 0$ , one gets

$$\mathbb{P}\left(\left|\hat{\beta}_n^2\hat{\sigma}_n^2 - \beta^2\sigma^2\right| > \epsilon\right) \leq \frac{C_{4;\beta^2\sigma^2}(\epsilon)}{n\epsilon^2}, \tag{15}$$

with

$$C_{4;\beta^2\sigma^2}(\epsilon) = K_2^{1/2} \left(\frac{2\sigma^2\beta^2 + \epsilon}{\sigma^2}\right)^2 + C_{\beta^2}(\epsilon) (2\sigma^2)^2.$$

We now apply the inverse Lemma 4 with inequality (15) to obtain a control of  $1/\hat{\beta}_n^2\hat{\sigma}_n^2$ . That is, for all  $\epsilon > 0$ , we have

$$\mathbb{P}\left(\left|\frac{1}{\hat{\beta}_n^2\hat{\sigma}_n^2} - \frac{1}{\beta^2\sigma^2}\right| > \epsilon\right) \leq \frac{C_{3;1/\beta^2\sigma^2}(\epsilon)}{n\epsilon^2},$$

with  $C_{3;1/\beta^2\sigma^2}$  defined by  $C_{3;1/\beta^2\sigma^2}(\epsilon) = \frac{C_{4;\beta^2\sigma^2}(\epsilon)}{\beta^8\sigma^8} \left(1 + (\beta^2\sigma^2\epsilon)^{2/5}\right)^5$ . Applying the product Lemma 5 with  $u = 1/(\beta^2\sigma^2)$  and  $v = \alpha^2$ , we obtain for all  $\epsilon > 0$ ,

$$C_{4;1/\rho^*}(\epsilon) = C_{3;1/\beta^2\sigma^2}(\epsilon) (2\alpha^2 + \epsilon\beta^2\sigma^2)^2 + C_{\alpha^2}(\epsilon) \frac{2^2}{\beta^4\sigma^4}.$$

Remark that when  $\epsilon$  tends to 0,  $C_{4;1/\rho^*}$  tends to  $\frac{2^4\alpha^4K_2^{1/2}}{\beta^4\sigma^8} + \frac{2^6\alpha^4(2^2K_1^2\sqrt{K_2} + C_{\delta^2})}{\beta^8\sigma^4} + \frac{2^5(2K_1^2\sqrt{K_2} + C_{\delta^2})}{\beta^4\sigma^4}$   $\square$

**Proof of Theorem 4.** Recall that  $\hat{a}_n^* = 1 + \frac{K_3}{\hat{\rho}_n^*}$  and  $a^* = 1 + \frac{K_3}{\rho^*}$ . For any  $u \geq 2n$ , we have for  $\epsilon > 0$

$$\mathbb{P}(n\bar{Z}_n^T \hat{S}_n^{*-1} \bar{Z}_n \geq u(1 + \hat{a}_n^* + 2\epsilon)) \leq \mathbb{P}(n\bar{Z}_n^T \hat{S}_n^{*-1} \bar{Z}_n \geq u(1 + a^* + \epsilon)) + \mathbb{P}(|\hat{a}_n - a^*| \geq \epsilon) \leq \text{(I)} + \text{(II)}. \tag{16}$$

We start by establishing a control for (I). Define  $\Delta_n = n\bar{Z}_n^T (\hat{S}_n^{*-1} - S_n^{*-1}) \bar{Z}_n$ , then we have (I) =  $\mathbb{P}(n\bar{Z}_n^T S_n^{*-1} \bar{Z}_n + \Delta_n \geq u(1 + a^* + \epsilon))$ . Since  $u \geq 2n > n$ , we have

$$\begin{aligned} \text{(I)} &\leq \mathbb{P}(n\bar{Z}_n^T S_n^{*-1} \bar{Z}_n + \Delta_n \geq u(1 + a^* + \epsilon), |\Delta_n| \leq \epsilon n) + \mathbb{P}(|\Delta_n| > \epsilon n) \\ &\leq \mathbb{P}(n\bar{Z}_n^T S_n^{*-1} \bar{Z}_n \geq u(1 + a^* + \epsilon) - \epsilon n) + \mathbb{P}(|\Delta_n| > \epsilon n) \leq \mathbb{P}(n\bar{Z}_n^T S_n^{*-1} \bar{Z}_n \geq u(1 + a^*)) + \mathbb{P}(|\Delta_n| > \epsilon n). \end{aligned} \tag{17}$$

Theorem 3 gives us an exponential bound controlling the first term of the right hand of inequality (17) when  $a = a^*$  and  $u \geq 2n$ .

