On eigenvalues of a high-dimensional spatial-sign covariance matrix

Weiming Li\textsuperscript{a}, Jianfeng Yao\textsuperscript{b} and Wang Zhou\textsuperscript{c}

\textsuperscript{a}Shanghai University of Finance and Economics, \textsuperscript{b}University of Hong Kong and \textsuperscript{c}National University of Singapore

Abstract

Sample spatial-sign covariance matrix is a much-valued alternative to sample covariance matrix in robust statistics to mitigate influence of outliers. Although this matrix is widely studied in the literature, almost nothing is known on its properties when the number of variables becomes large as compared to the sample size. This paper for the first time investigates the large-dimensional limits of the eigenvalues of a sample spatial sign matrix when both the dimension and the sample size tend to infinity. A first result of the paper establishes that the distribution of the eigenvalues converges to a deterministic limit that belongs to the family of celebrated generalized Marčenko-Pastur distributions. Using tools from random matrix theory, we further establish a new central limit theorem for a general class of linear statistics of these sample eigenvalues. In particular, asymptotic normality is established for sample spectral moments under mild conditions. This theory is established when the population is elliptically distributed. As applications, we first develop two new tools for estimating the population eigenvalue distribution of a large spatial sign covariance matrix, and then for testing the order of such population eigenvalue distribution when these distributions are finite mixtures. Using these inference tools and considering the problem of blind source separation, we are able to show by simulation experiments that in high-dimensional situations, the sample spatial-sign covariance matrix is still a valid and much better alternative to sample covariance matrix when samples contain outliers.

1 Introduction

When a multivariate data set is potentially contaminated by outliers, sample covariance matrix (SCM) becomes less reliable. A wide range of robust alternatives has been proposed in the literature starting from the early M-estimators (Maronna, 1976; Huber, 1977), the minimum volume ellipsoid and minimum determinant estimators (Rousseeuw, 1985), the Stahel-Donoho estimators (Humpel et al., 1986; Donoho and Gasko, 1992) and Tyler’s scatter matrix (Tyler, 1987). These estimators enjoy a high breakdown point and most of them are desirably affine equivariant. For book-length discussions on these classical estimators, we refer to Maronna et al. (2006) and Oja (2010), see also Magyar and Tyler (2014) for a sensible review. However many of these estimators are implicitly defined only, and this lack of an analytically tractable form leads to certain difficulty for their computation and theoretical analysis. Such difficulty is even more pronounced when the number of variables is large. This motivates a growing recent research for more tractable robust scatter estimators that might not be affine equivariant.
A particularly studied estimator is the spatial sign matrix, first introduced in Locantore et al. (1999) and Visuri et al. (2000). The former paper introduces an influential robust principal component analysis based on the spatial sign matrix. Two striking examples in the paper, shown on Figures 14-16 and Figures 15-17, respectively, demonstrate how the SCM leads to a much distorted principal components (PCs) in the presence of a single outlier, and how, at the same time, the spatial sign matrix is able to mitigate the impact of such highly influential outliers. A number of papers have followed since then, especially within the groups around H. Oja and D.E. Tyler, respectively, see Gervini (2008), Sirkia et al. (2009), Taskinen et al. (2010, 2012), Dürr et al. (2014, 2015, 2017) and Dürr and Vogel (2016).

Let us define the sample spatial sign covariance matrix (SSCM). The spatial sign \( s(z) \) of a nonnull \( p \)-dimensional vector \( z \in \mathbb{R}^p \) is \( s(z) = z/\|z\| \), i.e. its projection on the \( p \)-dimensional unit sphere. For completeness, we set \( s(0) = 0 \). This is called a sign because for the univariate case with \( p = 1 \), \( s(z) \in \{+1, -1\} \) are ordinary signs. Given a sample \( x_1, \ldots, x_n \) from a \( p \)-variate population \( x \), the sample SSCM is

\[
C_n = \frac{1}{n} \sum_{j=1}^{n} s(x_j - \hat{\mu})s(x_j - \hat{\mu})'.
\]  

Here \( \hat{\mu} \) is an estimate for the spatial median \( \mu \) of the population which is determined by the equation \( \mathbb{E}[s(x - \mu)] = 0 \) (zero of the mean spatial sign function). The population SSCM is \( \Sigma_x = \mathbb{E}[s(x - \mu)s(x - \mu)'] \). If the data is already centered, one may assume \( \hat{\mu} = \mu = 0 \) and consider the sample SSCM

\[
C_n = \frac{1}{n} \sum_{j=1}^{n} s(x_j)s(x_j)'.
\]  

Despite its simplicity, \( C_n \) is a rich scatter statistic for a multivariate population. It is indeed the exact counterpart of the usual SCM \( S_n = n^{-1} \sum_{j=1}^{n} x_jx_j' \) when one shifts from the Euclidean distance (or \( L^2 \) distance) to the Manhattan block distance (or \( L^1 \) distance) in \( \mathbb{R}^p \).

When the number of variables \( p \) is large compared to the sample size \( n \), the sample SSCM \( C_n \) will likely deviate from the population SSCM \( \Sigma_x \) due to the high-dimensional effect. Indeed for the usual covariance matrix \( \Sigma_0 = \mathbb{E}[xx'] \), such high-dimensional distortion between \( \Sigma_0 \) and the sample covariance matrix \( S_n \) is now well understood with the aid of random matrix theory, see Johnstone (2007) and Paul and Aue (2014). Typically, sample eigenvalues from \( S_n \) have a much wider spread than the population eigenvalues of \( \Sigma_0 \), and this deformation is precisely described by the famous Marčenko-Pastur law. A main result from the current paper shows that for the spatial sign matrix, such high-dimensional distortion again happens when \( p \) is large compared to \( n \).

Such high-dimensional distortion is particularly critical to the robust PCA or robust covariance matrix estimation proposed in Locantore et al. (1999) and Visuri et al. (2000). For example, the procedure (C1)-(C2)-(C3) on page 566 of Visuri et al. (2000) for a robust
estimator of the population covariance matrix works as follows when applied with spatial signs. Given a centered data sample \( x_1, \ldots, x_n \) from a \( p \)-dimensional population:

(C1) Construct eigenvector estimates using the eigenvectors of the sample SSCM \( C_n \), say, matrix \( U \).

(C2) Estimate the marginal variances (eigenvalues, principal values) of \((U'x_1, U'x_2, \ldots, U'x_n)\), using any univariate robust scale estimate (MAD, etc.). Write \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) for the estimates.

(C3) The estimate for \( \Sigma_0 \) is \( S = U\Lambda U' \).

In the high-dimensional context, estimates found in the steps (C1) and (C2) become seriously biased. For (C1), the eigenvectors in \( U \) can be far away from the their population counterpart (of the population SSCM \( \Sigma_x \)), so \textit{a fortiori} far away from the eigenvectors of the population covariance matrix \( \Sigma_0 \). For (C2), the marginal variances of the projections \((U'x_j)_{1 \leq j \leq n}\) could be very different of the eigenvalues of \( \Sigma_0 \). As for the robust PCA proposed in \textit{Locantore et al.} (1999), it will suffer from the same high-dimensional distortion because the procedure also uses the eigenvectors of the sample spatial sign matrix to estimate the eigenvectors of the population covariance matrix, exactly as in Step (C1) above.

Therefore it is necessary to correct such high-dimensional distortion appeared in the sample SSCM in order to preserve its long established attractiveness such as robustness. In this paper, using tools of random matrix theory, we investigate asymptotic spectral behaviors of the sample SSCM \( C_n \) in high-dimensional frameworks where both the dimension \( p \) and the sample size \( n \) tend to infinity. We restrict ourselves to the family of elliptical distributions for the population \( x \) for two reasons. Firstly, if \( x \) is elliptically distributed, the population SSCM \( \Sigma_x \) and the population covariance matrix \( \Sigma_0 \) share same eigenvectors while their respective eigenvalues are in an one-to-one correspondence through a well-known map (\textit{Boente and Fraiman}, 1999). Secondly, high-dimensional study as the one developed in this paper but for a more general population \( x \) seems out of reach at the moment. The first main result of the paper (Theorem 3.1) is an analogue of Marčenko-Pastur law for the limiting distribution of eigenvalues of \( C_n \). This law has been so far known for sample covariance matrices and sample correlation matrices only. The second main result of the paper provides a central limit theorem (CLT) for linear spectral statistics of \( C_n \) (Theorem 3.2). This CLT is the corner-stone for all subsequent applications we develop in the paper. These applications are designed to demonstrate the effectiveness of the theory we develop here for the SSCM \( C_n \).

There have been a few very recent works in the literature that deal with the high-dimensional SSCM \( C_n \) (or its variants), namely \textit{Zou et al.} (2014), \textit{Feng and Sun} (2016), \textit{Li et al.} (2016), and \textit{Chakraborty and Chaudhuri} (2017). A common feature in these papers is that given a specific null hypothesis on the population location or scatter, in a one-sample or two-sample design, the authors have in their disposal a specific test statistic which is an explicit function of \( C_n \) (or its variants). They thus directly study the statistic using traditional asymptotic methods such as projections (as in a U-statistic) or a martingale
decomposition. None of these papers studied the distribution of the eigenvalues of $C_n$ as done in this paper using random matrix theory. Meanwhile, some of these test statistics are indeed linear spectral statistics of $C_n$. Therefore in these cases, the CLT developed in this paper leads to an independent and new proof for these existing results. However this comparison will not be pursued here but in a later separated work.

The remaining of the paper is as follows. Section 2 summarizes some preliminary results from elliptical distributions and related random matrix theory. Section 3 establishes the two main theoretical results of the paper (Theorems 3.1 and 3.2). Application to spectral moments statistics is fully addressed with explicit limiting mean and covariance functions in the corresponding central limit theorem. Then in Section 4, relying on these results, we develop two statistical applications on the spectrum of $\Sigma_x$, the population SSCM, under a setting where the spectrum forms a discrete distribution with finite support. In one application, the spectrum of $\Sigma_x$ is estimated using the method of moments, and in the other application, we test the hypothesis that there are no more than $d_0$ distinct eigenvalues in the spectrum of $\Sigma_x$. In Section 5, we develop two applications of the general theory of Section 3 to robust statistics in the high-dimensional context. Technical proofs of the main theorems are gathered in Section 6. Some useful lemmas and their proofs are postponed to the last section.

## 2 Preliminaries

The family of elliptical distributions is an important extension to the multivariate normal distribution and has been broadly used in data analysis in various fields (Gupta et al., 2013). A random vector $x$ with zero mean is *elliptically distributed* if it has a stochastic representation:

\[ x = wA u, \quad (2.1) \]

where $A$ is a $p \times p$ deterministic and invertible matrix, $w \geq 0$ a scalar random variable representing the scale of $x$, and $u \in \mathbb{R}^p$ is the random direction, independent of $w$ and uniformly distributed on the unit sphere in $\mathbb{R}^p$. Besides the normal distribution, this family includes many other celebrated distributions, such as multivariate $t$-distribution, Kotz-type distributions, and Gaussian scale mixtures. Clearly the population covariance matrix is $\Sigma_0 = \mathbb{E}[xx'] = \mathbb{E}(w^2)T$ where $T = AA'$ is in fact the *shape* matrix of the population. In order to resolve the indeterminacy between the scales of $w$ and $T$, we will use throughout the paper the normalization $\text{tr}(T) = p$.

Let $x_1, \ldots, x_n$ be a sequence of independent and identically distributed (i.i.d.) random vectors from the elliptical population (2.1). We consider the sample SSCM $C_n$ defined in (1.2) and scale it as

\[ B_n = pC_n = \frac{1}{n} \sum_{j=1}^n y_j y_j' = \sqrt{p}s(x_j). \quad (2.2) \]
The reason for this scaling is that now $\text{tr}(B_n) = p$ and the eigenvalues of $B_n$ are of order $O(1)$ in average.

