Fluctuations of the diagonal entries of a large sample precision matrix

Nina Dörnemann, Holger Dette

November 2, 2022

Abstract

For a given $p \times n$ data matrix $\mathbf{X}_n$ with i.i.d. centered entries and a population covariance matrix $\mathbf{\Sigma}$, the corresponding sample precision matrix $\hat{\mathbf{\Sigma}}^{-1}$ is defined as the inverse of the sample covariance matrix $\hat{\mathbf{\Sigma}} = (1/n)\mathbf{\Sigma}^{1/2}\mathbf{X}_n\mathbf{X}_n^\top\mathbf{\Sigma}^{1/2}$. We determine the joint distribution of a vector of diagonal entries of the matrix $\hat{\mathbf{\Sigma}}^{-1}$ in the situation, where $p_n = p < n$ and $p/n \to y \in [0,1)$ for $n \to \infty$. Remarkably, our results cover both the case where the dimension is negligible in comparison to the sample size and the case where it is of the same magnitude. Our approach is based on a QR-decomposition of the data matrix, yielding a connection to random quadratic forms and allowing the application of a central limit theorem for martingale difference schemes. Moreover, we discuss an interesting connection to linear spectral statistics of the sample covariance matrix. More precisely, the logarithmic diagonal entry of the sample precision matrix can be interpreted as a difference of two highly dependent linear spectral statistics of $\hat{\mathbf{\Sigma}}$ and a submatrix of $\hat{\mathbf{\Sigma}}$. This difference of spectral statistics fluctuates on a much smaller scale than each single statistic.

Keywords: central limit theorem, random matrix theory, sample precision matrix
AMS subject classification: 60B20, 60F05

1 Introduction

Many statistical problems as they occur in biology or finance demand estimates of the covariance matrix or its inverse, for which the sample precision matrix is a popular choice. Spurred by the groundbreaking advances of data collecting devices, these applications nowadays call for analysis tools of high-dimensional data sets (see, e.g., Fan and Li, 2006; Johnstone, 2006, and references therein). Moreover, they motivate the investigation of the probabilistic properties of large sample covariance or precision matrices, where the dimension of the data and the sample size are of the same order. In the last decades, the scientific interest was mainly focused on the probabilistic properties of the spectrum of the sample covariance matrix. Since the pioneering work of Marčenko and Pastur (1967) on the empirical spectral distribution of $\hat{\mathbf{\Sigma}}$ for the case $p/n \to y \in (0, \infty)$, the asymptotic behavior of its eigenvalues and eigenvalue statistics has been studied by numerous authors. For example, we mention the works of Bai and Yin (1988) on the limiting spectral distribution in the case $y = 0$, Jonsson (1982), Bai and Silverstein (2004), Zheng et al. (2015b), Najim and Yao (2016) on linear spectral statistics, Baik and Silverstein (2006) on the eigenvalues of spiked population models, and of Johnstone (2001), Bai and Yin (2008) on the extreme eigenvalues of $\hat{\mathbf{\Sigma}}$, to name just a few.

From a statistical point of view the sample precision matrix plays a vital role in the analysis of high-dimensional linear models. In particular, the diagonal elements of the matrix $\hat{\mathbf{\Sigma}}^{-1}$ are proportional to the conditional variances of the least squares estimator of the individual coefficients in the linear model (provided that the errors are independent and homoscedastic and there is no intercept in the model). Despite its importance, the literature on the probabilistic properties of the spectrum or the diagonal elements of the sample precision matrix in the high-dimensional paradigm is more scarce. Using techniques from random matrix theory, Zheng et al. (2015a) established a central limit theorem for linear spectral statistics of a rescaled version of the sample precision matrix. In the case where the dimension exceeds the sample size, Bodnar et al. (2016) investigated the asymptotic properties of linear spectral statistics of the Moore-Penrose inverse of the sample covariance matrix.

The analysis of the fluctuation of the diagonal entries of the sample precision matrix is very challenging as the structure of the inverse is subtle, which demands sophisticated tools for the analysis of its diagonal entries as there exists no approachable representation of the entries of the sample precision matrix as a function of the data. Under the additional assumption of a multivariate normal distribution, the exact distribution of $(\hat{\mathbf{\Sigma}}^{-1})_{qr}$ is well-understood for fixed dimension and sample size ($1 \leq q, r \leq p$). In fact, $n^{-1}\hat{\mathbf{\Sigma}}^{-1}$ follows an inverse Wishart distribution (see Gupta and Nagar, 2018; Nydick, 2012; Von Rosen, 1988, for more details). Apart from this, the asymptotic properties of $(\hat{\mathbf{\Sigma}}^{-1})_{qq}$ for non-normal distributed data and a dimension growing with the sample sizes are not well understood so far.

We add to this line of research by establishing a central limit theorem for the diagonal entries of a large sample precision matrix. Our approach is based on the following consequence of Cramer's rule

$$(\hat{\mathbf{\Sigma}}^{-1})_{qq} = \frac{|\hat{\mathbf{\Sigma}}^{(-q)}|}{|\hat{\mathbf{\Sigma}}|}, \; 1 \leq q \leq p,$$

where $\hat{\mathbf{\Sigma}}^{(-q)}$ denotes the $(p-1) \times (p-1)$ submatrix of $\hat{\mathbf{\Sigma}}$ with the qth row and qth column being deleted. This representation reveals an explicit connection to a random quadratic form, which is shown to satisfy a central limit theorem for martingale difference schemes. Moreover, we also observe an immediate connection to linear spectral statistics of sample covariance matrices: the logarithm of the qth diagonal entry

$$\log(\hat{\mathbf{\Sigma}}^{-1})_{qq} = \log|\hat{\mathbf{\Sigma}}^{(-q)}| - \log|\hat{\mathbf{\Sigma}}| \tag{1.1}$$

is a difference of two linear spectral statistics of $\hat{\mathbf{\Sigma}}$ and its submatrix $\hat{\mathbf{\Sigma}}^{(-q)} \in \mathbb{R}^{(p-1)\times(p-1)}$. However, due to the strong dependence between the eigenvalues of $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Sigma}}^{(-q)}$, the asymptotic behavior of (1.1) cannot be described by the meanwhile classical CLT of Bai and Silverstein (2004) or one of the many follow-up works. Interestingly, as we will demonstrate below, the difference in (1.1) fluctuates on a scale $1/\sqrt{n}$ which is of significantly smaller order than the fluctuations of each single linear spectral statistic $\log|\hat{\mathbf{\Sigma}}|$ and $\log|\hat{\mathbf{\Sigma}}^{(-q)}|$. More precisely, after appropriate normalization, a finite-dimensional vector of diagonal entries follows a multivariate normal distribution. Similarly to linear spectral statistics of the sample covariance matrix, the limiting variance of $(\hat{\mathbf{\Sigma}}^{-1})_{qq}$ is determined by the fourth moment of the underlying data generating distribution.

The work most similar in spirit to ours is Cipolloni and Erdős (2020) who considered linear spectral statistics of the sample covariance matrix and its minor from i.i.d. data with finite moments of any order. Choosing the function $\log(x)$ in their main result and combining this with (1.1) and the delta method gives a CLT for a single diagonal entry of the sample precision matrix. In contrast to the work of these authors, our approach requires only the existence of the fourth moment and also allows a proof of the weak convergence of a vector of diagonal entries of the precision matrix.

The remaining part of this paper is organized as follows. A CLT for a single diagonal entry is given in Section 2 and is afterwards generalized to the joint convergence of several diagonal entries. All proofs of our main results are provided in Section 3 and Section 4. In Section 5, we give an outlook to future work concerning the sample precision matrix. Finally, Section A in the Appendix sheds some light on the QR-decomposition of the data matrix, which is an important tool used in the proofs.

2 A CLT for diagonal entries of the empirical precision matrix

Throughout this paper, let

$$\mathbf{X}_n = (x_{ij})_{\substack{i=1,\dots,p \\ j=1,\dots,n}} \in \mathbb{R}^{p\times n} \tag{2.1}$$

denote a random $p \times n$ matrix with i.i.d. centered entries having a continuous distribution, $\mathbf{\Sigma} = \mathbf{\Sigma}_n \in \mathbb{R}^{p\times p}$ nonrandom and (symmetric) positive definite matrix with symmetric square root $\mathbf{\Sigma}^{1/2}$. The matrix $\mathbf{\Sigma}$ denotes the population covariance matrix and for most of the following results, it is assumed to be a diagonal matrix (except for the normal case). We denote the sample covariance matrix by

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}_n\mathbf{X}_n^\top\mathbf{\Sigma}^{\frac{1}{2}} \in \mathbb{R}^{p\times p}.$$

If $p < n$, the inverse matrix $\hat{\mathbf{\Sigma}}^{-1}$ is almost surely well-defined and called the sample precision matrix. We are now in the position to formulate the first main result of this section.

Theorem 2.1 (CLT for diagonal entries of full-sample precision matrix)**.** *Let $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ be a diagonal matrix with positive diagonal entries. Assume that the random variables $\{x_{ij} \mid 1 \le i \le p,\ 1 \le j \le n\}$ in* (2.1) *are i.i.d. with continuous distribution, $\mathbb{E}[x_{11}] = 0,\ \operatorname{Var}(x_{11}) = 1$ and $1 < \mathbb{E}[x_{11}^4] = \nu_4 < \infty$. Then, it holds for $n \to \infty, p/n \to y \in [0, 1)$ and $q \in \{1, \dots, p\}$*

$$\frac{\sqrt{n-p+1}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}}\left(\frac{n-p+1}{n}\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq} - \left(\mathbf{\Sigma}^{-1}\right)_{qq}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho),\ n \to \infty,$$

where the asymptotic variance is given by $\rho = 2 + (\nu_4 - 3)(1 - y)$.