Now use the following matrix factorization  $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$  to control the second term in the right hand of inequality (17) with  $A = \hat{S}_n^*$  and  $B = \Sigma_n^{*2}$ . It is easy to see that  $B - A = (\rho^* - \hat{\rho}_n^*) I_p$ , then we obtain

$$\Delta_n = Tr(\Delta_n) = Tr(n\bar{Z}_n^T (\hat{S}_n^{*-1} - S_n^{*-1}) \bar{Z}_n) = n(\rho^* - \hat{\rho}_n^*) Tr(\bar{Z}_n^T \hat{S}_n^{*-1} S_n^{*-1} \bar{Z}_n).$$

Recall that  $S_n^* = S_n + \rho^* I_p = O_n \Lambda_n O_n^T + \rho^* I_p = O_n (\Lambda_n + \rho^* I_p) O_n^T$ , then using the same rotation matrix  $O_n$ , we obtain  $\hat{S}_n^{*-1} S_n^{*-1} = O_n D O_n^T$ , with  $D = \text{diag} \left( \frac{1}{(\lambda_1 + \rho^*)(\lambda_1 + \hat{\rho}_n^*)}, \dots, \frac{1}{(\lambda_n + \rho^*)(\lambda_n + \hat{\rho}_n^*)}, \frac{1}{\rho^* \hat{\rho}_n^*}, \dots, \frac{1}{\rho^* \hat{\rho}_n^*} \right)$  where  $\text{diag}$  is a diagonal matrix. It follows that

$$\Delta_n = n(\rho^* - \hat{\rho}_n^*) Tr \left( \bar{Z}_n^T O_n D^{\frac{1}{2}} D^{\frac{1}{2}} O_n^T \bar{Z}_n \right) = (\rho^* - \hat{\rho}_n^*) Tr \left( \left( D^{\frac{1}{2}} n^{\frac{1}{2}} O_n^T \bar{Z}_n \right)' \left( D^{\frac{1}{2}} n^{\frac{1}{2}} O_n^T \bar{Z}_n \right) \right) = (\rho^* - \hat{\rho}_n^*) \left\| D^{\frac{1}{2}} n^{\frac{1}{2}} \bar{Y}_n \right\|_2^2.$$

Since, for any  $x$  in  $\mathbb{R}^p$ ,  $\|D^{\frac{1}{2}} x\|_2^2 \leq \frac{1}{\rho^* \hat{\rho}_n^*} \|x\|_2^2$ , and because we have  $\|x\|_2^2 = p \|x\|^2$ , we get

$$|\Delta_n| \leq \frac{|\rho^* - \hat{\rho}_n^*|}{\rho^* \hat{\rho}_n^*} \left\| n^{\frac{1}{2}} \bar{Y}_n \right\|_2^2 \leq \left| \frac{1}{\rho^*} - \frac{1}{\hat{\rho}_n^*} \right| p \left\| n^{\frac{1}{2}} \bar{Y}_n \right\|^2.$$

**Lemma 7** gives a control of the first term on the right-hand side of this inequality so that it is sufficient to control the second term. Write

$$\left\| n^{\frac{1}{2}} \bar{Y}_n \right\|^2 = \frac{1}{pn} \sum_{j=1}^p \left( \sum_{i=1}^n Y_{i,j} \right)^2 = \overbrace{\frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n Y_{i,j}^2}^{I_1} + \overbrace{\frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j}}^{I_2}$$