In this paper, using tools of random matrix theory, we investigate limiting properties of the eigenvalues of $B_n$ in a high-dimensional setting. Precisely, both the dimension $p$ and the sample size $n$ tend to infinity with their ratio $p/n \to c$, a positive constant in $(0, \infty)$.

Let $M_p$ be a $p \times p$ matrix with eigenvalues $(\lambda_j)_{1 \leq j \leq p}$. Its empirical spectral distribution (ESD) is by definition the probability measure

$$F^{M_p} = \frac{1}{p} \sum_{j=1}^{p} \delta_{\lambda_j},$$

where $\delta_b$ denotes the Dirac mass at $b$. If this sequence has a limit when $p \to \infty$, the limit is referred as a limiting spectral distribution, or LSD, of the sequence. Our aim is to study the limiting properties of $F^{B_n}$ and CLT for linear spectral statistics (LSS) of the form

$$\int f(x) dF^{B_n}(x)$$

for a class of smooth test functions $f$. These properties may become powerful tools to recover spectral features of the scaled population SSCM, i.e. $\Sigma := p\mathbb{E}(s(x)s(x)') = p\Sigma_x$, and then those of the shape matrix $T$ since the matrices $\Sigma$ and $T$ share the same eigenvectors and their eigenvalues have a one-to-one correspondence (Boente and Fraiman, 1999). Moreover, as $p \to \infty$, the two matrices coincide in the sense that the spectral norm $||\Sigma - T|| \to 0$, as long as $||\Sigma||$ (or $||T||$) is uniformly bounded, see Lemma 7.1.

Spectral properties of a standard high-dimensional SCM have been extensively studied in random matrix theory since the pioneer work of Marčenko and Pastur (1967). The standard model in this literature has the form

$$\tilde{x} = \sigma A z,$$  \hspace{1cm} (2.3)

where $A$ is as before, $\sigma$ is a constant, and $z = (z_1, \ldots, z_p)' \in \mathbb{R}^p$ is a set of i.i.d. random variables satisfying $\mathbb{E}(z_1) = 0$, $\mathbb{E}(z_1^2) = 1$, and $\mathbb{E}(z_1^4) < \infty$. Let $\tilde{x}_1, \ldots, \tilde{x}_n$ be $n$ i.i.d. copies of $\tilde{x}$ and $\tilde{S}_n = \sum_{j=1}^{n} \tilde{x}_j \tilde{x}_j' / n$ be the corresponding SCM. It is well known that the ESD of $\tilde{S}_n$ converges to the celebrated Marčenko-Pastur (MP) law when $A = I_p$, and to a generalized MP law for general matrix $A$, as $(n, p) \to \infty$ with $p/n \to c > 0$ (Marčenko and Pastur, 1967; Silverstein, 1995). The CLT for LSS of $\tilde{S}_n$ was first studied in Jonsson (1982) by assuming the population is a standard multivariate normal. One breakthrough on the CLT was obtained by Bai and Silverstein (2004), where the population is allowed to be general with $\mathbb{E}(z_1^4) = 3$. This fourth moment condition was then weakened to $\mathbb{E}(z_1^4) < \infty$ in Pan and Zhou (2008). For more references, one can refer to Bai and Silverstein (2010), Bai et al. (2015), Gao et al. (2016), and references therein. However, these results do not apply to general elliptical populations since the two models in (2.1) and (2.3) have little in common, except for normal populations. In fact, for general elliptical populations, it has been reported that the ESD of the SCM $S_n = n^{-1} \sum_{j=1}^{n} x_j x_j'$ converges to a deterministic distribution that is not a generalized MP law, but has to be characterized by both the distribution of $w$ and the limiting spectrum of $T$ through a system of implicit equations.
The involvement of \( w \) seriously interferes with our understanding of the spectrum of \( T \) from the ESD of \( S_n \). This again motivates us to shift our attention to the SSCM \( B_n \) which discards the random radius \( (w_j) \) and focus only on the directions \( (Au_j) \).

3 High-dimensional theory for eigenvalues of \( B_n \)

3.1 Limiting spectral distribution of \( B_n \)

In this section we derive a LSD for the sequence \( B_n \) under the assumptions below.

Assumption (a). Both the sample size \( n \) and population dimension \( p \) tend to infinity in such a way that \( p = p(n) \) and \( c_n = p/n \to c \in (0, \infty) \).

Assumption (b). Sample observations are \( y_j = \sqrt{p} s(x_j) = \sqrt{p} s(Au_j), \ j = 1, \ldots, n, \) where \( A \) is a \( p \times p \) deterministic and invertible matrix with \( AA' = T \) and \( (u_j) \) are i.i.d. random vectors, uniformly distributed on the unit sphere in \( \mathbb{R}^p \).

Assumption (c). The spectral norm of \( \Sigma = \mathbb{E}(y_1y_1') \) is bounded and its spectral distribution \( H_p := F_{\Sigma} \) converges weakly to a probability distribution \( H \), called population spectral distribution (PSD). Moreover, the spectral moments of \( \Sigma \) are denoted by \( \gamma_{nj} = \int t^j dH_p(t) \) and their limits by \( \gamma_j = \int t^j dH(t) \).

From Lemma 7.1, it is clear that the spectral distributions of \( \Sigma \) and \( T \) are asymptotically identical. So one can certainly replace \( \Sigma \) with \( T \) in Assumption (c), which does not affect the LSD of \( F_{B_n} \). However we keep \( \Sigma \) because it is easy to describe the CLT for LSS using the spectral distribution \( H_p \) of \( \Sigma \).

For the characterization of the LSD of \( F_{B_n} \), we need to introduce the Stieltjes transform of a measure \( G \) on the real line, which is defined as

\[
m_G(z) = \int \frac{1}{x-z} dG(x), \quad z \in \mathbb{C}^+,
\]

where \( \mathbb{C}^+ \equiv \{ z \in \mathbb{C} : \Im(z) > 0 \} \).

Theorem 3.1. Suppose that Assumptions (a)-(c) hold. Then, almost surely, the empirical spectral distribution \( F_{B_n} \) converges weakly to a probability distribution \( F^{c,H} \), whose Stieltjes transform \( m = m(z) \) is the unique solution to the equation

\[
m = \int \frac{1}{t(1-c-czm) - z} dH(t), \quad z \in \mathbb{C}^+,
\]

in the set \( \{ m \in \mathbb{C} : -(1-c)/z + cm \in \mathbb{C}^+ \} \).
The LSD $F^{c,H}$ defined in (3.1) is a generalized MP law already appeared in the seminal paper Marčenko and Pastur (1967). Let $\overline{m} = \overline{m}(z)$ denote the Stieltjes transform of $F^{c,H} = cF^{c,H} + (1-c)\delta_0$. Then (3.1) can also be represented as (Silverstein, 1995)

$$z = -\frac{1}{m} + c\int \frac{t}{1+tm}dH(t), \quad z \in \mathbb{C}^+.$$  \hfill (3.2)

For procedures on finding the density function and the support set of $F^{c,H}$ from (3.1) and (3.2), one is referred to Bai and Silverstein (2010).

### 3.2 CLT for linear spectral statistics of $B_n$

Let $F^{c_n,H_p}$ be the LSD as defined in (3.1) with the parameters $(c, H)$ replaced by $(c_n, H_p)$. Let $G_n = F^{B_n} - F^{c_n,H_p}$. We now study the fluctuation of so-called LSS of the form

$$G_n(f) := \int f(x)dG_n(x) = \int f(x)d[F^{B_n}(x) - F^{c_n,H_p}(x)],$$

where $f$ is some given measurable function. Define also the interval

$$I_c := \left[\liminf_{p \to \infty} \Lambda_{\min}^p(1 - \sqrt{c})^2, \limsup_{p \to \infty} \Lambda_{\max}^p(1 + \sqrt{c})^2\right].$$ \hfill (3.3)

**Theorem 3.2.** Suppose that Assumptions (a)-(c) hold. Let $f_1, \ldots, f_k$ be $k$ functions analytic on an open set that includes the interval $I_c$ (3.3). Then the random vector $p\{G_n(f_1), \ldots, G_n(f_k)\}$ converges weakly to a Gaussian vector $(X_{f_1}, \ldots, X_{f_k})$ with mean function

$$EX_f = \frac{-1}{2\pi i} \oint_{C_1} f(z) \int \frac{cm'(z)t^2dH(t)}{m(z)(1 + m(z)t)^3}dz - \frac{cm(z)m'(z)}{\pi i} \oint_{C_1} f(z) \times \left[\oint \frac{(\gamma_2 t - \lambda^2)dH(t)}{1 + m(z)t^2} - \int \frac{tdH(t)}{1 + m(z)t} \int \frac{\lambda^2 dH(t)}{1 + m(z)t^2} \right]dz,$$

and covariance function

$$\text{Cov}(X_f, X_g) = \frac{-1}{2\pi i} \oint_{C_1} \oint_{C_2} f(z)g(\overline{z})m'(z)m'(\overline{z}) \left(\frac{m(z)}{m(\overline{z})}\right)^2dzd\overline{z} \pm 2\gamma_2c \int xf'(x)dF(x) \int xg'(x)dF^{c,H}(x) \pm \frac{1}{\pi i} \int f(z)m'(z)dz \int xg'(x)dF^{c,H}(x) \pm \frac{1}{\pi i} \int g(z)m'(z)dz \int xf'(x)dF^{c,H}(x).$$

Here $(f, g \in \{f_1, \ldots, f_k\})$ and the contours $C_1$ and $C_2$ are non-overlapping, closed, counterclockwise orientated in the complex plane and enclosing the support of the LSD $F^{c,H}$. 


A special case of interest is a multivariate normal population \( \mathbf{x} \) that satisfies both the elliptical model (2.1) and the linear transformation model (2.3). In this case, it is interesting to compare the limiting distribution in Theorem 3.2 based on the sample SSCM \( B_n \) with the classical CLT in Bai and Silverstein (2004) based on the SCM \( S_n \). One finds that some additional and new terms appear in Theorem 3.2, namely the second contour integral in the mean function and the second to fourth terms in the covariance function above do not exist in the classical CLT in Bai and Silverstein (2004).

Another closely related work is Hu et al. (2019), where the authors study elliptical population \( \mathbf{y} = \zeta \mathbf{A} \mathbf{u} \) by assuming \( \zeta \) being independent of \( \mathbf{u} \). Though sharing the same form, our model violates their independent assumption. Specifically, we take \( \zeta \equiv \sqrt{p}/||\mathbf{A}\mathbf{u}|| \) which is correlated with \( \mathbf{u} \). It will be shown that such correlation is not (asymptotically) negligible for the distribution of LSS.

### 3.3 Asymptotic distributions of spectral moments

Among all LSS, the following spectral moments of \( F^{B_n} \) are particularly important:

\[
\hat{\beta}_{nj} = \frac{1}{p} \text{tr}(B_n^j) = \int x^j dF^{B_n}(x), \quad j = 1, 2, \ldots.
\]

The first moment \( \hat{\beta}_{n1} \) is 1 since \( \text{tr}(B_n) \equiv \text{tr}(\Sigma) \equiv p \). All other moments \( (\hat{\beta}_{nj}), j \geq 2 \), are random. Define the moments of the related MP laws

\[
\beta_{nj} = \int x^j dF^{c\cdot H_p}(x) \quad \text{and} \quad \beta_j = \int x^j dF^{c\cdot H}(x), \quad j \geq 1.
\]

From Nica and Speicher (2006), the quantities \( (\beta_{nj}) \) and \( (\gamma_{nj}) \) (moments of \( H_p \)) are related through the recursive formulae:

\[
\beta_{nj} = \sum c_{i_1 \cdots i_j}^{i_1 + \cdots + i_{j-1}} (\gamma_{n2})^{i_2} \cdots (\gamma_{nj})^{i_j} \phi(i_1, \ldots, i_j), \quad j \geq 2, \quad (3.4)
\]

and \( \beta_{n1} = \gamma_{n1} \equiv 1 \), where the sum runs over the following partitions of \( j \):

\[
(i_1, \ldots, i_j) : j = i_1 + 2i_2 + \cdots + ji_j, \quad i_i \in \mathbb{N},
\]

and \( \phi(i_1, \ldots, i_j) = j!/[i_1! \cdots i_j!(j + 1 - i_1 - \cdots - i_j)!] \). The joint limiting distribution of moments \( (\hat{\beta}_{nj})_{2 \leq j \leq k} \) can be derived from Theorem 3.2 by considering the moment functions \( f_j(x) = x^j, j = 2, \ldots, k \). For this particular case, the mean and covariance functions in the limiting distribution can be explicitly calculated.