The proofs of this and of all other results in this paper are deferred to Section 3 and 4. At this point, we only sketch the main arguments for the proof of Theorem 2.1. We use a QR-decomposition of the data matrix to derive a representation of the diagonal entry as the inverse of a quadratic form. With this knowledge at hand, we prove a CLT for this quadratic form by an application of a central limit theorem for martingale difference schemes. By the delta method, we finally get asymptotic normality for $(\hat{\mathbf{\Sigma}}^{-1})_{qq}$ being its inverse. Note that QR-decompositions appear in other contexts in random matrix theory. For example, Wang et al. (2018) used this tool to derive the logarithmic law of the determinant of the sample covariance matrix for the case $p/n \to 1$, while Heiny and Parolya (2021) recently investigated the log-determinant of the sample correlation matrix under an infinite fourth moment. We also refer to Nguyen and Vu (2014) and Bao et al. (2015), who used the QR-decomposition to provide proofs of Girko's logarithmic law for a general random matrix with independent entries.

Remark 2.1.

1. Remarkably, our result also covers the low-dimensional case $y = 0$, where the dimension is negligible in comparison to the sample size. In this case, we may formulate the statement of Theorem 2.1 as

$$\frac{\sqrt{n}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}}\left(\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq} - \left(\mathbf{\Sigma}^{-1}\right)_{qq}\right) \xrightarrow{\mathcal{D}} W \sim \mathcal{N}(0, \nu_4 - 1),\ n \to \infty.$$

2. By the representation (1.1), the statistic $\log((\mathbf{\Sigma}^{-1})_{qq})$ can be interpreted as a difference of two linear spectral statistics of sample covariance matrices and a CLT for this random variable would yield a CLT for $(\mathbf{\Sigma}^{-1})_{qq}$ via the delta method. Recently, Cipolloni and Erdős (2020) considered the case $\mathbf{\Sigma} = \mathbf{I}$ and developed a CLT for the difference of linear spectral

statistics of a sample covariance matrix and its minor, which is applicable to a standardized and centered version of $\log(\boldsymbol{\Sigma}^{-1})_{qq}$. Their result requires i.i.d. entries x_{ij} with finite moments of all order, while we only assume a finite fourth moment in Theorem 2.1. Moreover, in comparison to Theorem 2.1, their asymptotic regime does not include the critical case $p/n \to 0$. Note that Cipolloni and Erdős (2020) do not assume the existence of the limit y of p/n. We only need this assumption to determine the limiting variance ρ, but it is not necessary for proving a CLT as in Theorem 2.1. One could instead normalize by a factor $1/\sqrt{\rho_n}$ defined in equation (3.7) in the proof of Theorem 2.1. We also emphasize that the techniques used for proving Theorem 2.1 sets us in the position to investigate the joint convergence of several diagonal elements of the sample precision matrix given in Theorem 2.2 below.

The variance and mean structure of the limiting distribution of linear spectral statistics of sample covariance matrices are usually expressed via contour integrals and depend on the limiting spectral distribution of $\boldsymbol{\Sigma}$ in a subtle way (see Bai and Silverstein, 2004; Najim and Yao, 2016; Pan and Zhou, 2008). So far, an explicit expression for these quantities has only been found in the null case $\boldsymbol{\Sigma} = \mathbf{I}$, and even for diagonal matrices as considered in Theorem 2.1, explicit expressions are out of reach. In this case, despite its close connection to these kinds of linear spectral statistics, the corresponding quantities of a diagonal entry $(\hat{\boldsymbol{\Sigma}}^{-1})_{qq}$ depend asymptotically on its population version $(\boldsymbol{\Sigma}^{-1})_{qq}$ in an explicit form. In particular, for $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\Sigma})$, the asymptotic mean and variance of a scaled diagonal entry $\sqrt{n-p}(\hat{\boldsymbol{\Sigma}}^{-1})_{qq}/(\boldsymbol{\Sigma}^{-1})_{qq}$ do not depend on $\boldsymbol{\Sigma}^{-1}$ anymore. Moreover, the following corollary, which is a direct consequence of Theorem 2.1 and Lemma 3.2 in Section 3.1, shows that these statements are correct for general population covariance matrices when imposing a normal assumption on the data.

Corollary 2.1. *Let $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ be a symmetric positive definite matrix and assume that the random variables $\{x_{ij} \mid 1 \le i \le p,\ 1 \le j \le n\}$ in (2.1) are i.i.d. with $x_{ij} \sim \mathcal{N}(0,1)$. Then, it holds for $n \to \infty, p/n \to y \in [0,1)$ and $q \in \{1,\dots,p\}$*

$$\frac{\sqrt{n-p+1}}{\left(\boldsymbol{\Sigma}^{-1}\right)_{qq}} \left(\frac{n-p+1}{n} \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{qq} - \left(\boldsymbol{\Sigma}^{-1}\right)_{qq} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2),\ n \to \infty.$$

Our final result of this section provides the joint asymptotic distribution of two diagonal entries and is proven in Section 4.

Theorem 2.2. *Let $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ be a diagonal matrix with positive diagonal entries. Assume that the random variables $\{x_{ij} \mid 1 \le i \le p,\ 1 \le j \le n\}$ in (2.1) are i.i.d. with continuous distribution, $\mathbb{E}[x_{11}] = 0$, $\operatorname{Var}(x_{11}) = 1$ and $1 < \mathbb{E}[x_{11}^4] = \nu_4 < \infty$ for $1 \le i \le p$, $1 \le j \le n$. Then, it holds for $n \to \infty, p/n \to y \in [0,1)$ and $1 \le q_1 \ne q_2 \le p$*

$$\left\{ \frac{\sqrt{n-p+1}}{\left(\boldsymbol{\Sigma}^{-1}\right)_{ii}} \left(\frac{n-p+1}{n} \left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{ii} - \left(\boldsymbol{\Sigma}^{-1}\right)_{ii} \right) \right\}_{i=q_1,q_2}^{\top} \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(\mathbf{0}, \rho \mathbf{I}_2\right),\ n \to \infty,$$

where $\rho = 2 + (\nu_4 - 3)(1 - y)$.

Remark 2.2. Note that Theorem 2.2 provides a nontrivial generalization of Theorem 2.1 since the diagonal entries of the empirical precision matrix are not independent. For more details on the concrete dependence structure, we refer the reader to Lemma 4.1 and 4.2 in Section 4. Moreover, it is notable that these random variables are asymptotically independent, which is a consequence of Theorem 2.1 and 2.2. In general, this property will not be valid beyond the diagonal case, and we can observe a proper dependency between two diagonal entries of the sample precision matrix. In particular, we know for the case of normally distributed data from the properties of the inverse Wishart distribution (see, e.g. Press, 2005; Von Rosen, 1988) that

$$\operatorname{Cov}\left(\sqrt{n-p}\frac{n-p}{n}(\hat{\mathbf{\Sigma}}^{-1})_{q_1,q_1}, \sqrt{n-p}\frac{n-p}{n}(\hat{\mathbf{\Sigma}}^{-1})_{q_2,q_2}\right) = 2(\mathbf{\Sigma}^{-1})_{q_1,q_2} + o(1)$$

for $1 \leq q_1, q_2 \leq p$ and $p/n = \mathcal{O}(1)$.

3 Proof of Theorem 2.1

In order to state the proofs rigorously, we need to introduce further notation. We denote the columns of the random matrix $\mathbf{X}_n$ by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the rows by $\mathbf{b}_1, \ldots, \mathbf{b}_p$, that is, we write

$$\mathbf{X}_n = (x_{ij})_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}} = (\mathbf{b}_1, \ldots, \mathbf{b}_p)^\top = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}. \tag{3.1}$$

In the case $\mathbf{\Sigma} = \mathbf{I}$, we denote the sample covariance matrix by

$$\hat{\mathbf{I}} = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^\top = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top \in \mathbb{R}^{p \times p}.$$

In order to pursue the approach based on Cramer's rule as described in the introduction, we will introduce several submatrices. If we set for some $q \in \{1, \ldots, p\}$

$$\tilde{\mathbf{X}}_n^{(-q)} = (\mathbf{b}_1, \ldots, \mathbf{b}_{q-1}, \mathbf{b}_{q+1}, \ldots, \mathbf{b}_p)^\top \in \mathbb{R}^{(p-1) \times n},$$

then

$$\hat{\mathbf{I}}^{(-q)} = \frac{1}{n}\tilde{\mathbf{X}}_n^{(-q)}\left(\tilde{\mathbf{X}}_n^{(-q)}\right)^\top \in \mathbb{R}^{(p-1) \times (p-1)}$$

can be obtained from $\hat{\mathbf{I}}$ by deleting the qth row and the qth column. Similarly, if we set $\mathbf{Y}_n = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}_n = (\mathbf{d}_1, \ldots, \mathbf{d}_p)^\top \in \mathbb{R}^{p \times n}$ and $\tilde{\mathbf{Y}}_n^{(-q)} = (\mathbf{d}_1, \ldots, \mathbf{d}_{q-1}, \mathbf{d}_{q+1}, \ldots, \mathbf{d}_p)^\top$, we define

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^\top \text{ and } \hat{\mathbf{\Sigma}}^{(-q)} = \frac{1}{n}\tilde{\mathbf{Y}}_n^{(-q)}\left(\tilde{\mathbf{Y}}_n^{(-q)}\right)^\top.$$

Additionally, the matrix $\mathbf{\Sigma}^{(-q)} \in \mathbb{R}^{(p-1)\times(p-1)}$ can be obtained from $\mathbf{\Sigma}$ by deleting the qth row and the qth column.
We continue by proving Theorem 2.1 using a CLT for martingale difference schemes. The auxiliary results for these proofs can be found in Section 3.1.