Since  $\mathbb{E}(I_1) = \mathbb{E} \left( \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n Y_{i,j}^2 \right) = \sigma^2$ , use Bienaymé-Tchebychev inequality and the independence of the  $Y_i$ 's to get for  $\epsilon > 0$ ,

$$\mathbb{P} \left( \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n Y_{i,j}^2 - \sigma^2 > \frac{\epsilon}{2} \right) \leq \mathbb{P} \left( \left| \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n Y_{i,j}^2 - \sigma^2 \right| > \frac{\epsilon}{2} \right) \leq \frac{4}{\epsilon^2} \text{Var} \left( \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n Y_{i,j}^2 \right) \leq \frac{4}{\epsilon^2} \frac{1}{np^2} \mathbb{E} \left( \left( \sum_{j=1}^p Y_{1,j}^2 \right)^2 \right). \tag{18}$$

Then, by Assumption  $A_2$ , we have  $\mathbb{E} \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j}^4 \right) \leq \sqrt{K_2}$ . Then, by Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} \frac{1}{np^2} \mathbb{E} \left( \left( \sum_{j=1}^p Y_{1,j}^2 \right)^2 \right) &\leq \frac{1}{np} \mathbb{E} \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j}^4 \right) + \frac{1}{np^2} \sum_{j=1}^p \sum_{\substack{k=1 \\ k \neq j}}^p \mathbb{E} \left( Y_{1,j}^2 Y_{1,k}^2 \right) \leq \frac{1}{np} \sqrt{K_2} + \frac{1}{np^2} \sum_{j=1}^p \sum_{\substack{k=1 \\ k \neq j}}^p \sqrt{\mathbb{E} \left( Y_{1,j}^4 \right)} \sqrt{\mathbb{E} \left( Y_{1,k}^4 \right)} \\ &\leq \frac{1}{np} \sqrt{K_2} + \frac{1}{n} \left( \frac{1}{p} \sum_{j=1}^p \sqrt{\mathbb{E} \left( Y_{1,j}^4 \right)} \right)^2 \leq \frac{1}{np} \sqrt{K_2} + \frac{1}{n} \mathbb{E} \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j}^4 \right) \leq \frac{1}{n} \sqrt{K_2} \left( \frac{1}{p} + 1 \right). \end{aligned} \tag{19}$$

Finally, combining inequalities (18), (19), we get the following control for  $I_1$ , for  $\eta > 0$

$$\mathbb{P} \left( I_1 - \mathbb{E}(I_1) > \frac{\eta}{2} \right) \leq \frac{4}{\eta^2} \frac{1}{n} \sqrt{K_2} \left( \frac{1}{p} + 1 \right). \tag{20}$$

Now, we focus on  $I_2$ . Using the independence between the observations  $Y_i$ 's, we have

$$\mathbb{E}(I_2) = \mathbb{E} \left( \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j} \right) = 0. \text{ By Bienaymé-Tchebychev inequality, we have, for } \eta > 0$$

$$\mathbb{P} \left( I_2 > \frac{\eta}{2} \right) \leq \frac{4}{\eta^2} \mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j} \right)^2 \right]. \tag{21}$$

Furthermore, since  $\frac{1}{n} = \frac{(n-1)^2}{4} \left( \frac{2}{n(n-1)} \right)^2$ , we can express the expectation above as the expectation of a U-statistic

$$\mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j} \right)^2 \right] = \frac{(n-1)^2}{4} \mathbb{E} \left[ \left( \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{i' \neq i}^n \frac{1}{p} \sum_{j=1}^p Y_{i,j} Y_{i',j} \right)^2 \right].$$

More precisely, this is a U-statistic of degree 2 with kernel  $w(Y_i, Y_{i'}) = \frac{1}{p} \sum_{j=1}^p Y_{i,j} Y_{i',j}$ , with  $\mathbb{E}[w(Y_i, Y_{i'})] = 0$  and degenerated gradients  $\mathbb{E}[w(Y_i, Y_{i'}) | Y_i] = 0$  and  $\mathbb{E}[w(Y_i, Y_{i'}) | Y_{i'}] = 0$ . Using the expression of the variance of this U-statistic as given in Lee (2019) [21], it follows that

$$\mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j} \right)^2 \right] = \frac{(n-1)^2}{4} \frac{1}{\frac{n(n-1)}{2}} \binom{n-2}{0} \text{Var} \left( w(Y_i, Y_{i'}) \right) = \frac{n-1}{2n} \mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j} Y_{2,j} \right)^2 \right]. \tag{22}$$