**Corollary 3.1.** Suppose that Assumptions (a)-(c) hold. Then the random vector

\[
p\left(\hat{\beta}_{n2} - \beta_{n2}, \ldots, \hat{\beta}_{nk} - \beta_{nk}\right) \overset{D}{\rightarrow} N_{k-1}(\nu, \Psi).
\]
The mean vector $v = (v_j)_{2 \leq j \leq k}$ is given by

$$v_j = \left. \frac{tP^j}{(j-2)!} \left( \frac{P_{2,3} - 2\gamma_2P_{1,2} + 2P_{2,1}}{1 - cz^2P_{2,2}} \right) \right|_{z=0}^{(j-2)}$$

where $P_{s,t} = \int x^s(1+zx)^{-t}dH(x)$, $P = (czP_{1,1} - 1)$, and $g^{(t)}(z)$ denotes the $t$th derivative of $g(z)$ with respect to $z$. The covariance matrix $\Psi = (\psi_{ij})_{2 \leq i,j \leq k}$ has entries

$$\psi_{ij} = \frac{2}{\ell^2} \sum_{\ell=0}^{i-1} (i-\ell)u_{i,\ell}u_{j,\ell+j-\ell} + 2c\gamma_2ij\beta_i\beta_j + 2j\beta_ju_{i,\ell+j+1} + 2i\beta_iu_{j,\ell+j+1},$$

where $u_{s,t} = \left. [P^t]^{(s)} / t! \right|_{z=0}$.

### 4 Applications to spectral inference

A natural question on spatial signs is how to infer the population SSCM $\Sigma$ from the sample SSCM $B_n$ when the dimension $p$ is large. If the question was for the pair of population and sample covariance matrices $(\Sigma_0, S_n)$, this falls in the widely studied problem of estimating a large covariance matrix. Noting the fundamental difference between an SSCM and a standard covariance matrix, we indeed found nothing in the literature for properties of a high-dimensional SSCM. (to our best knowledge).

In this section, we consider a scenario where the PSD $H$ of $\Sigma$ can be modeled as a finite mixture of point masses. Using the theory of Section 3, we propose two new inference tools for the PSD $H$. First an asymptotic normal estimator is found for such a finite-mixture PSD $H$. This estimator is particularly interesting for an elliptical population because the eigenvalues of $\Sigma$ and $\Sigma_0$ are then in a well-known one-to-one correspondence. This will finally lead to a robust estimator for $\Sigma_0$ much better than some existing proposals, for example, the estimator from the procedure (C1)-(C2)-(C3) of Visuri et al. (2000). The second inference tool we develop treats the question of determination of the order of the finite mixture in $H$.

Precisely, the family of PSDs under study is a class of parameterized discrete distributions with finite support on $\mathbb{R}^+$, that is,

$$H(\theta) = w_1 \delta_{a_1} + \cdots + w_d \delta_{a_d}, \quad \theta = (a_1, w_1, \ldots, a_{d-1}, w_{d-1}) \in \Theta,$$

where $\Theta = \left\{ \theta : 0 < a_1 < \cdots < a_d < \infty; \ 0 < w_i, \sum_{i=1}^d a_i^\ell w_i = 1, \ell = 0, 1 \right\}$. Here the restriction $\sum_{i=1}^d a_i w_i = 1$ is due to the normalization condition $\int tdH_n(t) = \text{tr}(\Sigma)/p \equiv 1$. Note that the model (4.1) depends on an integer parameter $d \geq 1$, referred as the order of $H$. Such finite mixtures have already been employed for the standard large covariance matrix $\Sigma_0$, see El Karoui (2008), Rao et al. (2008), Bai et al. (2010) and Li and Yao (2014). Similar to El Karoui (2008), we adopt the setting of fixed PSDs in this section, i.e. $(\Sigma, H_0) \equiv (c, H)$ for all $(n, p)$ large.
4.1 Estimation of a PSD

For the model in (4.1), we follow the moment method in Bai et al. (2010) for the PSD estimation. Given a known order \( d \), the method first estimates the moments \( (\gamma_j) \) of \( H \) through the recursive formulae in (3.4), and then solve a system of moment equations, \( \{\hat{\gamma}_j = \sum_{i=1}^{d} a_i^j w_i, \ j = 0, \ldots, 2d - 1\} \), to get a consistent estimator of \( \theta \).

In our situation, with notation \( \beta_j = (\beta_2, \ldots, \beta_j)' \) and \( \gamma_j = (\gamma_2, \ldots, \gamma_j)' \) for \( j \geq 2 \), we denote
\[
 g_1 : \gamma_{2d-1} \to \theta \quad \text{and} \quad g_{2,j} : \beta_j \to \gamma_j
\]
as the mappings between the corresponding vectors. These mappings are all one-to-one and the determinants of their Jacobian matrices are all nonzero, see Bai et al. (2010).

Therefore, applying Theorem 3.1, \( \hat{\beta}_j := (\hat{\beta}_{d,2}, \ldots, \hat{\beta}_d)' \overset{a.s.}{\to} \beta_j \) which implies that \( \hat{\theta}_n := g_1 \circ g_{2,2d-1}(\hat{\beta}_{2d-1}) \overset{a.s.}{\to} \theta \), as \( (n, p) \to \infty \). However, as shown by the CLT in Corollary 3.1, the estimator \( \hat{\beta}_j \) has a bias of the order \( O(1/p) \). So it’s natural to modify \( \hat{\beta}_j \) by subtracting its limiting mean in the CLT to obtain a better estimator of \( \theta \). Beyond this correction, the CLT can also provide confidence regions for the parameter \( \theta \).

Denote the modified estimators of \( \beta_j \), \( \gamma_j \), and \( \theta \) by
\[
\hat{\beta}_j^* = \hat{\beta}_j - \frac{1}{p}(\hat{v}_2, \ldots, \hat{v}_j)', \quad \hat{\gamma}_j^* = g_{2,j}(\hat{\beta}_j^*), \quad \text{and} \quad \hat{\theta}_n^* = g_1(\hat{\gamma}_j^*), \quad \text{(4.2)}
\]
respectively, where \( \hat{v}_\ell = v_\ell(\hat{\beta}_j) \) with \( v_\ell \) defined in Corollary 3.1 for \( \ell = 2, \ldots, j \). From Theorem 3.1, Corollary 3.1, and a standard application of the Delta method, one may easily get asymptotic properties of these estimators.

**Theorem 4.1.** Suppose that Assumptions (a)-(c) hold and the true value \( \theta \) is an inner point of \( \Theta \). Then we have
\[
 p(\hat{\gamma}_j^* - \gamma_j) \overset{D}{\to} N_{j-1}(0, J_{j,2,j}\Psi_{j,2,j}),
\]
\[
 p(\hat{\theta}_n^* - \theta) \overset{D}{\to} N_{2k-2}(0, J_1J_{2,2d-1}\Psi_{2d-1}J_{2,2d-1}'J_1'),
\]
where \( J_1 \) and \( J_{2,\ell} \) represent the Jacobian matrices \( \partial g_1/\partial \gamma_{2d-1} \) and \( \partial g_{2,\ell}/\partial \beta_\ell \), respectively, and \( \Psi_{\ell} \) is defined in Corollary 3.1 with \( k = \ell \).

4.2 Test for the order of a PSD

The aforementioned estimation procedure requires that the order \( d \) of the PSD be prespecified. In general, this prior knowledge should be testified in advance. To deal with this problem, we consider the hypotheses
\[
 H_0 : d \leq d_0 \quad \text{v.s.} \quad H_1 : d > d_0, \quad \text{(4.4)}
\]
where \( d_0 \geq 1 \) is a given positive integer. These hypotheses can also be regarded as a generalization of the well-known sphericity hypotheses on covariance matrices, i.e. the case \( d_0 = 1 \).

In Qin and Li (2016), a test procedure was outlined based on a moment matrix \( \Gamma \) and its estimator \( \hat{\Gamma} \) which can be formulated as

\[
\Gamma = \begin{pmatrix}
1 & \gamma_1 & \cdots & \gamma_{d_0} \\
\gamma_1 & \gamma_2 & \cdots & \gamma_{d_0+1} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{d_0} & \gamma_{d_0+1} & \cdots & \gamma_{2d_0}
\end{pmatrix} \quad \text{and} \quad \hat{\Gamma} = \begin{pmatrix}
1 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{d_0} \\
\hat{\gamma}_1 & \hat{\gamma}_2 & \cdots & \hat{\gamma}_{d_0+1} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\gamma}_{d_0} & \hat{\gamma}_{d_0+1} & \cdots & \hat{\gamma}_{2d_0}
\end{pmatrix}.
\]

Here we set \( \hat{\gamma}_1 = 1 \) and \( \hat{\gamma}_j = \hat{\gamma}_j^* \), as defined in (4.2), for \( j \geq 2 \). It has been proved that the determinant \( \det(\Gamma) \) of \( \Gamma \) is zero if the null hypothesis in (4.4) holds, otherwise \( \det(\Gamma) \) is strictly positive (Li and Yao, 2014). Therefore, the determinant \( \det(\hat{\Gamma}) \) can serve as a test statistic for (4.4) and the null hypothesis shall be rejected if the statistic is large. Applying Theorem 4.1 and the main theorem in Qin and Li (2016), the asymptotic distribution of \( \det(\hat{\Gamma}) \) is obtained immediately.

**Theorem 4.2.** Suppose that Assumptions (a)-(c) hold. Then the statistic \( \det(\hat{\Gamma}) \) is asymptotically normal, i.e.

\[
p \left( \det(\hat{\Gamma}) - \det(\Gamma) \right) \overset{D}{\rightarrow} N(0, \sigma^2),
\]

where \( \sigma^2 = \alpha'\Omega\alpha \) with \( \alpha = \text{vec}(\text{adj}(\Gamma)) \), the vectorization of the adjoint matrix of \( \Gamma \). The first two rows and columns of the \((2d_0+1)\times(2d_0+1)\) matrix \( \Omega \) consist of zero and the remaining sub-matrix \( J_{2d_0}\Psi_{2d_0}J_{2d_0}' \) is defined in (4.3). The \((d_0+1)^2\times(2d_0+1)\) matrix \( V = (v_{ij}) \) is a 0-1 matrix with only \( v_{i,a} = 1, a_i = i - \lfloor (i-1)/(d_0+1) \rfloor d_0, i = 1, \ldots, (d_0+1)^2 \), where \( \lfloor x \rfloor \) denotes the greatest integer not exceeding \( x \).

From Theorem 4.1, the limiting variance \( \sigma^2 \) in (4.5) is a continuous function of \( \gamma_{4d_0} \). While, under the null hypothesis, this variance is a function of \( \gamma_{2d_0-1} \), denoted by \( \sigma_{H_0}^2(\gamma_{2d_0-1}) \). Let \( \hat{\sigma}_{H_0}^2 = \sigma_{H_0}^2(\hat{\gamma}_{2d_0-1}) \). Then it is a strongly consistent estimator of \( \sigma_{H_0}^2(\gamma_{2d_0-1}) \).