Proof of Theorem 2.1. Using Lemma 3.1 and noting that the distribution of $\mathbf{X}_n$ is invariant under a permutation of the qth and the pth row, we see that

$$\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}} = \left(\hat{\mathbf{I}}^{-1}\right)_{qq} \overset{\mathcal{D}}{=} \left(\hat{\mathbf{I}}^{-1}\right)_{pp} = \frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{pp}}{\left(\mathbf{\Sigma}^{-1}\right)_{pp}}.$$

Thus, we may assume $q = p$ without loss of generality. From now on, the proof is divided in several steps.

Step 1: QR decomposition

In this step, we rewrite $|\hat{\mathbf{I}}|$ and $|\hat{\mathbf{I}}^{(-p)}|$ in a more handy form via the QR decomposition. More details on this decomposition can be found in Section A.
As explained in detail in Section A, we get by proceeding the QR-decomposition for $\mathbf{X}_n^\top$

$$\mathbf{X}_n^\top = \mathbf{Q}\mathbf{R},\ \mathbf{X}_n = \mathbf{R}^\top\mathbf{Q}^\top, \tag{3.2}$$

where $\mathbf{Q} = (\mathbf{e}_1, \dots, \mathbf{e}_p) \in \mathbb{R}^{n\times p}$ denotes a matrix with orthonormal columns satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$ is an upper triangular matrix with entries $r_{ij} = (\mathbf{e}_i, \mathbf{b}_j)$ for $i \le j$ and $r_{ij} = 0$ for $i > j$, $i, j \in \{1, \dots, p\}$. Note that, since $\left(\tilde{\mathbf{X}}_n^{(-p)}\right)^\top$ is the same as $\mathbf{X}_n^\top$ but with the pth column $\mathbf{b}_p$ removed, we have

$$\left(\tilde{\mathbf{X}}_n^{(-p)}\right)^\top = \mathbf{Q}\tilde{\mathbf{R}},\ \tilde{\mathbf{X}}_n^{(-p)} = \tilde{\mathbf{R}}^\top\mathbf{Q}^\top, \tag{3.3}$$

where $\tilde{\mathbf{R}} = (r_{ij})_{\substack{1\le i\le p,\\ 1\le j\le p-1}} \in \mathbb{R}^{p\times(p-1)}$ and we set $\tilde{\mathbf{R}}^{(-p)} = (r_{ij})_{1\le i,j\le p-1} \in \mathbb{R}^{(p-1)\times(p-1)}$. Using (3.2), we write

$$|\mathbf{X}_n\mathbf{X}_n^\top| = |\mathbf{R}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{R}| = |\mathbf{R}^\top\mathbf{R}| = |\mathbf{R}|^2 = \prod_{i=1}^{p} r_{ii}^2$$

and similarly, by using (3.3) and the Cauchy-Binet formula,

$$\left|\tilde{\mathbf{X}}_n^{(-p)}\left(\tilde{\mathbf{X}}_n^{(-p)}\right)^\top\right| = |\tilde{\mathbf{R}}^\top\tilde{\mathbf{R}}| = |\tilde{\mathbf{R}}^{(-p)}|^2 = \prod_{\substack{i=1,\\ i\ne p}}^{p} r_{ii}^2.$$

Thus, we obtain from Lemma 3.1 and Cramer's rule

$$\left(\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}}\right)^{-1} = \left(\left(\hat{\mathbf{I}}^{-1}\right)_{pp}\right)^{-1} = \frac{|\hat{\mathbf{I}}|}{|\hat{\mathbf{I}}^{(-q)}|} = \frac{1}{n} r_{pp}^2. \tag{3.4}$$

Before continuing with Step 2 of the proof of Theorem 2.1, we visit as an illustrating example the normal case where the distribution of r_{pp}^2 is explicitly known.

Illustration: The normal case

If we assume additionally that $x_{ij} \sim \mathcal{N}(0,1)$ i.i.d. for $i \in \{1,\dots,p\}, j \in \{1,\dots,n\}$, then it is well-known that $r_{pp}^2 \sim \mathcal{X}_{n-p+1}$ (see, e.g., Goodman (1963) or directly use (A.1)), that is,

$$r_{pp}^2 \stackrel{\mathcal{D}}{=} \sum_{j=1}^{n-p+1} Z_j^2,$$

where Z_j are i.i.d. standard normal distributed random variables, $j \in \{1,\dots,n-p+1\}$. Thus, we are able to apply a CLT for r_{pp}^2, namely,

$$\sqrt{n-p+1}\left(\frac{1}{n-p+1} r_{pp}^2 - 1\right) = \frac{1}{\sqrt{n-p+1}} \sum_{j=1}^{n-p+1} (Z_j^2 - 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2).$$

Applying the delta method, we get

$$\sqrt{n-p+1}\left(\frac{n-p+1}{r_{pp}^2} - 1\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2).$$

Thus, using (3.4), we conclude

$$\begin{aligned}
&\frac{\sqrt{n-p+1}}{\left(\mathbf{\Sigma}^{-1}\right)_{pp}}\left(\frac{n-p+1}{n}\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{pp} - \left(\mathbf{\Sigma}^{-1}\right)_{pp}\right) \\
=&\sqrt{n-p+1}\left(\frac{n-p+1}{n}\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{pp}}{\left(\mathbf{\Sigma}^{-1}\right)_{pp}} - 1\right) \\
=&\sqrt{n-p+1}\left(\frac{n-p+1}{r_{pp}^2} - 1\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2).
\end{aligned} \tag{3.5}$$

Note that in the normal case, we have $\nu_4 = 3$. Thus, we have recovered the assertion of Theorem 2.1 in this special case.

Step 2: CLT for quadratic forms

In this step, we will show that the random variable r_{pp}^2 meets the conditions of a CLT for martingale difference schemes. In Section A, it is shown that (see (A.1))

$$r_{pp}^2 = \mathbf{b}_p^\top \mathbf{P}(p-1)\mathbf{b}_p,$$

where $\mathbf{P}(0) = \mathbf{I}_n$ and for $q > 1$

$$\mathbf{P}(q) = \mathbf{I} - \tilde{\mathbf{X}}_{n,q}^\top \left(\tilde{\mathbf{X}}_{n,q}\tilde{\mathbf{X}}_{n,q}^\top\right)^{-1} \tilde{\mathbf{X}}_{n,q} \in \mathbb{R}^{n\times n} \tag{3.6}$$

denotes the projection matrix on the orthogonal complement of the subspace generated by the first q rows of $\mathbf{X}_n$, that is, that is,

$$\tilde{\mathbf{X}}_{n,q} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)^\top \in \mathbb{R}^{q\times n}.$$

Note that the random vector $\mathbf{b}_p$ is defined in (3.1). For the following analysis, we denote $\mathbf{P}(p-1) = \mathbf{P} = (p_{ik})_{1\le i,k\le n}$, which only depends on the random variables $\mathbf{b}_1, \ldots, \mathbf{b}_{p-1}$ and is independent of $\mathbf{b}_p$.
We write

$$\begin{aligned}
\sqrt{\frac{n-p+1}{\rho_n}}\frac{1}{n-p+1}\left(r_{pp}^2 - (n-p+1)\right) =& \frac{1}{\sqrt{\rho_n(n-p+1)}}\left(\mathbf{b}_p^\top \mathbf{P}\mathbf{b_p} - \mathbb{E}_{\mathbf{b}_q}\left[\mathbf{b}_q^\top \mathbf{P}\mathbf{b_p}\right]\right) \\
=& \frac{1}{\sqrt{\rho_n(n-p+1)}}\sum_{i=1}^n Z_{pi},
\end{aligned}$$

where for $i \in \{1, \ldots, n\}$, $n \in \mathbb{N}$

$$\begin{aligned}
Z_{pi} =& 2b_{pi}\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(b_{pi}^2 - \mathbb{E}[b_{pi}^2]\right), \\
\rho_n =& 2 + \frac{\nu_4 - 3}{n-p+1}\sum_{i=1}^n p_{ii}^2.
\end{aligned} \tag{3.7}$$

For $i \in \{1, \ldots, n\}$, let $\mathbb{E}_i$ denote the conditional expectation with respect to the σ-field $\mathcal{F}_{pi}$ generated by $\{\mathbf{b}_1, \ldots, \mathbf{b}_{p-1}\} \cup \{b_{pk} : 1 \le k \le i\}$. Furthermore, $\mathbb{E}_0[X] = \mathbb{E}[X]$ denotes the usual expectation.
Since b_{pk} is measurable with respect to $\mathcal{F}_{p,i-1}$ for $k \in \{1, \ldots, i-1\}$ and b_{pj} is independent of $\mathcal{F}_{p,i-1}$ for $j \in \{i, \ldots, n\}$, and $\mathbf{P}$ is measurable with respect to $\mathcal{F}_{pi}$ for all $i \in \{1, \ldots, n\}$, we obtain

$$\mathbb{E}_{i-1}[Z_{pi}] = 2\sum_{k=1}^{i-1} \mathbb{E}_{i-1}[b_{pi}p_{ki}]b_{pk} + \mathbb{E}_{i-1}\left[p_{ii}\left(b_{pi}^2 - \mathbb{E}[b_{pi}^2]\right)\right]$$

$$=2\mathbb{E}[b_{pi}]\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(\mathbb{E}[b_{pi}^2]-\mathbb{E}[b_{pi}^2]\right) = 0,\ 2\le i\le n.$$