Now, we have by independence  $\mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j} Y_{2,j} \right)^2 \right] = \mathbb{E} \left[ \frac{1}{p^2} \sum_{j=1}^p \sum_{k=1}^p Y_{1,j} Y_{2,j} Y_{1,k} Y_{2,k} \right] = \frac{1}{p^2} \sum_{j=1}^p \sum_{k=1}^p \mathbb{E} \left[ (Y_{1,j} Y_{1,k})^2 \right]$ . Recall that  $\mathbb{E} [Y_{1,j} Y_{1,k}] = 0$  if  $j \neq k$ . By using Hölder inequalities repetitively and by Assumption  $A_2$ , we have  $\frac{1}{p} \sum_{j=1}^p \left[ \mathbb{E} \left( Y_{1,j}^2 \right) \right]^2 \leq \left( \frac{1}{p} \sum_{j=1}^p \mathbb{E} \left( Y_{1,j}^8 \right) \right)^{\frac{1}{2}} \leq K_2^{\frac{1}{2}}$ , yielding

$$\mathbb{E} \left[ \left( \frac{1}{p} \sum_{j=1}^p Y_{1,j} Y_{2,j} \right)^2 \right] \leq \frac{1}{p} \sqrt{K_2}. \tag{23}$$

Finally, combining Eqs. (21), (22) and (23), we obtain a control for  $I_2$  as follows

$$\mathbb{P} \left( I_2 > \frac{\eta}{2} \right) = \mathbb{P} \left( \frac{1}{pn} \sum_{j=1}^p \sum_{i=1}^n \sum_{i' \neq i}^n Y_{i,j} Y_{i',j} > \frac{\eta}{2} \right) \leq \frac{1}{\eta^2} \frac{2(n-1)}{pn} \sqrt{K_2}. \tag{24}$$

Finally, Assumption  $A_1$  implies

$$\begin{aligned} \mathbb{P} (|A_n| > \epsilon n) &= \mathbb{P} \left( p \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| \left\| n^{1/2} \bar{Y}_n \right\|^2 > \epsilon n \right) \leq \mathbb{P} \left( \left\| n^{1/2} \bar{Y}_n \right\|^2 \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| > \frac{\epsilon}{K_1} \right) \\ &\leq \mathbb{P} \left( \left( \left\| n^{1/2} \bar{Y}_n \right\|^2 - \sigma^2 \right) \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| > \frac{\epsilon}{2K_1} \right) + \mathbb{P} \left( \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| > \frac{\epsilon}{2\sigma^2 K_1} \right). \end{aligned}$$

Using the fact that  $\mathbb{P}(AB > \epsilon) \leq \mathbb{P}(A > \sqrt{\epsilon}) + \mathbb{P}(B > \sqrt{\epsilon})$ , and the definition of the function  $g_n$  in Lemma 7, we have

$$\begin{aligned} \forall \epsilon > 0, \quad \mathbb{P} (|A_n| > \epsilon n) &\leq \mathbb{P} \left( \left\| n^{1/2} \bar{Y}_n \right\|^2 - \sigma^2 > \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \frac{\epsilon}{2\sigma^2 K_1} \right) \\ &\leq \mathbb{P} \left( I_1 - \sigma^2 > \frac{1}{2} \sqrt{\frac{\epsilon}{2K_1}} \right) + \mathbb{P} \left( I_2 > \frac{1}{2} \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \frac{\epsilon}{2\sigma^2 K_1} \right). \end{aligned}$$

Therefore, by inequalities (20) and (24), considering  $\eta = \sqrt{\frac{\epsilon}{2K_1}}$ , we get for any  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{P} (|A_n| > \epsilon n) &\leq \frac{4}{\left( \sqrt{\frac{\epsilon}{2K_1}} \right)^2} \times \left[ \frac{\sqrt{K_2}}{n} \left( \frac{1}{p} + 1 \right) + \frac{1}{2} \frac{n-1}{n} \frac{\sqrt{K_2}}{p} \right] + g_n \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \frac{\epsilon}{2\sigma^2 K_1} \right) \\ &\leq \frac{4K_1 \sqrt{K_2}}{\epsilon n} \left( 2 + \frac{1}{p} + K_1 \right) + g_n \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \frac{\epsilon}{2\sigma^2 K_1} \right). \end{aligned} \tag{25}$$