**Corollary 4.1.** Suppose that Assumptions (a)-(c) hold. Then, under the null hypothesis,

\[
T_n := \frac{p\det(\hat{\Gamma})}{\hat{\sigma}_{H_0}} \overset{D}{\rightarrow} N(0, 1),
\]

as \( n \to \infty \). In addition, the asymptotic power of \( T_n \) tends to 1.

Corollary 4.1 follows directly from Theorem 4.2 and its proof is thus omitted. This corollary includes as a particular case the sphericity test. For this case, the test statistic reduces to \( T_n = n(\hat{\gamma}_0^* - 1)/2 \) and its null distribution is consistent with that in Paindaveine and Verdebout (2016) which is obtained by a direct and completely different method.
4.3 Simulation experiments

Simulations are carried out to evaluate the performance of the proposed estimation and test for discrete PSDs in (4.1). Samples \((z_{ij})\) are drawn from \(N(0, 1)\) and all empirical statistics are calculated from 10,000 independent replications.

The estimation procedure is tested for the following two PSDs.

- Model 1: \(H_1 = 0.5\delta_{0.5} + 0.5\delta_{1.5}\) and \(c = 2\);
- Model 2: \(H_2 = 0.3\delta_{0.2} + 0.4\delta_{1} + 0.3\delta_{1.8}\) and \(c = 1/4\).

The sample sizes are \(n = 100, 200, 400\) for Model 1 and \(n = 400, 800, 1600\) for Model 2, respectively. In addition to empirical means and standard deviations of all estimators, we also calculate 95% confidence intervals for all parameters and report their coverage probabilities. Results are collected in Tables 1 and 2. The consistency of all estimators is clearly demonstrated.

Table 1: Estimation for Model 1 with sample size \(n = 100, 200, 400\) and \(c = 2\). The number of independent replications is 10,000 and the nominal coverage probability (C. P.) is fixed at 95%.

| \(\theta\) | \(n = 100\) Mean | St. D. | C. P. | \(n = 200\) Mean | St. D. | C. P. | \(n = 400\) Mean | St. D. | C. P. |
|---|---|---|---|---|---|---|---|---|---|
| \(a_1 = 0.5\) | 0.4939 | 0.1145 | 0.9315 | 0.4960 | 0.0550 | 0.9491 | 0.5000 | 0.0269 | 0.9486 |
| \(w_1 = 0.5\) | 0.4915 | 0.1135 | 0.9137 | 0.4968 | 0.0588 | 0.9423 | 0.4997 | 0.0292 | 0.9488 |
| \(a_2 = 1.5\) | 1.5030 | 0.1330 | 0.9288 | 1.4990 | 0.0668 | 0.9426 | 1.4998 | 0.0329 | 0.9487 |
| \(w_2 = 0.5\) | 0.5085 | 0.1135 | 0.9137 | 0.5032 | 0.0588 | 0.9423 | 0.5003 | 0.0292 | 0.9488 |

Table 2: Estimation for Model 2 with sample size \(n = 400, 800, 1600\) and \(c = 1/4\). The number of independent replications is 10,000 and the nominal coverage probability (C. P.) is fixed at 95%.

| \(\theta\) | \(n = 400\) Mean | St. D. | C. P. | \(n = 800\) Mean | St. D. | C. P. | \(n = 1600\) Mean | St. D. | C. P. |
|---|---|---|---|---|---|---|---|---|---|
| \(a_1 = 0.2\) | 0.1887 | 0.0429 | 0.9227 | 0.1988 | 0.0147 | 0.9358 | 0.2003 | 0.0071 | 0.9367 |
| \(w_1 = 0.3\) | 0.2824 | 0.0447 | 0.9403 | 0.2956 | 0.0184 | 0.9525 | 0.2990 | 0.0090 | 0.9483 |
| \(a_2 = 1.0\) | 0.9960 | 0.1347 | 0.9345 | 0.9924 | 0.0661 | 0.9486 | 0.9991 | 0.0337 | 0.9433 |
| \(w_2 = 0.4\) | 0.4064 | 0.0373 | 0.9453 | 0.4012 | 0.0209 | 0.9239 | 0.4002 | 0.0110 | 0.9351 |
| \(a_3 = 1.8\) | 1.7824 | 0.0856 | 0.9236 | 1.7919 | 0.0440 | 0.9413 | 1.7960 | 0.0227 | 0.9392 |
| \(w_3 = 0.3\) | 0.3113 | 0.0696 | 0.9221 | 0.3031 | 0.0365 | 0.9429 | 0.3008 | 0.0189 | 0.9420 |

Next we examine the test procedure for the order of a PSD. The following two models are employed in this experiment:

- Model 3: \(H_3 = 0.5\delta_{1-x} + 0.5\delta_{1+x}\);
Table 3: Empirical size and power of $T_n$ in percentage under Model 3 and Model 4 with the sample size $n = 400$. The number of independent replications is 10,000 and the nominal significance level is 0.05.

| $x$ | 0  | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 | 0.12 | 0.14 | 0.16 | 0.18 |
|-----|----|------|------|------|------|------|------|------|------|------|
| $c = \frac{1}{2}$ | 5.24 | 5.81 | 9.13 | 17.91 | 34.86 | 62.30 | 87.31 | 98.01 | 99.90 | 100 |
| $c = 1$ | 5.33 | 5.92 | 8.43 | 18.09 | 35.62 | 63.12 | 88.14 | 98.69 | 99.96 | 100 |
| $c = 2$ | 4.76 | 6.39 | 9.69 | 17.39 | 35.23 | 63.57 | 88.15 | 98.67 | 99.97 | 100 |

| $x$ | 0  | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 |
|-----|----|------|------|------|------|------|------|------|------|------|
| $c = \frac{1}{2}$ | 4.75 | 7.19 | 17.49 | 43.96 | 79.28 | 97.06 | 99.87 | 100 | 100 | 100 |
| $c = 1$ | 5.05 | 6.31 | 12.22 | 26.78 | 53.74 | 80.74 | 95.07 | 99.52 | 99.97 | 100 |
| $c = 2$ | 4.88 | 5.65 | 8.56 | 16.33 | 30.09 | 49.17 | 71.60 | 86.54 | 95.20 | 98.61 |

- Model 4: $H_4 = 0.25\delta_{0.5 - x} + 0.25\delta_{0.5 + x} + 0.25\delta_{1.5 - x} + 0.25\delta_{1.5 + x}$.

Here the parameter $x \in [0, 0.5]$ represents the distance between the null and alternative hypotheses. In particular, Model 3 is used for testing $H_0 : d \leq 1$ (sphericity test) with $x$ ranging from 0 to 0.2 by step 0.18 and Model 4 is for testing $H_0 : d \leq 2$ with $x$ ranging from 0 to 0.45 by step 0.05. The sample size is $n = 400$, the dimension-sample size ratios are $c = 1/2, 1, 2$, and the significance level is fixed at $\alpha = 0.05$. Results summarized in Table 3 show that the proposed test has accurate empirical size and its power tends to 1 as the parameter $x$ increases under the two models.

5 Application to robust statistics

In this section we develop a few applications of the general theory of Section 3 to robust statistics using the sample SSCM $B_n$.

5.1 Robustness

We examine the robustness of several estimators for the shape matrix $T$ when sample data include outliers. Four estimators derived from $S_n$ and $B_n$ are considered in this comparison. Let be the spectral decomposition of $S_n$ and $B_n$:

$$S_n = U_s \Lambda_s U'_s \quad \text{and} \quad B_n = U_b \Lambda_b U'_b,$$

where the $\Lambda$’s are diagonal matrices of eigenvalues, sorted in ascending order, and the $U$’s are matrices of corresponding eigenvectors, respectively. In addition, we define a regularization function $r(\cdot)$ as

$$r(A) = p \frac{A}{tr(A)},$$
for any $p \times p$ matrix $A$ with non-zero trace. Obviously, this function normalizes $A$ such that $\text{tr}(A) = p$. With the above notations, the four estimators of $T$ we examine are as follows:

(i) Regularized SCM $\hat{T}_1$, 
$$\hat{T}_1 = r(S_n);$$

(ii) Spectrum-corrected SCM $\hat{T}_2$, 
$$\hat{T}_2 = r(U_s \Lambda_2 U_s'),$$
where $\Lambda_2 = \text{diag}(\lambda_{21}, \ldots, \lambda_{2p})$ is a collection of ascendingly sorted estimators of population eigenvalues using a moment method developed in Li and Yao (2014);

(iii) Robust sample SSCM $\hat{T}_3$ constructed from the procedures (C1)-(C2)-(C3) of Visuri et al. (2000),
$$\hat{T}_3 = r(U_b \Lambda_3 U_b'),$$
where $\Lambda_3 = \text{diag}(\lambda_{31}, \ldots, \lambda_{3p})$ with $\lambda_{3k}$ being the square of the MAD of the $k$th row of $(U_b'x_1, \ldots, U_b'x_n)$ for $k = 1, \ldots, p$.

(iv) Spectrum-corrected sample SSCM $\hat{T}_4$, 
$$\hat{T}_4 = r(U_b \Lambda_4 U_b'),$$
where the correction $\Lambda_4 = \text{diag}(\lambda_{41}, \ldots, \lambda_{4p})$ is obtained following three steps:

- Step 1: Estimate the PSD $H_p$ of $\Sigma$ from the ESD $F^{B_n}$ through the procedure in Section 4.1 to get, say, $\hat{H}_p$;
- Step 2: Estimate the eigenvalues of $T$ from $\hat{H}_p$ using the correspondence between the eigenvalues of $\Sigma$ and $T$ as given in Lemma 7.1;
- Step 3: Sort the obtained estimates of the eigenvalues in ascending order to obtain $\Lambda_4$.

The performance of the four estimators $\{\hat{T}_j\}_{1 \leq j \leq 4}$ are tested under two models below.

**Model 1:** Contaminated normal distribution of elliptical form:
$$(1 - \varepsilon)N(0, T) + \varepsilon N(0, 16T),$$
where the population shape matrix $T$ is a diagonal matrix,
$$T = \text{diag}(0.5, \ldots, 0.5, 1.5, \ldots, 1.5).$$

This model implies there are about $100\varepsilon\%$ outlying observations with large amplitude. The mixing parameter $\varepsilon$ takes two values 0 (uncontaminated) and 0.01 (contaminated by 1% outliers).
**Model 2:** Contaminated normal distribution of non-elliptical form:

\[(1 - \varepsilon)N(0, \mathbf{T}) + \varepsilon N(0, 16\mathbf{\tilde{T}})\]

where the population shape matrix \(\mathbf{T}\) is the same as in Model 1 and the mixing parameter \(\varepsilon\) takes values 0.01 and 0.05. For outliers, their shape matrix is

\[\mathbf{\tilde{T}} = \text{Diag}(1.5, \ldots, 1.5, 0.5, \ldots, 0.5)\]

The population dimension is \(p = 2, 40, 80, 120, 160, 200\), and the sample size is \(n = 100\). All statistics are averaged from 1000 independent replications. The number of outliers is fixed at 100\(\varepsilon\) for a given \(\varepsilon\).