Note that Z_{pi} is measurable with respect to $\mathcal{F}_{pi}$ $(1 \le i \le n)$. These observations imply that for each $n \in \mathbb{N}$, $(Z_{pi})_{1\le i\le n}$ forms a martingale difference sequence with respect to the filtration $(\mathcal{F}_{pi})_{1\le i\le n}$. This representation of a random quadratic form as a martingale difference scheme generalizes the one of Bhansali et al. (2007). Note that we are not able to apply their Theorem 2.1 directly in order to prove asymptotic normality, since in our case $\mathbf{P}$ is a random matrix and the random vectors $\mathbf{b}_p$ vary with $n \in \mathbb{N}$. Thus, we have to give a direct proof showing that it satisfies the conditions of the central limit theorem for martingale difference sequences provided in Lemma 3.3 in Section 3.1. More precisely, we will show that for all $\delta > 0$

$$\sigma_n^2 = \frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}_{i-1}[Z_{pi}^2] \xrightarrow{\mathbb{P}} 1, \tag{3.8}$$

$$r_n(\delta) = \frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}\left[Z_{pi}^2 I_{\{|Z_{pi}|\ge\delta\sqrt{(n-p+1)\rho_n}\}}\right] \to 0, \tag{3.9}$$

as $n\to\infty$.
As a preparation for the following steps, we note that

$$\max_{l=1,\ldots,n}\sum_{m=1}^{n} p_{lm}^2 \le ||\mathbf{P}||^2 \le 1, \tag{3.10}$$

$$\operatorname{tr}\left(\mathbf{P}^2\right) = \sum_{i,k=1}^{n} p_{ki}p_{ik} = ||\mathbf{P}||_2^2 = \operatorname{tr}\mathbf{P} = n-p+1, \tag{3.11}$$

where $||\mathbf{P}||$ denotes the spectral norm of $\mathbf{P}$ and $||\mathbf{P}||_2$ denotes the Euclidean norm of $\mathbf{P}$. The first inequality in (3.10) is a well-known estimate for general symmetric matrices and can be shown by choosing the unit vectors for the maximum appearing in the definition of the spectral norm, while the equality in (3.11) follows from the fact that $\mathbf{P}^2 = \mathbf{P}$.

Step 2.1: Calculation of the variance

We begin with a proof of (3.8). For this purpose, we calculate

$$\begin{aligned}
\sigma_n^2 =& \frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}_{i-1}\left[Z_{pi}^2\right]\\
=& \frac{4}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}_{i-1}\left[\left(\sum_{k=1}^{i-1}p_{ki}b_{pk}\right)^2\right]\\
&+\frac{4}{\rho_n(n-p+1)}\sum_{i=1}^{n}\left\{\left(\mathbb{E}\left[b_{pi}^3\right]-\mathbb{E}[b_{pi}]\mathbb{E}\left[b_{pi}^2\right]\right)\sum_{k=1}^{i-1}b_{pk}\mathbb{E}_{i-1}[p_{ki}p_{ii}]\right\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}_{i-1}[p_{ii}^2]\mathbb{E}\Big[b_{pi}^2-\mathbb{E}[b_{pi}^2]\Big]^2\\
=&\frac{4}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}_{i-1}\left[\left(\sum_{k=1}^{i-1}p_{ki}b_{pk}\right)^2\right]+\frac{4\mathbb{E}\left[b_{p1}^3\right]}{\rho_n(n-p+1)}\sum_{i=1}^{n}\left\{\sum_{k=1}^{i-1}b_{pk}p_{ki}p_{ii}\right\}\\
&+\frac{(\nu_4-1)}{\rho_n(n-p+1)}\sum_{i=1}^{n}p_{ii}^2.
\end{aligned}
\tag{3.12}
$$

Here, we used that b_{pk} is measurable with respect to $\mathcal{F}_{pi}$ for $k\in\{1,\dots,i\}$ and b_{pj} is independent of $\mathcal{F}_{pi}$ for $j\in\{i+1,\dots,n\}$, and $\mathbf{P}$ is measurable with respect to $\mathcal{F}_{pi}$ for all $i\in\{1,\dots,n\}$. Moreover, we obtain using (3.11)

$$
\begin{aligned}
1=&\rho_n^{-1}\left(2+\frac{\nu_4-3}{n-p+1}\sum_{i=1}^{n}p_{ii}^2\right)\\
=&\frac{2}{\rho_n(n-p+1)}\sum_{\substack{i,k=1,\\ i\neq k}}^{n}p_{ki}^2+\frac{\nu_4-1}{\rho_n(n-p+1)}\sum_{i=1}^{n}p_{ii}^2\\
=&\frac{4}{\rho_n(n-p+1)}\sum_{i=1}^{n}\sum_{k=1}^{i-1}p_{ki}^2+\frac{\nu_4-1}{\rho_n(n-p+1)}\sum_{i=1}^{n}p_{ii}^2.
\end{aligned}
\tag{3.13}
$$

Denoting $\nu_4=1+\varepsilon$ for some small $\varepsilon>0$, we note that ρ_n is uniformly bounded away from 0, since for all $n\in\mathbb{N}$

$$
\rho_n=2-\frac{2-\varepsilon}{n-p+1}\sum_{i=1}^{n}p_{ii}^2\geq 2-\frac{2-\varepsilon}{n-p+1}\sum_{i=1}^{n}p_{ii}=\varepsilon>0.
\tag{3.14}
$$

In the following, we will show that (3.8) holds true with $\sigma^2=1$. For this purpose, we write using (3.12), (3.13) and (3.14)

$$
\begin{aligned}
|\sigma_n^2-1|\leq&\frac{4}{\rho_n(n-p+1)}\left|\sum_{i=1}^{n}\left(\mathbb{E}_{i-1}\left[\sum_{k=1}^{i-1}p_{ki}b_{pk}\right]^2-\sum_{k=1}^{i-1}p_{ki}^2\right)\right|\\
&+\frac{4\mathbb{E}|b_{p1}|^3}{\rho_n(n-p+1)}\left|\sum_{i=1}^{n}\left\{\sum_{k=1}^{i-1}b_{pk}p_{ki}p_{ii}\right\}\right|\\
\lesssim&\frac{1}{n-p+1}\left(\delta_{n,1}+\delta_{n,2}+\delta_{n,3}\right),
\end{aligned}
\tag{3.15}
$$

where

$$
\begin{aligned}
\delta_{n,1}&=\left|\sum_{i=1}^{n}\sum_{1\leq k<j\leq i-1}p_{ki}p_{ji}b_{pk}b_{pj}\right|,\\
\delta_{n,2}&=\left|\sum_{i=1}^{n}\sum_{k=1}^{i-1}\left(b_{pk}^2-1\right)p_{ki}^2\right|,
\end{aligned}
$$

$$\delta_{n,3} = \left| \sum_{i=1}^{n} \sum_{k=1}^{i-1} b_{pk} p_{ki} p_{ii} \right|.$$

Similarly as in Bhansali et al. (2007), one can show that $\delta_{n,i}/(n-p+1) = o_{\mathbb{P}}(1)$, as $n \to \infty$ for $i \in \{1, 2, 3\}$, by bounding the second moments of $\delta_{n,1}, \delta_{n,2}, \delta_{n,3}$. Exemplarily, we demonstrate this for the term $\delta_{n,3}$. Notice that an application of Lemma 2.1 in Bhansali et al. (2007) and (3.11) yields

$$\left(\sum_{i,i'=1}^{n} \left(\sum_{k=1}^{\min(i,i')-1} p_{ik} p_{i'k} \right)^2 \right)^{\frac{1}{2}} \lesssim \sqrt{n-p+1} ||\mathbf{P}|| \leq \sqrt{n-p+1}. \tag{3.16}$$

Using the Cauchy-Schwarz inequality, (3.16) and (3.11),

$$\begin{aligned} \mathbb{E}[\delta_{n,3}^2] &= \mathbb{E} \left[\sum_{i,i'=1}^{n} p_{ii} p_{i'i'} \sum_{k=1}^{\min(i,i')-1} p_{ki} p_{ki'} \right] \\ &\leq \mathbb{E} \left[\left(\sum_{i}^{n} p_{ii}^2 \right) \left(\sum_{i,i'=1}^{n} \left(\sum_{k=1}^{\min(i,i')-1} p_{ki} p_{ki'} \right)^2 \right)^{\frac{1}{2}} \right] \\ &\lesssim (n-p+1)^{\frac{3}{2}} = o\left((n-p+1)^2 \right), \ n \to \infty. \end{aligned}$$

Proceeding similarly for the remaining terms $\delta_{n,1}$ and $\delta_{n,2}$, we get $\sigma_n^2 = 1 + o_{\mathbb{P}}(1)$ as $n \to \infty$. By an application of Lemma 3.4 given at the end of this section, the normalizing term ρ_n converges in probability towards ρ as $n \to \infty$.