We now complete the proof of the theorem by handling the term (II) in inequality (16). By Lemma 7, we get for any  $\epsilon > 0$ ,

$$\mathbb{P} (|\hat{a}_n - a^*| > \epsilon) = \mathbb{P} \left( \left| \frac{1}{\hat{\rho}_n^*} - \frac{1}{\rho^*} \right| > \frac{\epsilon}{K_3} \right) = g_n \left( \frac{\epsilon}{K_3} \right). \tag{26}$$

With inequalities (16), (17), (25), and (26), and using the expression of  $G$  to bound  $g_n$  given in Lemma 7, we obtain for  $\epsilon > 0$ , and  $u \geq 2n$ ,

$$\begin{aligned} \mathbb{P} (n \bar{Z}_n^T \hat{S}_n^{*-1} \bar{Z}_n \geq u (1 + \hat{a}_n^* + 2\epsilon)) &\leq \mathbb{P} (n \bar{Z}_n^T S_n^{*-1} \bar{Z}_n \geq u (1 + a^*)) + \frac{4K_1 \sqrt{K_2}}{\epsilon n} \left( 2 + \frac{1}{p} + K_1 \right) \\ &\quad + g_n \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + g_n \left( \frac{\epsilon}{2\sigma^2 K_1} \right) + g_n \left( \frac{\epsilon}{K_3} \right) \leq \frac{2e^3}{9} \left( \frac{u-n}{2} \right)^{\frac{n}{2}} \frac{e^{-\frac{u-n}{2}}}{\Gamma \left( \frac{n}{2} + 1 \right)} + \frac{1}{n} \frac{C(\epsilon)}{\epsilon}, \end{aligned}$$

where  $C(\epsilon)$  is independent of  $n$  such that

$$\forall \epsilon > 0, \quad C(\epsilon) = 4K_1 \sqrt{K_2} \left( 2 + \frac{1}{p} + K_1 \right) + 2K_1 G \left( \sqrt{\frac{\epsilon}{2K_1}} \right) + \frac{4K_1^2 \sigma^4}{\epsilon} G \left( \frac{\epsilon}{2\sigma^2 K_1} \right) + \frac{K_3^2}{\epsilon} G \left( \frac{\epsilon}{K_3} \right). \quad \square$$

### Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2024.105342>.

### References

[1] M. Abramovitch, I. Stegun, Handbook of Mathematical Tables, National Bureau of Standards, Washington, DC, 1970.  
 [2] J.O. Ajadi, Z. Wang, I.M. Zwetsloot, A review of dispersion control charts for multivariate individual observations, Qual. Eng. 33 (1) (2021) 60–75.  
 [3] Z. Bai, J. Hu, C. Wang, C. Zhang, Test on the linear combinations of covariance matrices in high-dimensional data, Stat. Pap. 62 (2) (2019) 701–719.