For each estimator \(\hat{\mathbf{T}} \in \{\hat{\mathbf{T}}_1, \hat{\mathbf{T}}_2, \hat{\mathbf{T}}_3, \hat{\mathbf{T}}_4\}\), we calculate three distances from \(\hat{\mathbf{T}}\) to its target matrix \(\mathbf{T}\), including the Frobenius distance, the Kullback-Leibler (KL) distance and the spectral distance. The KL distance is not applicable to \(\hat{\mathbf{T}}_1\) and \(\hat{\mathbf{T}}_3\) for cases with \(p \geq n\) because their determinants are thus zero. Figures 1, 2 and 3 summarize the results. They show that, when there is no outlier, \(\hat{\mathbf{T}}_1\) and \(\hat{\mathbf{T}}_3\) are comparable and both suffers from large biases caused by the high-dimensional effect. Such bias can be much alleviated by means of spectral correction as demonstrated by \(\hat{\mathbf{T}}_2\) and \(\hat{\mathbf{T}}_4\) which have almost the same accuracy. Note that the remaining bias is clearly present and it is a pity that there is no effective way at present to remove this remaining bias entirely. Therefore, the performance of \(\hat{\mathbf{T}}_2\) and \(\hat{\mathbf{T}}_4\) with \(\varepsilon = 0\) can serve as a benchmark for the four estimators when comparing their robustness against outliers. As shown in the figures, in the presence of outliers, the estimator \(\hat{\mathbf{T}}_3\) is more robust than \(\hat{\mathbf{T}}_1\), but both of them are still heavily biased for large \(p\). This is explained by the fact that both of them are not adapted to high dimensions. For the two other estimators \(\hat{\mathbf{T}}_2\) and \(\hat{\mathbf{T}}_4\) with high-dimensional correction, the estimator \(\hat{\mathbf{T}}_2\) (based on the SCM \(\mathbf{S}_n\)) becomes unstable depending on the percentage of outliers and the magnitude of the dimension \(p\). In contrast, the estimator \(\hat{\mathbf{T}}_4\) is robust against outliers and is nearly identical to the benchmark in all cases under study. Finally, the rescaled SCM \(\hat{\mathbf{T}}_1\) stays the worst estimator in all the scenarios since it is neither robust nor high-dimensional adapted.

### 5.2 Application to blind source separation

Inspired by the signal processing application developed in Section 5 of Visuri et al. (2000), we develop an application of our theory to robust signal prewhitening using the sample SSCM \(\mathbf{B}_n\). Note that examples developed in the reference are low-dimensional, all the simulation experiments have dimension \(p = 2\) while experiments will be here conducted for high-dimensional data with dimension varying from 200 to 800.

Consider the blind source separation (BSS) problem for an observed “colored” signal \(\mathbf{x}\) of the form

\[\mathbf{x} = \mathbf{\Lambda}\mathbf{\xi} + \mathbf{e},\]

(5.1)
where \( \Lambda \) is a \( p \times p \) invertible mixing matrix, \( \xi \) a \( p \) dimensional source signal vector, and \( \varepsilon \) is additive noise vector. The source signal \( \xi \) is typically assumed to be an uncorrelated non-Gaussian vector with zero mean (Cardoso, 1989).

Let \( x_1, \ldots, x_n \) be \( n \) i.i.d. copies of \( x \) defined in (5.1) with the above settings. Blind source separation aims at a data-driven linear filter \( \hat{Q} \) such that the filtered outputs

\[
W_n = \{ \hat{Q}x_1, \ldots, \hat{Q}x_n \}
\]
can be considered as a satisfactory reconstruction of the unobserved signals \( \{\xi_1, \ldots, \xi_n\} \). This filtering is also referred as prewhitening in signal processing literature.

A natural choice for the filter \( \hat{Q} \) is \( \hat{Q} = \hat{T}^{-\frac{1}{2}} \) for some estimator \( \hat{T} \) of the shape matrix \( T := A^2 \). So the estimators \( \{\hat{T}_1, \hat{T}_2, \hat{T}_3, \hat{T}_4\} \) discussed in Section 5.2 are all possible candidates. For any estimator \( \hat{T}_\ell \), the observations are filtered using filter matrix \( \hat{Q}_\ell := \hat{T}_\ell^{-\frac{1}{2}} \), that is, the (Moore-Penrose) inverse of the square root of the matrix \( \hat{T}_\ell \). Note that
the filters $\hat{Q}_1, \hat{Q}_2$ are based on the SCM $S_n$, thus deemed to be less robust than the filters $\hat{Q}_3$ and $\hat{Q}_4$ which are based on the SSCM $B_n$. Moreover, $\hat{Q}_4$ is expected to perform better than $\hat{Q}_3$ in high-dimensional situations.

Each filter $\hat{Q}_\ell$ is assessed by comparing the filtered vectors $(\hat{Q}_\ell x_i)_i$ with the true
signals \((\xi_i)_{i=1}^n\) through an averaged Euclidean distance given by

\[d_\ell = \frac{1}{n} \sum_{i=1}^n \frac{1}{w_i} \|\hat{Q}_\ell x_i - \xi_i\|_2, \quad 1 \leq \ell \leq 4.\]  

(5.2)

Here each error term is scaled according to its actual scale \(w\) which is known from the simulation setting. In situations where forecasting is of interest, one may use the filter matrix to whiten new observations. Therefore it’s worth to quantify the performance of the whitennings in out-of-sample scenarios. Moreover, the distance in (5.2) is still applicable except that the uncontaminated observations \((x_i)\) is new and independent of the estimate \(\hat{Q}_\ell\).

The simulation design is as follows:

(a) The signal \(\xi\) follows a standardized multivariate double exponential distribution. Precisely, see (2.1),

\[\xi = w u, \quad w \sim \text{Gamma}(p,1) / \sqrt{p + 1},\]

where Gamma\((k, \theta)\) represents a Gamma distribution with shape and scale parameters \((k, \theta)\). The scaling factor \(1/\sqrt{p + 1}\) is chosen such that Cov\((\xi) = I_p\).

(b) The mixing matrix is

\[\Lambda = U \begin{pmatrix} \sqrt{0.5}I_p/2 & 0 \\ 0 & \sqrt{1.5}I_p/2 \end{pmatrix} U',\]

assuming \(p\) even, where \(U\) is an orthogonal matrix. The matrix \(\Lambda\) is normalized by \(\text{tr}(\Lambda^2)/p = 1\). The matrix \(U\) is chosen randomly and fixed once generated.

(c) The additive noise vector follows a Gaussian mixture:

\[e \sim (1 - \varepsilon) N(0, \sigma_1^2 I_p) + \varepsilon N(0, \sigma_2^2 I_p),\]

where \((\sigma_1^2, \sigma_2^2) = (0.1, 25)\) and the mixing parameter \(\varepsilon\) ranges from zero to 0.1. The model thus has in average 100\(\varepsilon\)% outlying observations with large amplitude.

The dimensions are \((p, n) = (50, 100)\) and \((200, 100)\). The in-sample/out-of-sample distances of \((\hat{Q}_\ell)\) are calculated and averaged from 1000 independent replications. Figure 4 illustrates the results. It clearly shows that the whitening based on \(\hat{Q}_4\) outperforms its competitors in all studied cases with the presence of outliers. It is particularly interesting to see that among the four filters, the ones using sample SSCM have their errors remaining stable when the number of outliers increases. In addition, the filter \(\hat{Q}_4\) is superior to \(\hat{Q}_3\) by virtue of the spectral correction.
In-sample performance of whitening with (p,n)=(50,100)

Out-of-sample performance of whitening with (p,n)=(50,100)

In-sample performance of whitening with (p,n)=(200,100)

Out-of-sample performance of whitening with (p,n)=(200,100)

Figure 4: Weighted Euclidean distance between whitened vectors and source vectors. The distance is averaged from 1000 independent replications. The horizontal axis $k = \varepsilon n$ denotes the average number of outliers in the model.
6 Proofs

6.1 Proof of Theorem 3.1

By definition, a uniformly distributed random vector \( u \) on the unit sphere in \( \mathbb{R}^p \) has expression \( u = z/||z|| \) with \( z \sim N(0, I_p) \). Accordingly, the spatial-sign population \( y \) can be formulated as

\[
y = \sqrt{p}Au/||Au|| = \sqrt{p}Az/||Az||.
\]

Since the eigenvalues of \( B_n \) are invariant under orthogonal transformation, one may further assume the matrix \( A \) to be diagonal. Applying Lemma 7.2 to the population \( y \), one may get

\[
\text{Var}(y'C_py) = o(p^2) \tag{6.1}
\]

for \( \{C_p\} \) any bounded sequence of symmetric matrices. Hence, Theorem 3.1 follows from (6.1) and Theorem 1.1 in Bai and Zhou (2008).

6.2 Proof of Theorem 3.2

General strategy of the proof comes from the groundbreaking work Bai and Silverstein (2004), which analyzes the resolvent of sample covariance matrix. New challenges here lie in the following two aspects.

1. The spatial-sign population possesses nonlinear correlation among its components, which cannot be accommodated by the linear transform model studied in Bai and Silverstein (2004). As a result, many technical steps in Bai and Silverstein (2004) have to be updated for the new model.

2. The identity in Lemma 7.2 is fundamentally important in the proof. Comparing it with Equation (1.15) in Bai and Silverstein (2004), the counterpart for the linear transform model, the new identity is more complex, which will result in more complicated calculations in deriving the asymptotic distribution of LSS.

6.2.1 Sketch of the proof of Theorem 3.2

We begin with defining a rectangular contour enclosing the support of the LSD \( F^{c,H} \). Since the support is a subset of the interval \([s_l, s_r]\) with

\[
s_l = \liminf_{p \to \infty} \lambda^\Sigma_{\min}(1 - \sqrt{c})^2 I_{(0,1)}(c) \quad \text{and} \quad s_r = \limsup_{p \to \infty} \lambda^\Sigma_{\max}(1 + \sqrt{c})^2,
\]

we choose two numbers \( x_l < x_r \) such that \([s_l, s_r] \subset (x_l, x_r)\). Let \( v_0 > 0 \) be arbitrary, then the contour can be described as

\[
\mathcal{C} = \{x \pm iv : x \in [x_l, x_r]\} \cup \{x + iv : x \in \{x_r, x_l\}, v \in [-v_0, v_0]\}. \tag{6.3}
\]
Let \( m_0(z) \) and \( \tilde{m}_0(z) \) be the Stieltjes transforms of \( F^{c_n,H_p} \) and \( c_nF^{c_n,H_p} + (1 - c_n)\delta_0 \), we then define a random process on \( \mathcal{C} \) as

\[
M_n(z) = p[m_n(z) - m_0(z)] = n[m_n(z) - \tilde{m}_n(z)], \quad z \in \mathcal{C}.
\]

From Cauchy's integral formula, for any \( k \) analytic functions \( (f_\ell) \) and complex numbers \( (\alpha_\ell) \), we have

\[
\sum_{\ell=1}^{k} p\alpha_\ell \int f_\ell(x)dG_n(x) = -\sum_{\ell=1}^{k} \frac{\alpha_\ell}{2\pi i} \oint_{\mathcal{C}} f_\ell(z)M_n(z)dz,
\]

when all sample eigenvalues fall in the interval \((x_l, x_r)\), which holds asymptotically with probability one. In order to deal with the small probability event where some eigenvalues fall outside of the interval in finite dimensional situations, for \( z = x + iv \in \mathcal{C} \), Bai and Silverstein (2004) suggested truncating \( M_n(z) \) as

\[
\tilde{M}_n(z) = \begin{cases} 
M_n(z) & z \in \mathcal{C}_n, \\
M_n(x + in^{-1}\varepsilon_n) & x \in \{x_l, x_r\} \text{ and } v \in [0, n^{-1}\varepsilon_n], \\
M_n(x - in^{-1}\varepsilon_n) & x \in \{x_l, x_r\} \text{ and } v \in [-n^{-1}\varepsilon_n, 0],
\end{cases}
\]

(6.4)

where

\[
\mathcal{C}_n = \{x \pm iv_0 : x \in [x_l, x_r]\} \cup \{x \pm iv : x \in \{x_l, x_r\}, v \in [n^{-1}\varepsilon_n, v_0]\},
\]

a regularized version of \( \mathcal{C} \) excluding a small segment near the real line, and the positive sequence \( (\varepsilon_n) \) decreases to zero satisfying \( \varepsilon_n > n^{-a} \) for some \( a \in (0, 1) \). From similar arguments on Page 563 in Bai and Silverstein (2004), for any \( \ell \in \{1, \ldots, k\} \), one may get that

\[
\oint_{\mathcal{C}} f_\ell(z)M_n(z)dz = \oint_{\mathcal{C}} f_\ell(z)\tilde{M}_n(z)dz + o_p(1).
\]

Hence, the proof of Theorem 3.2 can be completed by the convergence of \( \tilde{M}_n(z) \) on \( \mathcal{C} \) as stated in the following lemma.