Step 2.2: Verifying the Lindeberg-type condition (3.9)

Using a truncation argument as in Bhansali et al. (2007), it is sufficient to prove (3.9) under the assumption $\mathbb{E}[b_{11}^8] < \infty$. Then, we obtain by using (3.14)

$$r_n(\delta) \leq \frac{1}{(n-p+1)^2 \rho_n^2 \delta^2} \sum_{i=1}^{n} \mathbb{E}\left[Z_{pi}^4 \right] \lesssim J_1 + J_2,$$

where

$$\begin{aligned} J_1 =& \frac{1}{(n-p+1)^2 \delta^2} \sum_{i=1}^{n} \mathbb{E} \left[b_{pi}^4 \left(\sum_{k=1}^{i-1} p_{ki} b_{pk} \right)^4 \right] \\ \lesssim& \frac{1}{(n-p+1)^2 \delta^2} \sum_{i=1}^{n} \mathbb{E} \left[\left(\sum_{j,k=1}^{i-1} p_{ki} p_{ji} b_{pk} b_{pj} \right)^2 \right] \end{aligned}$$

$$
\begin{aligned}
&\lesssim \frac{1}{(n-p+1)\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{k=1}^{i-1}p_{ki}^2 b_{pk}^2\right)^2\right] \\
&\quad + \frac{1}{(n-p+1)\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{\substack{j,k=1\\ j<k}}^{i-1}p_{ki}p_{ji}b_{pk}b_{pj}\right)^2\right],
\end{aligned}
$$

$$
J_2 = \frac{1}{(n-p+1)^2\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[p_{ii}^4\left(b_{pi}^2-\mathbb{E}[b_{pi}^2]\right)^4\right] \lesssim \frac{1}{(n-p+1)\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[p_{ii}^4\right].
$$

This implies using (3.10) and (3.11)

$$
\begin{aligned}
J_1+J_2 &\lesssim \frac{1}{(n-p+1)^2\delta^2}\sum_{i=1}^{n}\left(\sum_{j,k=1}^{i-1}\mathbb{E}[p_{ki}^2p_{ji}^2]+\mathbb{E}[p_{ii}^4]\right) \lesssim \frac{1}{(n-p+1)^2\delta^2}\sum_{i,j,k=1}^{n}\mathbb{E}[p_{ki}^2p_{ji}^2] \\
&\lesssim \frac{1}{(n-p+1)^2\delta^2}\sum_{j,k=1}^{n}\mathbb{E}\left[p_{jk}^2\max_{l=1,\dots,n}\sum_{m=1}^{n}p_{lm}^2\right] \\
&\lesssim \frac{1}{(n-p+1)^2\delta^2}\sum_{j,k=1}^{n}\mathbb{E}[p_{jk}^2] = o(1).
\end{aligned}
$$

Step 3: Conclusion via delta method

In Step 2, we have shown that an appropriately centered and standardized version of r_{pp}^2 satisfies a CLT. By applying the delta method and using (3.4), we conclude that

$$
\begin{aligned}
&\frac{\sqrt{n-p+1}}{\left(\mathbf{\Sigma}^{-1}\right)_{pp}}\left(\frac{n-p+1}{n}\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{pp}-\left(\mathbf{\Sigma}^{-1}\right)_{pp}\right) = \sqrt{n-p+1}\left(\frac{n-p+1}{r_{pp}^2}-1\right) \\
&\xrightarrow{\mathcal{D}} \mathcal{N}(0,\rho),\ \ n\to\infty,
\end{aligned}
$$

which finishes the proof of Theorem 2.1. □

3.1 Auxiliary results

As the following result reveals, the diagonal entries of the sample precision matrix for standardized data are closely connected to those for data with inhomogeneous variances.

Lemma 3.1. *For $1\le q\le p$ and a diagonal matrix $\mathbf{\Sigma}\in\mathbb{R}^{p\times p}$, it holds*

$$
\left(\hat{\mathbf{I}}^{-1}\right)_{qq} = \frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}}.
$$

Proof of Lemma 3.1. Applying Cramer's rule and noting that $|\hat{\mathbf{\Sigma}}| = |\mathbf{\Sigma}||\hat{\mathbf{I}}|$, we get

$$\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}} = \frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|}\frac{|\hat{\mathbf{\Sigma}}^{(-q)}|}{|\mathbf{\Sigma}^{(-q)}|} = \frac{1}{|\hat{\mathbf{I}}|}\frac{|\hat{\mathbf{\Sigma}}^{(-q)}|}{|\mathbf{\Sigma}^{(-q)}|}. \tag{3.17}$$

Let $(\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}$ denote the $(p-1)\times p$ submatrix of $\mathbf{\Sigma}^{1/2}$ where the qth row is deleted. Similarly, $(\mathbf{\Sigma}^{1/2})^{(\cdot,-q)}$ denotes the $p \times (p-1)$ submatrix of $\mathbf{\Sigma}^{1/2}$ where the qth column is deleted. Using these definitions, we see that

$$\hat{\mathbf{\Sigma}}^{(-q)} = (\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\mathbf{X}_n\mathbf{X}_n^\top(\mathbf{\Sigma}^{1/2})^{(\cdot,-q)} = (\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\mathbf{X}_n\left((\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\mathbf{X}_n\right)^\top. \tag{3.18}$$

(In order to enforce (3.18), $\mathbf{\Sigma}$ does not need to be a diagonal matrix.) Since $\mathbf{\Sigma}$ is a diagonal matrix, it holds

$$(\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\mathbf{X}_n = \left(\mathbf{\Sigma}^{(-q)}\right)^{1/2}\tilde{\mathbf{X}}_n^{(-q)},$$

which implies

$$|\hat{\mathbf{\Sigma}}^{(-q)}| = |\mathbf{\Sigma}^{(-q)}||\hat{\mathbf{I}}^{(-q)}|. \tag{3.19}$$

Using (3.17), (3.19) and Cramer's rule again, we obtain

$$\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}} = \frac{|\hat{\mathbf{I}}^{(-q)}|}{|\hat{\mathbf{I}}|} = \left(\hat{\mathbf{I}}^{-1}\right)_{qq}.$$

□

The connection given in Lemma 3.1 can be generalized to the case of dependent coordinates if we assume that the data follows a standard normal distribution.

Lemma 3.2. *If $\mathbf{\Sigma}$ is a general (not necessarily diagonal) $p \times p$ population covariance matrix and $x_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ $(1 \le i \le p,\ 1 \le j \le n)$, then for any $1 \le q \le p$*

$$\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}} \overset{\mathcal{D}}{=} \left(\hat{\mathbf{I}}^{-1}\right)_{qq}.$$

Proof of Lemma 3.2. Recall formula (3.18) from the proof of Lemma 3.1. It follows from our normal assumption that

$$(\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^{(-q)}),\ 1 \le i \le n,$$

where we used that

$$(\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\left((\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}\right)^\top = (\mathbf{\Sigma}^{1/2})^{(-q,\cdot)}(\mathbf{\Sigma}^{1/2})^{(\cdot,-q)} = \mathbf{\Sigma}^{(-q)}.$$

This implies that

$$(\mathbf{\Sigma}^{(-q)})^{1/2}\tilde{\mathbf{X}}_n^{(-q)} \overset{\mathcal{D}}{=} \tilde{\mathbf{Y}}_n^{(-q)}.$$

Using Cramers rule, we get

$$\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{qq}}{\left(\mathbf{\Sigma}^{-1}\right)_{qq}} = \frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|}\frac{|\hat{\mathbf{\Sigma}}^{(-q)}|}{|\mathbf{\Sigma}^{(-q)}|} \overset{\mathcal{D}}{=} \frac{|\hat{\mathbf{I}}^{(-q)}|}{|\hat{\mathbf{I}}|} = \left(\hat{\mathbf{I}}^{-1}\right)_{qq}.$$

The proof of Lemma 3.2 concludes. □

In order to prove asymptotic normality of the quadratic forms appearing in the previous proofs, we make use of the following CLT for martingale difference schemes.

Lemma 3.3 (Theorem 35.12 in Billingsley (1995))**.** *Suppose that for each $n \in \mathbb{N}$, $Z_{n1}, ..., Z_{nr_n}$ form a real martingale difference sequence with respect to the increasing σ-field (F_{nj}) having second moments. If, as $n \to \infty$*

$$\sum_{j=1}^{r_n} \mathbb{E}[Z_{nj}^2 | F_{n,j-1}] \xrightarrow{\mathbb{P}} \sigma^2, \tag{3.20}$$

where $\sigma^2 > 0$, and for each $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathbb{E}[Z_{nj}^2 I_{\{|Z_{nj}|>\varepsilon\}}] \to 0, \tag{3.21}$$

then

$$\sum_{j=1}^{r_n} Z_{nj} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2).$$

We conclude this section by proving the following lemma, which was used in the proof of Theorem 2.1 and provides the limiting variance.