- [4] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, *Statist. Sinica* (1996) 311–329.
- [5] J.C.A. Barata, M.S. Hussein, The Moore–Penrose pseudoinverse: A tutorial review of the theory, *Braz. J. Phys.* 42 (2012) 146–165.
- [6] P. Bertail, E. Gautherat, H. Harari-Kermadec, Exponential bounds for multivariate self-normalized sums, *Electron. Commun. Probab.* 13 (2008) 628–640.
- [7] T. Bodnar, A.K. Gupta, N. Parolya, On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix, *J. Multivariate Anal.* 132 (2014) 215–228.
- [8] G. Chasapis, R. Liu, T. Kocoz, Rademacher–Gaussian tail comparison for complex coefficients and related problems, *Proc. Amer. Math. Soc.* 150 (03) (2022) 1339–1349.
- [9] L. Chen, D. Paul, R.L. Prentice, P. Wang, A regularized Hotelling’s  $T^2$  test for pathway analysis in proteomic studies, *J. Amer. Statist. Assoc.* 106 (496) (2011) 1345–1360.
- [10] A.P. Dempster, A high dimensional two sample significance test, *Ann. Math. Stat.* (1958) 995–1010.
- [11] K. Dong, H. Pang, T. Tong, M.G. Genton, Shrinkage-based diagonal Hotelling’s tests for high-dimensional small sample size data, *J. Multivariate Anal.* 143 (2016) 127–142.
- [12] P. Doukhan, J.R. León, Cumulants for stationary mixing random sequences and applications to empirical spectral density, *Probab. Math. Statist.* 10 (1989) 11–26.
- [13] M.L. Eaton, B. Efron, Hotelling’s  $T^2$  test under symmetry conditions, *J. Amer. Statist. Assoc.* 65 (330) (1970) 702–711.
- [14] G.H. Golub, C.F. Van Loan, *Matrix Computations*, third ed., The Johns Hopkins University Press, Baltimore, 1996.
- [15] W. Hoeffding, Probability inequalities for sums of bounded random variables, in: *The Collected Works of Wassily Hoeffding*, Springer, 1994, pp. 409–426.
- [16] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, *Ann. Statist.* 29 (2) (2001) 295–327.
- [17] S. Lahiri, S. Mukhopadhyay, A penalized empirical likelihood method in high dimensions, *Ann. Statist.* 40 (5) (2012) 2511–2540.
- [18] B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection, *Ann. Statist.* 28 (2000) 1302–1338.
- [19] O. Ledoit, M. Wolf, A well conditioned estimator for large dimensional covariance matrices, *J. Multivariate Anal.* 88 (2004) 365–411.
- [20] O. Ledoit, M. Wolf, Quadratic shrinkage for large covariance matrices, *Bernoulli* 28 (3) (2022) 1519–1547.
- [21] A. Lee, *U-Statistics: Theory and Practice*, first ed., Routledge, 1990.
- [22] H. Li, A. Aue, D. Paul, J. Peng, P. Wang, An adaptable generalization of Hotelling’s  $T^2$  test in high dimension, *Ann. Statist.* 48 (3) (2020) 1815–1847.
- [23] F. Merlevède, J. Najim, P. Tian, Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications, *Linear Algebra Appl.* 577 (2019) 317–359.
- [24] X. Mestre, On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices, *IEEE Trans. Signal Process.* 56 (2008) 5353–5368.
- [25] W.K. Newey, R.J. Smith, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* 72 (2017) 219–255.
- [26] D. Panchenko, Symmetrization approach to concentration inequalities for empirical processes, *Ann. Probab.* 31 (2003) 2068–2081.
- [27] I. Pinelis, Extremal probabilistic problems and Hotelling’s  $T^2$  test under a symmetry condition, *Ann. Statist.* 22 (1) (1994) 357–368.
- [28] M. Pourahmadi, High-dimensional covariance estimation: With high-dimensional data, in: *Wiley Series in Probability and Statistics*, John Wiley & Sons Inc, 2013.
- [29] M.S. Srivastava, Multivariate theory for analyzing high dimensional data, *J. Japan Statist. Soc.* 37 (1) (2007) 53–86.
- [30] M.S. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, *J. Multivariate Anal.* 99 (3) (2008) 386–402.
- [31] M. Talagrand, The missing factor in Hoeffding’s inequalities, *Ann. I’HP Probab. Stat.* 31 (4) (1995) 689–702.
- [32] A.N. Tikhonov, On the regularization of ill-posed problems, in: *Doklady Akademii Nauk*, Vol. 153, Russian Academy of Sciences, 1963, pp. 49–52.
- [33] N.D. Tracy, J.C. Young, R.L. Mason, Multivariate control charts for individual observations, *J. Qual. Technol.* 24 (2) (1992) 88–95.
- [34] J.A. Tropp, et al., An introduction to matrix concentration inequalities, *Found. Trends Mach. Learn.* 8 (1–2) (2015) 1–230.
- [35] J. Yao, A. Kammoun, J. Najim, Eigenvalue estimation of parameterized covariance matrices of large dimensional data, *IEEE Trans. Signal Process.* 60 (11) (2012) 5893–5905.