**Lemma 6.1.** Under Assumptions (a)-(c), the random process \( \tilde{M}_n(\cdot) \) on \( \mathcal{C} \) converges weakly to a two-dimensional Gaussian process \( M(\cdot) \). The mean and covariance functions are

\[
\mathbb{E}M(z) = \int \frac{c(m'(z)t)^2dH(t)}{m(z)(1 + m(z)t)^3} + 2cm(z)m'(z) \left[ \int \frac{\gamma_2 t - t^2 dH(t)}{1 + m(z)t} \times \right.
\]

\[
\left. \int \frac{tdH(t)}{(1 + m(z)t)^2} - \int \frac{tdH(t)}{1 + m(z)t} \int \frac{t^2 dH(t)}{(1 + m(z)t)^2} \right],
\]

(6.5)
and
\[
\text{Cov}(M(z), M(\bar{z})) = \frac{2m'(z)m'(\bar{z})}{(m(z) - m(\bar{z}))^2} - \frac{2}{(z - \bar{z})^2} + \frac{2\gamma_2}{c} (m(z) + zm'(z)) (m(\bar{z}) + \bar{z}m'(\bar{z})) - \frac{2}{c} \left( \frac{m'(z)}{m^2(z)} - 1 \right) (m(z) + zm'(z)),
\]
respectively, for \( z \neq \bar{z} \in \mathbb{C} \), where \( \gamma_2 = \int t^2 dH(t) \) is the second moment of the PSD \( H \).

### 6.2.2 Proof of Lemma 6.1
We prove in this part the convergence of \( \hat{M}_n(z) \) following the main strategy in Bai and Silverstein (2004). Without loss of generality, we assume \( \|\Sigma\| \leq 1 \) for all \( p \) and denote by \( K \) any constant appearing in inequalities. Some quantities are listed below which will be used frequently throughout of the proof.

\[
r_j = \frac{1}{\sqrt{n}} y_j, \quad B_n = \sum_{j=1}^{n} r_j r_j', \quad D(z) = B_n - z I,
\]
\[
D_j(z) = D(z) - r_j r_j', \quad D_{ij}(z) = D(z) - r_i r_i' - r_j r_j', \quad (i \neq j),
\]
\[
\varepsilon_j(z) = r_j' D_j^{-1}(z) r_j - \frac{1}{n} \text{tr} \Sigma D_j^{-1}(z),
\]
\[
\delta_j(z) = r_j' D_j^{-2}(z) r_j - \frac{1}{n} \text{tr} \Sigma D_j^{-2}(z),
\]
\[
\beta_j(z) = \frac{1}{1 + r_j' D_j^{-1}(z) r_j}, \quad \tilde{\beta}_j(z) = \frac{1}{1 + n^{-1} \text{tr} \Sigma D_j^{-1}(z)},
\]
\[
b_n(z) = \frac{1}{1 + n^{-1} \text{tr} \Sigma D_j^{-1}(z)}.
\]

Note that the last three quantities are bounded in absolute value by \( |z|/v \) for any \( z = u + iv \in \mathbb{C}^+ \).

Splitting \( \hat{M}_n(z) \) into two parts as
\[
\hat{M}_n(z) = M_n^{(1)}(z) + M_n^{(2)}(z),
\]
where
\[
M_n^{(1)}(z) = p[m_n(z) - \mathbb{E}m_n(z)] \quad \text{and} \quad M_n^{(2)}(z) = p[\mathbb{E}m_n(z) - m_0(z)].
\]
We will deal with the convergence of \( M_n^{(1)}(z) \) and \( M_n^{(2)}(z) \), respectively.

**Step 1: Finite dimensional convergence of \( M_n^{(1)}(z) \) in distribution.** For any \( w \) complex numbers \( z_1, \ldots, z_w \in \mathbb{C}_n \), this step finds joint limiting distribution of
\[
\left[ M_n^{(1)}(z_1), \ldots, M_n^{(1)}(z_w) \right]
\]
(6.7)
by the martingale CLT in Lemma 7.4. To this end, we extend the identity (1.15) and Lemma 2.2 in Bai and Silverstein (2004) to Lemmas 7.2 and 7.3, respectively. These two lemmas are used to calculate the limiting covariance function and verify Lindeberg’s condition when applying the martingale CLT.

Let $E_0(\cdot)$ denote expectation and $E_j(\cdot)$ denote conditional expectation with respect to the $\sigma$-field generated by $r_1, \ldots, r_j$, $j = 1, \ldots, n$. From the martingale decomposition and the identity

$$D^{-1}(z) - D_j^{-1}(z) = -D_j^{-1}(z) r_j' r_j D_j^{-1}(z) \beta_j(z), \quad (6.8)$$

we get

$$M_n^{(1)}(z) = \sum_{j=1}^{n} (E_j - E_{j-1}) \text{tr} [D^{-1}(z) - D_j^{-1}(z)]$$

$$= -\sum_{j=1}^{n} (E_j - E_{j-1}) \beta_j(z) r_j' r_j,$$

$$= \sum_{j=1}^{n} (E_j - E_{j-1}) \frac{d\log(\beta_j(z)/\beta_j(z))}{dz},$$

$$= \frac{d}{dz} \sum_{j=1}^{n} (E_j - E_{j-1}) \log[1 - \beta_j(z) \epsilon_j(z) + \beta_j(z) \beta_j(z) \epsilon_j^2(z)], \quad (6.9)$$

where the last equality is from the identity $\beta_j(z) = \bar{\beta}_j(z) - \beta_j^2(z) \epsilon_j(z) + \beta_j^2(z) \beta_j(z) \epsilon_j^2(z)$.

By Lemma 7.3 and Lemma 2.1 in Bai and Silverstein (2004), we have

$$E \left| \sum_{j=1}^{n} (E_j - E_{j-1}) \bar{\beta}_j(z) \beta_j(z) \epsilon_j^2(z) \right|^2 \to 0.$$

Thus applying the Taylor expansion to the log function in (6.9), one may conclude

$$M_n^{(1)}(z) = -\frac{d}{dz} \sum_{j=1}^{n} (E_j - E_{j-1}) \bar{\beta}_j(z) \epsilon_j(z) + o_p(1)$$

$$= -\frac{d}{dz} \sum_{j=1}^{n} E_j \bar{\beta}_j(z) \epsilon_j(z) + o_p(1).$$

Therefore, we turn to consider the martingale difference sequence

$$Y_{nj}(z) := \frac{d}{dz} E_j \bar{\beta}_j(z) \epsilon_j(z), \quad j = 1, \ldots, n.$$
The Lyapunov condition for this sequence is guaranteed by the fact that
\[
\sum_{j=1}^{n} \mathbb{E}|Y_{nj}(z)|^4 = \sum_{j=1}^{n} \mathbb{E}\left|E_j \left(\delta_j(z)\beta_j(z) - \varepsilon_j(z)\beta_j^2(z)\frac{1}{n} \text{tr} \Sigma D_j^{-1}(z)\right)^4\right| \leq K \sum_{j=1}^{n} \left(\frac{|z|^{4E}|\delta_j(z)|^4}{v^4} + \frac{|z|^8 p^4E|\varepsilon_j(z)|^4}{v^{16}n^4}\right) \to 0,
\]
where the convergence is from Lemma 7.3. As a conclusion of the martingale CLT, the random vector in (6.7) will tend to a zero-mean Gaussian vector \((M^{(1)}(z_1), \ldots, M^{(1)}(z_w))\) with covariance function
\[
\text{Cov}(M^{(1)}(z), M^{(1)}(\tilde{z})) = \lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}_{j-1} [Y_{nj}(z)Y_{nj}(\tilde{z})], \quad z \neq \tilde{z} \in \{z_1, \ldots, z_w\}, \quad (6.10)
\]
provided these limits exit. Moreover, from the arguments on Page 571 of Bai and Silverstein (2004) and discussions on Page 439 in Bai and Zhou (2008), we have also
\[
\mathbb{E}|\beta_j(z) - b_n(z)| \to 0 \quad \text{and} \quad b_n(z) + z \mu_0(z) \to 0, \quad (6.11)
\]
by which the remaining task is to show that
\[
z \tilde{\mu}_0(z) \mu_0(\tilde{z}) \sum_{j=1}^{n} \mathbb{E}_{j-1} (E_j \varepsilon_j(z)E_j \varepsilon_j(\tilde{z})) \quad (6.12)
\]
converges in probability to a deterministic quantity whose second mixed partial derivative yields the limit in (6.10).

From Lemma 7.2, one may get
\[
(6.12) = 2(T_1 + \gamma_{n2}T_2 - T_3 - T_4),
\]
where
\[
T_1 = \frac{z \tilde{\mu}_0(z) \mu_0(\tilde{z})}{n^2} \sum_{j=1}^{n} \text{tr} \left[E_j \Sigma D_j^{-1}(z)E_j (\Sigma D_j^{-1}(\tilde{z}))\right],
\]
\[
T_2 = \frac{z \tilde{\mu}_0(z) \mu_0(\tilde{z})}{pn^2} \sum_{j=1}^{n} \text{tr} \left[E_j \Sigma D_j^{-1}(z)\text{tr} [E_j \Sigma D_j^{-1}(\tilde{z})]\right],
\]
\[
T_3 = \frac{z \tilde{\mu}_0(z) \mu_0(\tilde{z})}{pn^2} \sum_{j=1}^{n} \text{tr} \left[E_j \Sigma^2 D_j^{-1}(z)\text{tr} [E_j \Sigma^2 D_j^{-1}(\tilde{z})]\right],
\]
\[
T_4 = \frac{z \tilde{\mu}_0(z) \mu_0(\tilde{z})}{pn^2} \sum_{j=1}^{n} \text{tr} \left[E_j \Sigma D_j^{-1}(z)\text{tr} [E_j \Sigma^2 D_j^{-1}(\tilde{z})]\right].
\]
The quantities $T_1$ and $T_2$ have been respectively studied in Bai and Silverstein (2004) and Hu et al. (2019) under their models. Following their steps with the help of Lemma 7.3, one may obtain

$$T_1 \xrightarrow{i.p.} \int_0^\infty \frac{a(z,\hat{z})}{1 - z} dz, \quad a(z,\hat{z}) = 1 - \frac{m(z)m(\hat{z})(z - \hat{z})}{m(z) - m(\hat{z})},$$

$$T_2 \xrightarrow{i.p.} c \int \frac{tm(z)dH(t)}{1 + tm(z)} \int \frac{tm(\hat{z})dH(t)}{1 + tm(\hat{z})},$$

and thus

$$\frac{\partial^2 T_1}{\partial z \partial \hat{z}} \xrightarrow{i.p.} \frac{m'(z)m'(\hat{z})}{(m(z) - m(\hat{z}))^2} - \frac{1}{(z - \hat{z})^2},$$

$$\frac{\partial^2 T_2}{\partial z \partial \hat{z}} \xrightarrow{i.p.} \frac{1}{c} \left( m(z) + zm'(z) \right) \left( m(\hat{z}) + \hat{z}m'(\hat{z}) \right).$$

Notice that statistics $T_3$ and $T_4$ will reduce to $T_2$ if $\Sigma^2$ is replaced with $\Sigma$. Therefore, similar to the derivation of the limit of $T_2$, it’s straightforward to get

$$T_3 \xrightarrow{i.p.} c \int \frac{t^2m(z)dH_p(t)}{1 + tm(z)} \int \frac{tm(\hat{z})dH_p(t)}{1 + tm(\hat{z})},$$

$$T_4 \xrightarrow{i.p.} c \int \frac{tm(z)dH_p(t)}{1 + tm(z)} \int \frac{t^2m(\hat{z})dH_p(t)}{1 + tm(\hat{z})}.$$

Their corresponding derivatives are

$$\frac{\partial^2 T_3}{\partial z \partial \hat{z}} \xrightarrow{i.p.} \frac{1}{c} \left( \frac{m'(z)}{m^2(z)} - 1 \right) \left( m(\hat{z}) + \hat{z}m'(\hat{z}) \right),$$

$$\frac{\partial^2 T_4}{\partial z \partial \hat{z}} \xrightarrow{i.p.} \frac{1}{c} \left( \frac{m'(\hat{z})}{m^2(\hat{z})} - 1 \right) \left( m(z) + zm'(z) \right).$$

Collecting the above results, we get the covariance function in the lemma.