Lemma 3.4. *It holds*

$$\rho_n \xrightarrow{\mathbb{P}} \rho, \;\; n \to \infty,$$

where ρ is defined in Theorem 2.1 and ρ_n in (3.7)*.*

Proof of Lemma 3.4. Assume that $y = 0$. For this case, we note that

$$\frac{1}{n}\sum_{i=1}^{n} p_{ii}^2 = \frac{1}{n}\sum_{i=1}^{n}(1-p_{ii})^2 - 1 + \frac{2}{n}\sum_{i=1}^{n} p_{ii} = \frac{2(n-p+1)}{n} - 1 + o_{\mathbb{P}}(1)$$
$$= 1 + o_{\mathbb{P}}(1),\ n \to \infty, \tag{3.22}$$

where we used

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(1-p_{ii})^2 \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[1-p_{ii}] = \frac{1}{n}\operatorname{tr}(\mathbf{I}-\mathbf{P}) = \frac{p-1}{n} = o(1),\ n \to \infty.$$

Then, (3.22) implies

$$\rho_n = 2 + \frac{(\nu_4-3)n}{n-p+1} + o_{\mathbb{P}}(1) = \nu_4 - 1 = \rho.$$

Let $y \in (0,1)$. Then we have from Theorem 3.2 in Anatolyev and Yaskov (2017)

$$\frac{1}{n}\sum_{i=1}^{n}(1-p_{ii}-y)^2 \xrightarrow{\mathbb{P}} 0,\ n \to \infty,$$

which implies

$$\frac{1}{n}\sum_{i=1}^{n} p_{ii}^2 = \frac{1}{n}\sum_{i=1}^{n}(1-p_{ii}-y)^2 - (1-y)^2 + \frac{2(1-y)}{n}\sum_{i=1}^{n} p_{ii}$$
$$= \frac{2(1-y)(n-p+1)}{n} - (1-y)^2 + o_{\mathbb{P}}(1) = (1-y)^2 + o_{\mathbb{P}}(1),\ n \to \infty.$$

We conclude for $n \to \infty$

$$\rho_n = 2 + \frac{(\nu_4-3)(1-y)^2 n}{n-p+1} + o_{\mathbb{P}}(1) = 2 + (\nu_4-3)(1-y) + o_{\mathbb{P}}(1) = \rho + o_{\mathbb{P}}(1).$$

□

4 Proof of Theorem 2.2

Proof of Theorem 2.2. Since the distribution of $\hat{\mathbf{I}}^{-1}$ is invariant under interchanging rows of $\mathbf{X}_n$, we have using Lemma 3.1

$$\left(\frac{\left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{q_1,q_1}}{\left(\boldsymbol{\Sigma}^{-1}\right)_{q_1,q_1}}, \frac{\left(\hat{\boldsymbol{\Sigma}}^{-1}\right)_{q_2,q_2}}{\left(\boldsymbol{\Sigma}^{-1}\right)_{q_2,q_2}}\right) = \left(\left(\hat{\mathbf{I}}^{-1}\right)_{q_1,q_1}, \left(\hat{\mathbf{I}}^{-1}\right)_{q_2,q_2}\right) \stackrel{\mathcal{D}}{=} \left(\left(\hat{\mathbf{I}}^{-1}\right)_{p-1,p-1}, \left(\hat{\mathbf{I}}^{-1}\right)_{pp}\right)$$
$$= \left(\left(\hat{\mathbf{I}}^{-1}\right)_{p-1,p-1}, \left(\hat{\mathbf{I}}^{-1}\right)_{pp}\right).$$

Thus, we may assume $q_1 = p-1$ and $q = p$ without loss of generality. Similar to the proof of Theorem 2.1, we start by investigating the asymptotic properties of

$$\begin{aligned}
W_n =& \Bigg\{ \frac{1}{\sqrt{n-p+1}} \left(n \left(\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{pp}}{\left(\mathbf{\Sigma}^{-1}\right)_{pp}} \right)^{-1} - (n-p+1) \right), \\
& \frac{1}{\sqrt{n-p+1}} \left(n \left(\frac{\left(\hat{\mathbf{\Sigma}}^{-1}\right)_{p-1,p-1}}{\left(\mathbf{\Sigma}^{-1}\right)_{p-1,p-1}} \right)^{-1} - (n-p+1) \right) \Bigg\}^\top \\
=& \Bigg\{ \frac{1}{\sqrt{n-p+1}} \left(n \left(\hat{\mathbf{I}}^{-1}\right)_{pp}^{-1} - (n-p+1) \right), \\
& \frac{1}{\sqrt{n-p+1}} \left(n \left(\hat{\mathbf{I}}^{-1}\right)_{p-1,p-1}^{-1} - (n-p+1) \right) \Bigg\}^\top \\
=& \frac{1}{\sqrt{n-p+1}} \Bigg\{ \mathbf{b}_p^\top \mathbf{P}(p-1) \mathbf{b}_p - (n-p+1), \\
& \mathbf{b}_{p-1}^\top (\mathbf{P}(p-2) - \mathbf{Q}(p)) \mathbf{b}_{p-1} - (n-p+1) \Bigg\}^\top,
\end{aligned}$$

where we used Lemma 3.1 and Lemma 4.2 and the projection matrix $\mathbf{Q}(p)$ is defined in (4.1). From now on, the proof is divided in several steps.

Approximation and MDS

Note that for any rank-one projection matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ independent of $\mathbf{b}_p$, we have

$$\operatorname{Var}(\mathbf{b}_p^\top \mathbf{Q} \mathbf{b}_p) \lesssim 1 \ \forall n \in \mathbb{N},$$

and consequently, by Slutsky's lemma, it is sufficient to investigate

$$\begin{aligned}
W_n^{(1)} &= \frac{1}{\sqrt{n-p+1}} \left\{ \mathbf{b}_p^\top \mathbf{P}(p-2) \mathbf{b}_p - (n-p+2), \mathbf{b}_{p-1}^\top \mathbf{P}(p-2) \mathbf{b}_{p-1} - (n-p+2) \right\} \\
&= W_n + o_{\mathbb{P}}(1).
\end{aligned}$$

Throughout the rest of this proof, we denote $\mathbf{P}(p-2) = \mathbf{P} = (p_{ij})_{1 \le i,j \le n}$. By an application of the Cramer-Wold device, we note that it is sufficient to prove a one-dimensional central limit theorem for

$$W_n^{(2)} = \frac{1}{\sqrt{n-p+1}} \left\{ a \left(\mathbf{b}_p^\top \mathbf{P} \mathbf{b}_p - (n-p+2) \right) + b \left(\mathbf{b}_{p-1}^\top \mathbf{P} \mathbf{b}_{p-1} - (n-p+2) \right) \right\}, \ a, b \in \mathbb{R},$$

in order to ensure that the vector $W_n^{(1)}$ converges to a two-dimensional normal distribution. We write

$$\frac{1}{\sqrt{\rho_n}} W_n^{(2)} = \frac{1}{\sqrt{(n-p+1)\rho_n}} \sum_{i=1}^{n} W_{pi},$$

where

$$\begin{aligned} W_{pi} =& a\left(2b_{pi}\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(b_{pi}^2 - 1\right)\right) + b\left(2b_{p-1,i}\sum_{k=1}^{i-1} p_{ki}b_{p-1,k} + p_{ii}\left(b_{p-1,i}^2 - 1\right)\right), \\ \rho_n =& 2 + \frac{\nu_4 - 3}{n-p+1}\sum_{i=1}^{n} p_{ii}^2. \end{aligned}$$

For $p \in \mathbb{N}$, $1 \le i \le n$, let $\mathcal{A}_{pi}$ denote the σ field generated by $\{\mathbf{b}_1, \dots, \mathbf{b}_{p-2}\} \cup \{b_{pk}, b_{p-1,k} : 1 \le k \le i\}$. Similar to in the proof of Theorem 2.1, one can show that $(W_{pi})_{1\le i\le n}$ forms a martingale difference sequence with respect to the σ-fields $(\mathcal{A}_{pi})_{1\le i\le n}$ for each $p \in \mathbb{N}$. In order to apply the central limit theorem given in Lemma 3.3, we need to verify the conditions (3.20) and (3.21).

Calculation of the variance

We begin with a proof of condition (3.20). Note that

$$\begin{aligned} &\frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}[W_{pi}^2|\mathcal{A}_{i-1}] \\ =&\frac{a^2}{\rho_n(n-p+1)}\mathbb{E}\left[\left(2b_{pi}\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(b_{pi}^2 - 1\right)\right)^2 \Bigg|\mathcal{A}_{i-1}\right] \\ &+\frac{b^2}{\rho_n(n-p+1)}\mathbb{E}\left[\left(2b_{p-1,i}\sum_{k=1}^{i-1} p_{ki}b_{p-1,k} + p_{ii}\left(b_{p-1,i}^2 - 1\right)\right)^2 \Bigg|\mathcal{A}_{i-1}\right] \\ &+\frac{2ab}{\rho_n(n-p+1)}\mathbb{E}\Bigg[\left(2b_{p-1,i}\sum_{k=1}^{i-1} p_{ki}b_{p-1,k} + p_{ii}\left(b_{p-1,i}^2 - 1\right)\right) \\ &\times\left(2b_{pi}\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(b_{pi}^2 - 1\right)\right)\Bigg|\mathcal{A}_{i-1}\Bigg] \\ =&\frac{a^2}{\rho_n(n-p+1)}\mathbb{E}\left[\left(2b_{pi}\sum_{k=1}^{i-1} p_{ki}b_{pk} + p_{ii}\left(b_{pi}^2 - 1\right)\right)^2 \Bigg|\mathcal{A}_{i-1}\right] \\ &+\frac{b^2}{\rho_n(n-p+1)}\mathbb{E}\left[\left(2b_{p-1,i}\sum_{k=1}^{i-1} p_{ki}b_{p-1,k} + p_{ii}\left(b_{p-1,i}^2 - 1\right)\right)^2 \Bigg|\mathcal{A}_{i-1}\right] \\ =&a^2 + b^2 + o_{\mathbb{P}}(1),\ n\to\infty, \end{aligned}$$

where we used (3.8) from the proof of Theorem 2.1.