**Step 2: Tightness of $M_n^{(1)}(z)$.** The tightness of $M_n^{(1)}(z)$ can be established by verifying the moment condition (12.51) of Billingsley (1968):

$$\sup_{n, z_1, z_2 \in C_n} \frac{\mathbb{E}|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} < \infty.$$  \hfill (6.13)

The first task here is to control the probability of the event that extreme eigenvalues of $B_n$ falling outside of the interval $[s_l, s_r]$ defined by (6.2). This is done in Lemma 7.5. By virtue of this lemma and arguments on Page 579 in Bai and Silverstein (2004), one may assume that moments of $D^{-1}(z)$, $D_j^{-1}(z)$ and $D_{ij}^{-1}(z)$ are all bounded in $n$ and $z \in C_n$, that is, for any positive $q$,

$$\max\{\mathbb{E}|D^{-1}(z)|^q, \mathbb{E}|D_j^{-1}(z)|^q, \mathbb{E}|D_{ij}^{-1}(z)|^q\} \leq K_q.$$  \hfill (6.14)
Then using such boundedness, the inequality in Lemma 7.3 can be extended as

\[ \left| E \left[ a(v) \prod_{l=1}^{q} (y^T B_l(v) y - \text{tr } \Sigma B_l(v)) \right] \right| \leq K p^{q/2}, \quad (6.15) \]

where the matrices \( B_l(v) \) are independent of \( y \) and

\[
\max \{|a(v)|, \|B_l(v)\|\} \leq K \left[ 1 + p^* I \left( \|B_n\| \geq \eta_r \text{ or } \lambda_{\min}^B \leq \eta_k \right) \right]
\]

for some positive \( s, \eta_r \in (s_r, x_r) \) and \( \eta_k \in (x_k, s_k) \), where \( B \) is \( B_n \) or \( B_\theta \) with some \( r_j \)'s removed. Note that the inequality (6.15) is parallel to the inequality (3.2) in Bai and Silverstein (2004) for the linear transform model.

Finally, following closely the procedure in Section 3 of Bai and Silverstein (2004), and applying Lemmas 7.3 and 7.5 together with (6.14) and (6.15), one may verify (6.13). The details are thus omitted.

**Step 3: Convergence of \( M_n^{(2)} (z) \).** To finish the proof, it is enough to show that the sequence of \( M_n^{(2)} (z) \) is bounded and equicontinuous, and converges to the mean function (6.5). The boundedness and equicontinuity can be verified following the arguments on Pages 592-593 of Bai and Silverstein (2004), and we only focus on the convergence of \( M_n^{(2)} (z) \) in this step. A novel method to derive the limiting mean function is proposed, which is quite different from the idea in Bai and Silverstein (2004). This new procedure is more straightforward and easier to follow.

We first list some results that will be used in this part:

\[
\sup_{z \in C_n} \mathbb{E}|\varepsilon_j(z)|^q \leq K n^{-q/2}, \quad \sup_{z \in C_n} \mathbb{E}|\gamma_j(z)|^q \leq K n^{-q/2}, \quad \text{for } q \text{ even,} \quad (6.16)\]

\[
\sup_{n, z \in C_n} |b_n(z) + zm_{0}(z)| \to 0, \quad \sup_{n, z \in C_n} \|zI - b_n(z) \Sigma\| < \infty, \quad (6.17)\]

\[
\sup_{n, z \in C_n} \mathbb{E} |\text{tr } D^{-1}(z)|M - \text{tr } D^{-1}(z)|M|^2 \leq K \|M\|^2, \quad (6.18)\]

where \( M \) is any nonrandom \( p \times p \) matrix. These results can be verified step by step following similar discussions in Bai and Silverstein (2004) and we omit the details.

Writing \( V(z) = zI - b_n(z) \Sigma \), we decompose \( M_n^{(2)} (z) \) as

\[
M_n^{(2)} (z) = [pE m_n(z) + \text{tr } V^{-1}(z)] - [\text{tr } V^{-1}(z) + pm_0(z)] := S_n(z) - T_n(z) \quad (6.19)\]

\[
= [nE m_n(z) + nb_n(z)/z] - [nb_n(z)/z + nm_0(z)] := S_n(z) - T_n(z). \quad (6.20)\]

Notice that

\[
T_n(z) = p \int \frac{dH_p(t)}{z-b_n(z)t} - p \int \frac{dH_p(t)}{z+zm_n(z)t} \]

\[
= p \left[ b_n(z) + zm_n(z) \right] \int \frac{tdH_p(t)}{(z-b_n(z)t)(z+zm_n(z)t)} \]

\[
= c_n \int \frac{tdH_p(t)}{(z-b_n(z)t)(1+zm_n(z)t)}. \]

27
We have
\[
\frac{M^{(2)}_n(z) - S_n(z)}{M^{(2)}_n(z) - \mathcal{S}_n(z)} = \frac{T_n(z)}{\overline{T}_n(z)} = \frac{c_n}{z} \int \frac{tdH_p(t)}{(1 + \overline{w}_0(z)t)^2} + o(1),
\]
where the second identity uses the convergence in (6.17).

Our next task is to study the limits of \(S_n(z)\) and \(\mathcal{S}_n(z)\). For simplicity, we suppress the expression \(z\) in the sequel when it is served as independent variables of some functions. All expressions and convergence statements hold uniformly for \(z \in C_n\).

We first simplify the expression of \(S_n\). Using the identity \(r'_j D^{-1} - \mathbf{1}_{1} = r'_j D^{-1} - \mathbf{1}_{j1} \beta_j\), we have
\[
S_n = \mathbb{E} \text{tr}(D^{-1} + V^{-1})
\]
\[
= \mathbb{E} \text{tr} \left[ V^{-1} \left( \sum_{j=1}^{n} r_j r'_j - b_n \Sigma \right) D^{-1} \right]
\]
\[
= n\mathbb{E} \beta_1 r'_1 D^{-1} V^{-1} r_1 - b_n \mathbb{E} \text{tr} \Sigma D^{-1} V^{-1}. \quad (6.21)
\]

From (6.8) and \(\beta_1 = b_n - b_n \beta_1 \gamma_1\),
\[
\mathbb{E} \text{tr} V^{-1} \Sigma (D^{-1} - D^{-1}) = \mathbb{E} \text{tr} V^{-1} \Sigma D^{-1} r_1 r'_1 D^{-1} \beta_1
\]
\[
= b_n \mathbb{E} (1 - \beta_1 \gamma_1) r'_1 D^{-1} V^{-1} \Sigma D^{-1} r_1,
\]
where \(|\mathbb{E} \beta_1 r'_1 D^{-1} V^{-1} \Sigma D^{-1} r_1| \leq Kn^{-1/2}\). From this and (6.21), we get
\[
S_n = n\mathbb{E} \beta_1 r'_1 D^{-1} V^{-1} r_1 - b_n \mathbb{E} \text{tr} \Sigma D^{-1} V^{-1} + \frac{1}{n} b_n^2 \mathbb{E} \text{tr} D^{-1} V^{-1} \Sigma D^{-1} \Sigma + o(1).
\]

Plugging \(\beta_1 = b_n - b_n^2 \gamma_1 + b_n^3 \gamma_2^2 - \beta_1 b_n^2 \gamma_1^3\) into the first term in the above equation, we obtain
\[
\frac{n\mathbb{E} \beta_1 r'_1 D^{-1} V^{-1} r_1}{b_n^2 \mathbb{E} \text{tr} D^{-1} V^{-1} \Sigma - n b_n^2 \mathbb{E} \gamma_1 r'_1 D^{-1} V^{-1} r_1}
\]
\[
= b_n^2 \mathbb{E} \text{tr} D^{-1} V^{-1} \Sigma - n b_n^2 \mathbb{E} \gamma_1 r'_1 D^{-1} V^{-1} r_1 - n b_n^2 \mathbb{E} \gamma_1^3 r'_1 D^{-1} V^{-1} r_1.
\]
Note that, from (6.15), (6.16), and (6.18),

\[ E \gamma_1 r'_1 D_1^{-1} V^{-1} r_1 = E \left[ r'_1 D_1^{-1} r_1 - \frac{1}{n} \text{tr} D_1^{-1} \Sigma \right] \left[ r'_1 D_1^{-1} V^{-1} r_1 \right. \\
\left. - \frac{1}{n} \text{tr} D_1^{-1} V^{-1} \Sigma \right] + \frac{1}{n} \text{Cov}(\text{tr} D_1^{-1} \Sigma, \text{tr} D_1^{-1} V^{-1} \Sigma) \]

\[ = E \left[ r'_1 D_1^{-1} r_1 - \frac{1}{n} \text{tr} D_1^{-1} \Sigma \right] \left[ r'_1 D_1^{-1} V^{-1} r_1 \right. \\
\left. - \frac{1}{n} \text{tr} D_1^{-1} V^{-1} \Sigma \right] + o \left( \frac{1}{n} \right), \]

\[ E \gamma_2 r'_1 D_1^{-1} V^{-1} r_1 = E \gamma_2 \left[ r'_1 D_1^{-1} V^{-1} r_1 - \frac{1}{n} \text{tr} D_1^{-1} V^{-1} \Sigma \right] \\
+ \frac{1}{n} \text{Cov}(\gamma_2, \text{tr} D_1^{-1} V^{-1} \Sigma) + \frac{1}{n} E \gamma_2 \text{tr} D_1^{-1} V^{-1} \Sigma \\
= \frac{1}{n} E \gamma_2 \text{tr} D_1^{-1} V^{-1} \Sigma + o \left( \frac{1}{n} \right), \]

\[ E \gamma_3 r'_1 D_1^{-1} V^{-1} r_1 = o \left( \frac{1}{n} \right). \]

We thus arrive at

\[ S_n = -nb_n^2 E \left[ r'_1 D_1^{-1} r_1 - \frac{1}{n} \text{tr} D_1^{-1} \Sigma \right] \left[ r'_1 D_1^{-1} V^{-1} r_1 - \frac{1}{n} \text{tr} D_1^{-1} V^{-1} \Sigma \right] \\
+ b_n^3 E \gamma_2 \text{tr} D_1^{-1} V^{-1} \Sigma + \frac{1}{n} b_n^3 E \text{tr} D_1^{-1} \Sigma D_1^{-1} \Sigma + o(1). \]