Verification of the Lindeberg-type condition

For a proof of condition (3.21), we use the results from Step 2.2 in the proof of Theorem 2.1 and obtain

$$
\begin{aligned}
&\frac{1}{\rho_n(n-p+1)}\sum_{i=1}^{n}\mathbb{E}\left[W_{pi}^2 I_{\{|W_{pi}|\geq\delta\sqrt{(n-p+1)\rho_n}\}}\right] \leq \frac{1}{(n-p+1)^2\rho_n^2\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[W_{pi}^4\right]\\
\lesssim&\frac{a^4}{(n-p+1)^2\rho_n^2\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(2b_{pi}\sum_{k=1}^{i-1}p_{ki}b_{pk}+p_{ii}\left(b_{pi}^2-1\right)\right)^4\right]\\
+&\frac{b^4}{(n-p+1)^2\rho_n^2\delta^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(2b_{p-1,i}\sum_{k=1}^{i-1}p_{ki}b_{p-1,k}+p_{ii}\left(b_{p-1,i}^2-1\right)\right)^4\right] = o(1),\ n\to\infty.
\end{aligned}
$$

$\square$

Conclusion via delta method

Summarizing the steps above, we obtain from Lemma 3.3

$$
W_n \xrightarrow{\mathcal{D}} \mathcal{N}_2(\mathbf{0},\rho\mathbf{I}_2), n\to\infty.
$$

By an application of the multivariate delta method, we have

$$
\begin{aligned}
(Z_{n,p},Z_{n,p-1})^\top =&\left\{\sqrt{n-p+1}\left(\frac{n-p+1}{\mathbf{b}_p^\top\mathbf{P}(p-1)\mathbf{b}_p}-1\right),\right.\\
&\left.\sqrt{n-p+1}\left(\frac{n-p+1}{\mathbf{b}_{p-1}^\top(\mathbf{P}(p-2)-\mathbf{Q}(p))\mathbf{b}_{p-1}}-1\right)\right\}^\top\\
\xrightarrow{\mathcal{D}}&\mathcal{N}_2(\mathbf{0},\rho\mathbf{I}_2), n\to\infty.
\end{aligned}
$$

4.1 Auxiliary results

The following lemma gives a concrete representation for any diagonal element of the sample precision matrix in terms of the entries of the triangular matrix $\mathbf{R}$.

Lemma 4.1. *For $1\leq q\leq p$, it holds*

$$
n\left(\hat{\mathbf{I}}^{-1}\right)_{qq}^{-1} = r_{qq}^2\prod_{i=q+1}^{p}\frac{r_{ii}^2}{r_{ii,q}^2},
$$

where the matrix $\mathbf{R}$ *is defined in the proof of Theorem 2.1 and,*

$$r_{ii,q}^2 = \mathbf{b}_i^\top \mathbf{P}(i-1,q)\mathbf{b}_i, \; 1 \le i \ne q \le p.$$

Here, $\mathbf{P}(i-1,q)$ *denotes the projection matrix on the orthogonal complement of span(*$\{\mathbf{b}_1,\dots,\mathbf{b}_{i-1}\}\setminus\{\mathbf{b}_q\}$ *). In particular, if* $q=p-1$*, we obtain*

$$n\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{p-1,p-1} = \frac{r_{pp}^2 r_{p-1,p-1}^2}{\mathbf{b}_p^\top \mathbf{P}(p-2)\mathbf{b}_p}.$$

Proof of Lemma 4.1. Recall the QR-decomposition of $\mathbf{X}_n^\top$ given in Section A and the resulting formula

$$|\mathbf{X}_n\mathbf{X}_n^\top| = \prod_{i=1}^p r_{ii}^2.$$

Note that the first $(q-1)$ step in the QR-decomposition of the matrices $\tilde{\mathbf{X}}_n^\top = (\tilde{\mathbf{X}}_n^{(-q)})^\top$ and $\mathbf{X}_n^\top$ coincide, which implies

$$|\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^\top| = \prod_{i=1}^{q-1} r_{ii}^2 \prod_{i=q+1}^{p} r_{ii,q}^2.$$

Combining these formulas with Cramer's rule, we conclude

$$n\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{qq} = \frac{|\mathbf{X}_n\mathbf{X}_n^\top|}{|\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^\top|} = r_{qq}^2 \prod_{i=q+1}^{p} \frac{r_{ii}^2}{r_{ii,q}^2}.$$

□

Recall from the proof of Theorem 2.1 (or see Section A for more details) that

$$\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{pp} = \frac{1}{n} r_{pp}^2 = \frac{1}{n}\mathbf{b}_p^\top \mathbf{P}(p-1)\mathbf{b}_p,$$

while it follows from the fact the entries x_{ij} of the matrix $\mathbf{X}_n$ are i.i.d. random variables that

$$\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{qq} \stackrel{\mathcal{D}}{=} \left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{pp}, \; 1 \le q \le p.$$

These quantities can also be written as a quadratic form, but its concrete structure is unknown so far. The next lemma provides such a representation and specifies the dependency structure between two diagonal elements. Moreover, it helps us to understand the dependence structure between two diagonal entries and, thus, is crucial for proving Theorem 2.2. For convenience, we restrict ourselves to the case $q = p-1$.

Lemma 4.2. *It holds*

$$n\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{p-1,p-1} = \mathbf{b}_{p-1}^\top \left(\mathbf{P}(p-2) - \mathbf{Q}(p)\right)\mathbf{b}_{p-1},$$

where $\mathbf{P}(p-2) - \mathbf{Q}(p)$ is a projection matrix of rank $n-p+1$ and independent of $\mathbf{b}_{p-1}$. More precisely, $\mathbf{Q}(p)$ denotes the matrix corresponding to the projection to $\mathbf{P}(p-2)\mathbf{b}_p$, that is,

$$\mathbf{Q}(p) = \frac{\mathbf{P}(p-2)\mathbf{b}_p\mathbf{b}_p^\top\mathbf{P}(p-2)}{\mathbf{b}_p^\top\mathbf{P}(p-2)\mathbf{b}_p}. \tag{4.1}$$

Proof of Lemma 4.2. Recall from Lemma 4.1 that

$$n\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{p-1,p-1} = \frac{r_{pp}^2 r_{p-1,p-1}^2}{\mathbf{b}_p^\top\mathbf{P}(p-2)\mathbf{b}_p}.$$

Note that $\mathbf{P}(p-1)\mathbf{b}_p = \mathbf{P}(p-2)\mathbf{b}_p - \mathrm{proj}_{e_{p-1}}(\mathbf{b}_p)$, where the projection of a vector $\mathbf{a} \in \mathbb{R}^n$ to a vector $\mathbf{e} \in \mathbb{R}^n$ is given by

$$\mathrm{proj}_{\mathbf{e}}(\mathbf{a}) = \frac{(\mathbf{e}, \mathbf{a})}{(\mathbf{e}, \mathbf{e})}\mathbf{e}$$

and (for details, see Section A)

$$\mathbf{u}_{p-1} = \mathbf{P}(p-2)\mathbf{b}_{p-1},\ \mathbf{e}_{p-1} = \frac{\mathbf{u}_{p-1}}{||\mathbf{u}_{p-1}||_2}.$$

Thus, we obtain

$$\begin{aligned} n\left(\hat{\mathbf{I}}^{-1}\right)^{-1}_{p-1,p-1} =& \mathbf{b}_{p-1}^\top\mathbf{P}(p-2)\mathbf{b}_{p-1}\left(1 - \frac{\mathbf{b}_p^\top \,\mathrm{proj}_{\mathbf{e}_{p-1}}(\mathbf{b}_p)}{\mathbf{b}_p^\top\mathbf{P}(p-2)\mathbf{b}_p}\right) \\ =& \mathbf{b}_{p-1}^\top\mathbf{P}(p-2)\mathbf{b}_{p-1}\left(1 - \mathbf{b}_p^\top \frac{(\mathbf{u}_{p-1}, \mathbf{b}_p)}{(\mathbf{u}_{p-1}, \mathbf{u}_{p-1})\mathbf{b}_p^\top\mathbf{P}(p-2)\mathbf{b}_p}\mathbf{u}_{p-1}\right) \\ =& \mathbf{b}_{p-1}^\top\mathbf{P}(p-2)\mathbf{b}_{p-1} - \frac{\mathbf{b}_p^\top(\mathbf{u}_{p-1}, \mathbf{b}_p)\mathbf{u}_{p-1}}{\mathbf{b}_p^\top\mathbf{P}(p-2)\mathbf{b}_p} \\ =& \mathbf{b}_{p-1}^\top\mathbf{P}(p-2)\mathbf{b}_{p-1} - \mathbf{b}_{p-1}^\top\mathbf{Q}(p)\mathbf{b}_{p-1}. \end{aligned}$$

Note that $\mathbf{Q}(p)^2 = \mathbf{Q}(p)$ and $\mathbf{P}(p-2)\mathbf{Q}(p) = \mathbf{Q}(p)\mathbf{P}(p-2) = \mathbf{Q}(p)$. Consequently, we obtain

$$\begin{aligned} \left(\mathbf{P}(p-2) - \mathbf{Q}(p)\right)^2 =& \mathbf{P}(p-2)^2 + \mathbf{Q}(p)^2 - \mathbf{P}(p-2)\mathbf{Q}(p) - \mathbf{Q}(p)\mathbf{P}(p-2) \\ =& \mathbf{P}(p-2) + \mathbf{Q}(p) - 2\mathbf{Q}(p) = \mathbf{P}(p-2) - \mathbf{Q}(p). \end{aligned}$$

This implies that $\mathbf{P}(p-2)-\mathbf{Q}(p)$ is a projection matrix independent of $\mathbf{b}_{p-1}$ of rank

$$\operatorname{tr}\left(\mathbf{P}(p-2)-\mathbf{Q}(p)\right)=n-p+2-1=n-p+1.$$

$\square$

5 Conclusions

In this paper, we have provided a multivariate central limit theorem for the diagonal entries of a sample precision matrix if the dimension-to-sample-size ratio satisfies $p/n \to y \in [0,1)$ as $n \to \infty$ and the population covariance matrix is diagonal. An important direction of future research is to find the asymptotic distribution of the diagonal entries for a general structure of the population covariance matrix. We emphasize that this question results in a substantially more complicated problem, since the method of the proofs used in this work is tailored to the diagonal case. In particular, we reduce the diagonal case $\mathbf{\Sigma} = \operatorname{diag}(\mathbf{\Sigma})$ to the null case $\mathbf{\Sigma} = \mathbf{I}$. For a general distribution and a general population covariance matrix, this step is no longer correct. Then again, if we conduct a QR-decomposition for $(\mathbf{\Sigma}^{1/2}\mathbf{X}_n)^\top$ instead of $\mathbf{X}_n^\top$ (as in step 1 of the proof of Theorem 2.1), we obtain a quadratic form where the random vectors depend on the projection matrix in an implicit form. Our proofs, especially the martingale argument for applying a CLT, rely crucially on the fact that the random vector $\mathbf{b}_p$ (defined in (3.1)) is independent of the random projection matrix $\mathbf{P}(p-1)$ (defined in (3.6)). Similarly, the techniques used in Cipolloni and Erdős (2020), which can be used to derive a central limit theorem for a single diagonal entry of the sample precision matrix by a representation as a difference of two linear spectral statistics (see Remark 2.1), require the even stronger assumption $\mathbf{\Sigma} = \mathbf{I}$. Additionally, it is not straightforward to adapt the tools provided by Bai and Silverstein (2004) due to the different normalizations appearing in the CLT for a single linear spectral statistic and the difference of two. The development of novel techniques that meet the challenges of the dependent case $\mathbf{\Sigma} \neq \operatorname{diag}(\mathbf{\Sigma})$ will be the objective of our future work.

Acknowledgements. This work was partially supported by the DFG Research unit 5381 *Mathematical Statistics in the Information Age*, project number 460867398. The authors would like to thank Giorgio Cipolloni and László Erdős for some helpful discussions.

References

Anatolyev, S. and Yaskov, P. (2017). Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory*, 33(3):717–738.

Bai, Z. and Silverstein, J. W. (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability*, 32(1):553–605.

Bai, Z. D. and Yin, Y. Q. (1988). Convergence to the semicircle law. *The Annals of Probability*, 16(2):863–875.

Bai, Z.-D. and Yin, Y.-Q. (2008). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. In *Advances In Statistics*, pages 108–127. World Scientific.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408.

Bao, Z., Pan, G., Zhou, W., et al. (2015). The logarithmic law of random determinant. *Bernoulli*, 21(3):1600–1628.

Bhansali, R., Giraitis, L., and Kokoszka, P. (2007). Convergence of quadratic forms with nonvanishing diagonal. *Statistics & probability letters*, 77(7):726–734.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., third edition.

Bodnar, T., Dette, H., and Parolya, N. (2016). Spectral analysis of the moore–penrose inverse of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 148:160–172.

Cipolloni, G. and Erdős, L. (2020). Fluctuations for differences of linear eigenvalue statistics for sample covariance matrices. *Random Matrices: Theory and Applications*, 09(03):2050006.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *Proceedings of the International Congress of Mathematicians, Madrid*, 3.

Goodman, N. (1963). The distribution of the determinant of a complex wishart distributed matrix. *The Annals of Mathematical Statistics*, 34(1):178–180.

Gupta, A. K. and Nagar, D. K. (2018). *Matrix variate distributions*. Chapman and Hall/CRC.

Heiny, J. and Parolya, N. (2021). Log determinant of large correlation matrices under infinite fourth moment. *arXiv preprint arXiv:2112.15388*.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295 – 327.

Johnstone, I. M. (2006). High dimensional statistical inference and random matrices. *Proceedings of the International Congress of Mathematicians, Madrid*.

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1):1–38.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Annals of Applied Probability*, 26(3):1837–1887.

Nguyen, H. H. and Vu, V. (2014). Random matrices: Law of the determinant. *The Annals of Probability*, 42(1):146–167.

Nydick, S. W. (2012). The wishart and inverse wishart distributions. *Electronic Journal of Statistics*, 6(1-19).

Pan, G. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, 18(3):1232–1270.

Press, S. J. (2005). *Applied multivariate analysis: using Bayesian and frequentist methods of inference*. Courier Corporation.

Von Rosen, D. (1988). Moments for the inverted wishart distribution. *Scandinavian Journal of Statistics*, pages 97–109.

Wang, X., Han, X., and Pan, G. (2018). The logarithmic law of sample covariance matrices near singularity. *Bernoulli*, 24(1):80–114.

Zheng, S., Bai, Z., and Yao, J. (2015a). Clt for linear spectral statistics of a rescaled sample precision matrix. *Random Matrices: Theory and Applications*, 4(04):1550014.

Zheng, S., Bai, Z., and Yao, J. (2015b). Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Annals of Statistics*, 43(2):546–591.

A Details on the QR-decomposition of $\mathbf{X}_n^\top$

In this section, we give more details on the QR-decomposition of the matrix $\mathbf{X}_n^\top$ (compare Section 2 in Wang et al., 2018) and provide an explicit representation of the diagonal elements of $\mathbf{R}$ as a quadratic form in the rows of $\mathbf{X}_n$.

To begin with, we describe the QR-decomposition of a general full-column rank matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_p) \in \mathbb{R}^{n\times p}$ by applying the Gram-Schmidt procedure to the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$. Recall the definition of the projection of a vector $\mathbf{a} \in \mathbb{R}^n$ on a vector $\mathbf{e} \in \mathbb{R}^n, \mathbf{e} \neq \mathbf{0}$, is given by

$$\mathrm{proj}_{\mathbf{e}}(\mathbf{a}) = \frac{(\mathbf{e}, \mathbf{a})}{(\mathbf{e}, \mathbf{e})}\mathbf{e}.$$

It holds

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{a}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \\
\mathbf{u}_2 &= \mathbf{a}_2 - \mathrm{proj}_{\mathbf{u}_1} \mathbf{a}_2, & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}, \\
\mathbf{u}_3 &= \mathbf{a}_3 - \mathrm{proj}_{\mathbf{u}_1} \mathbf{a}_3 - \mathrm{proj}_{\mathbf{u}_2} \mathbf{a}_3, & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|}, \\
&\ \vdots & &\ \vdots \\
\mathbf{u}_n &= \mathbf{a}_n - \sum_{j=1}^{n-1} \mathrm{proj}_{\mathbf{u}_j} \mathbf{a}_n, & \mathbf{e}_n &= \frac{\mathbf{u}_n}{\|\mathbf{u}_n\|}.
\end{aligned}$$

Rearranging these equations, we may write $\mathbf{A} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q} = (\mathbf{e}_1, \ldots, \mathbf{e}_p) \in \mathbb{R}^{n\times p}$ denotes a matrix with orthonormal columns satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$ is an upper triangular matrix with entries $r_{ij} = (\mathbf{e}_i, \mathbf{a}_j)$ for $i \leq j$ and $r_{ij} = 0$ for $i > j$, $i, j \in \{1, \ldots, p\}$.

In order to ensure formal correctness of the QR decomposition for the matrix $\mathbf{X}_n^\top = (\mathbf{b}_1, \ldots, \mathbf{b}_p)$, we note that the matrix $\mathbf{X}_n^\top$ has full column rank since we assumed that each x_{ij} follows a continuous distribution for $1 \leq i \leq p,\ 1 \leq j \leq n$. Performing the QR decomposition for the special choice $\mathbf{A} = \mathbf{X}_n^\top = (\mathbf{b}_1, \ldots, \mathbf{b}_p)$, we get

$$\mathbf{X}_n^\top = \mathbf{Q}\mathbf{R},$$

where $\mathbf{Q} = (\mathbf{e}_1, \ldots, \mathbf{e}_p) \in \mathbb{R}^{n\times p}$ denotes a matrix with orthonormal columns satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} \in \mathbb{R}^{p\times p}$ is an upper triangular matrix with entries $r_{ij} = (\mathbf{e}_i, \mathbf{b}_j)$ for $i \leq j$ and $r_{ij} = 0$ for $i > j$, $i, j \in \{1, \ldots, p\}$. Using the definitions $r_{qq}^2 = (\mathbf{e}_i, \mathbf{b}_i)^2$ for $1 \leq q \leq p$ and $\mathbf{P}(0) = \mathbf{I}$, we have

$$r_{11}^2 = (\mathbf{e}_1, \mathbf{b}_1)^2 = ||\mathbf{b}_1||_2^2 = \mathbf{b}_1^\top\mathbf{P}(0)\mathbf{b}_1,$$

and for $2 \leq q \leq p$

$$r_{qq}^2 = (\mathbf{e}_q, \mathbf{b}_q)^2 = \left(\frac{\mathbf{u}_q^\top\mathbf{b}_q}{||\mathbf{u}_q||_2}\right)^2 = \left(\frac{\mathbf{b}_q^\top\mathbf{P}(q-1)\mathbf{b}_q}{||\mathbf{P}(q-1)\mathbf{b}_q||_2}\right)^2 = \mathbf{b}_q^\top\mathbf{P}(q-1)\mathbf{b}_q, \tag{A.1}$$

where the projection matrix $\mathbf{P}(q-1)$ is defined in (3.6) and satisfies $\mathbf{P}(q-1)^2 = \mathbf{P}(q-1)$.