On the other hand, by the identity \( r'_j D^{-1} = r'_j D_j^{-1} \beta_j \), we have

\[ p + z \text{tr} D^{-1} = \text{tr}(B_n D^{-1}) = \sum_{j=1}^n \beta_j r'_j D_j^{-1} r_j = n - \sum_{j=1}^n \beta_j, \]

which implies \( n \zeta m_n = - \sum_{j=1}^n \beta_j \). From this, together with \( \beta_1 = b_n - b_n^2 \gamma_1 + b_n^3 \gamma_1^2 - \beta_1 b_n^4 \gamma_1^3 \)

(6.15), we get

\[ S_n = -\frac{n}{z} E (\beta_1 - b_n) = -\frac{n}{z} b_n^3 E \gamma_1^2 + o(1). \]

Applying Lemma 7.2 to the simplified \( S_n \) and \( S_n \), and then replacing \( D_j \) with \( D \) in the
derived results yield

\[ S_n = -\frac{b_n^2}{n} \left[ E \text{tr} \ D^{-1} \Sigma D^{-1} V^{-1} \Sigma + \frac{2}{p} \left( \gamma_2 E \text{tr} \ Sigma D^{-1} \text{tr} \ Sigma D^{-1} V^{-1} \right) \right. \]

\[ \left. - E \text{tr} \Sigma^2 D^{-1} \text{tr} \Sigma D^{-1} V^{-1} - E \text{tr} \Sigma D^{-1} \text{tr} \Sigma^2 D^{-1} V^{-1} \right] \]

\[ + \frac{2b_n^3}{n^2} \left[ E \text{tr} \ D^{-1} \Sigma D^{-1} \Sigma + \frac{1}{p} \left( \gamma_2 E \text{tr} \Sigma D^{-1} \text{tr} \Sigma D^{-1} \right. \right. \]

\[ \left. \left. - 2E \text{tr} \Sigma^2 D^{-1} \text{tr} \Sigma D^{-1} \right) \right] E \text{tr} D^{-1} V^{-1} \Sigma + o(1), \quad (6.22) \]

\[ \Sigma_n = -\frac{2b_n^3}{zn} \left[ E \text{tr} \ D^{-1} \Sigma D^{-1} \Sigma + \frac{1}{p} \left( \gamma_2 E \text{tr} \Sigma D^{-1} \text{tr} \Sigma D^{-1} + \right. \right. \]

\[ \left. \left. - 2E \text{tr} \Sigma^2 D^{-1} \text{tr} \Sigma D^{-1} \right) \right] + o(1). \quad (6.23) \]

To study the limits of \( S_n \) and \( \Sigma_n \) in (6.22) and (6.23), we need to figure out the difference between \( D^{-1} \) and \( V^{-1} \). Write

\[ D^{-1} + V^{-1} = b_n \tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3, \quad (6.24) \]

where

\[ \tilde{R}_1 = \sum_{j=1}^{n} V^{-1} (r_j r_j' - n^{-1} \Sigma) D_j^{-1}, \]

\[ \tilde{R}_2 = \sum_{j=1}^{n} V^{-1} r_j r_j' D_j^{-1} (\beta_j - b_n), \]

\[ \tilde{R}_3 = \frac{1}{n} \sum_{j=1}^{n} b_n V^{-1} \Sigma (D_j^{-1} - D^{-1}). \]

Similar to the arguments on page 590 in Bai and Silverstein (2004) we have, for any \( p \times p \) matrix \( M \),

\[ |E \text{tr} \tilde{R}_2 M| \leq n^{1/4} K (E\|M\|^2)^{1/4} \quad \text{and} \quad |\text{tr} \tilde{R}_3 M| \leq K (E\|M\|^2)^{1/2}. \quad (6.25) \]

In addition, for nonrandom matrix \( M \),

\[ |E \text{tr} \tilde{R}_1 M| \leq n^{1/2} K \|M\|. \quad (6.26) \]

Taking a step further, for \( M \) nonrandom, we write

\[ \text{tr} \tilde{R}_1 \Sigma D^{-1} M = \tilde{R}_{11} + \tilde{R}_{12} + \tilde{R}_{13}, \quad (6.27) \]
where

\[ \tilde{R}_{11} = \text{tr} \sum_{j=1}^{n} V^{-1} r_j r'_j D_j^{-1} \Sigma (D^{-1} - D_j^{-1}) M, \]

\[ \tilde{R}_{12} = \text{tr} \sum_{j=1}^{n} V^{-1} (r_j r'_j - n^{-1} \Sigma) D_j^{-1} \Sigma D_j^{-1} M, \]

\[ \tilde{R}_{13} = -\frac{1}{n} \text{tr} \sum_{j=1}^{n} V^{-1} \Sigma D_j^{-1} \Sigma (D^{-1} - D_j^{-1}) M. \]

It’s clear that \( E\tilde{R}_{12} = 0 \) and moreover, using (6.14), (6.15) and (6.18), we get

\[ |E\tilde{R}_{13}| \leq K ||M||, \]

(6.28)

\[ E\tilde{R}_{11} = -b_n n^{-1} E(\text{tr} D_1^{-1} \Sigma D_1^{-1} \Sigma) (\text{tr} D_1^{-1} MV^{-1} \Sigma) + o(1) \]

(6.29)

Applying (6.11), (6.24)-(6.29), one may calculate the limit of each component of \( S_n \) and \( \hat{S}_n \) in (6.22) and (6.23). Specifically, we have

\[ \frac{1}{n} E \text{tr} D^{-1} \Sigma^k = - \int \frac{c_n t^k dH_p(t)}{z(1 + m_0 t)} + o(1), \quad k = 1, 2, \]

\[ \frac{1}{n} E \text{tr} D^{-1} V^{-1} \Sigma^k = - \int \frac{c_n t^k dH_p(t)}{z^2(1 + m_0 t)^2} + o(1), \quad k = 1, 2, \]

\[ \frac{1}{n} E \text{tr} D^{-1} \Sigma D^{-1} \Sigma \]

\[ = - \frac{1}{n} E \text{tr} V^{-1} \Sigma D^{-1} \Sigma - \frac{b_n^2}{n^2} E \text{tr} D^{-1} \Sigma D^{-1} \Sigma E \text{tr} V^{-1} \Sigma D^{-1} \Sigma + o(1) \]

\[ = - \frac{1}{n} E \text{tr} V^{-1} \Sigma D^{-1} \Sigma \left[ 1 + \frac{b_n^2}{n} E \text{tr} V^{-1} \Sigma D^{-1} \Sigma \right]^{-1} + o(1), \]

\[ = \int \frac{c_n t^2 dH_p(t)}{z^2(1 + m_0 t)^2} \left[ 1 - \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right]^{-1} + o(1), \]
and
\[
\frac{1}{n} \mathbb{E} \text{tr} D^{-1} \Sigma D^{-1} V^{-1} \Sigma
\]
\[
= -\frac{1}{n} \mathbb{E} \text{tr} V^{-1} \Sigma D^{-1} V^{-1} \Sigma \left[ 1 + \frac{b_2}{n} \mathbb{E} \text{tr} D^{-1} \Sigma D^{-1} \Sigma \right] + o(1)
\]
\[
= -\frac{1}{n} \mathbb{E} \text{tr} V^{-1} \Sigma D^{-1} V^{-1} \Sigma \left[ 1 + \frac{b_2}{n} \mathbb{E} \text{tr} V^{-1} \Sigma D^{-1} \Sigma \right]^{-1} + o(1),
\]

\[
= \int \frac{c_n t^2 dH_p(t)}{z^3(1 + m_0 t)^3} \left[ 1 - \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right]^{-1} + o(1).
\]

Combining the above results with (6.22) and (6.23), we obtain
\[
S_n = -\int \frac{c_n m_0^2 t^2 dH_p(t)}{z(1 + m_0 t)} \left[ 1 - \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right]^{-1}
\]
\[
- 2 \left[ \int \frac{\gamma_2 t - t^2 dH_p(t)}{z(1 + m_0 t)} \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} - \int \frac{tdH_p(t)}{z(1 + m_0 t)} \right] \times
\]
\[
\left. \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right| + 2c_n m_0^3 \left\{ \int \frac{t^2 dH_p(t)}{z(1 + m_0 t)} \left[ 1 - \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right]^{-1} \right.
\]
\[
+ \gamma_2 \left[ \int \frac{tdH_p(t)}{1 + m_0 t} \right]^{-2} - 2 \int \frac{t^2 dH_p(t)}{1 + m_0 t} \int \frac{tdH_p(t)}{1 + m_0 t} \right\} \int \frac{tdH_p(t)}{1 + m_0 t} + o(1),
\]
\[
= 2c_n m_0^3 \left\{ \int \frac{t^2 dH_p(t)}{1 + m_0 t} \left[ 1 - \int \frac{c_n m_0^2 t^2 dH_p(t)}{(1 + m_0 t)^2} \right]^{-1} + \gamma_2 \left[ \int \frac{tdH_p(t)}{1 + m_0 t} \right]^2 \right.
\]
\[
- 2 \int \frac{t^2 dH_p(t)}{1 + m_0 t} \int \frac{tdH_p(t)}{1 + m_0 t} \right\} + o(1).
\]

Therefore we get
\[
M_n^{(2)}(z) = \frac{S_n - S_n T_n / \overline{T}_n}{1 - T_n / \overline{T}_n}
\]
\[
\rightarrow \left[ 1 - \int \frac{c t dH(t)}{z(1 + m t)^2} \right]^{-1} \left\{ \int \frac{c m^2 t^2 dH(t)}{z(1 + m t)^2} \left[ 1 - \int \frac{c m^2 t^2 dH(t)}{(1 + m t)^2} \right]^{-1} \right.
\]
\[
- \frac{2c m^2}{z} \left( \int \frac{\gamma_2 t - t^2 dH(t)}{1 + m t} \int \frac{tdH(t)}{(1 + m t)^2} - \int \frac{tdH(t)}{1 + m t} \int \frac{t^2 dH(t)}{(1 + m t)^2} \right) \},
\]
as \(n \to \infty\). Using the identity
\[
\left[ 1 - \int \frac{c t dH(t)}{z(1 + m t)^2} \right]^{-1} = -z m \left[ 1 - \int \frac{c m^2 t^2 dH(t)}{(1 + m t)^2} \right]^{-1} = -\frac{z m'}{m}
\]
we finally obtain the mean function of the lemma.
6.3 Proof of Corollary 3.1

Choose a contour \( C \) for the integrals such that \( \max_{z \in S_H, z \in C} |\dot{m}(z)| < 1 \), where \( S_H \) is the support of \( H \). Let \( m(C) = \{ m(z) : z \in C \} \) denote the image of \( C \) under \( m(z) \). Then \( m(C) \) is a simple and closed contour having clockwise direction and enclosing zero (Qin and Li, 2017).

By the identity in (3.2), the integral in the mean function of Theorem 3.2 becomes

\[
v_j = - \frac{c}{2\pi i} \oint_{m(C)} \frac{P_j(m)P_{j,3}(m)}{m^{j+1}(1 - cm^2 P_{j,2}(m))} \, dm
\]

From this and the Cauchy integral theorem, we get the mean function. The covariance function can be obtained following the proof of Theorem 1 in Qin and Li (2017).

7 Appendix

**Lemma 7.1.** Let \( \mathbf{x} = (x_1, \ldots, x_p)' \sim N_p(0, \mathbf{T}) \) where \( \mathbf{T} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) \) is a diagonal matrix with the spectral norm \( ||\mathbf{T}|| \) bounded. Write \( r_k = \sum_{i=1}^p \sigma_i^{2k}/p, k = 1, 2 \). Then we have for \( 1 \leq i \neq j \leq p \).

\[
\begin{align*}
\mathbb{E} \left( \sum_{i=1}^p \frac{x_i^2}{x_i^2 + p} \right) &= \frac{\sigma_i^2}{r_1} + \frac{2\sigma_i^2 r_2 - 2\sigma_i^4 r_1}{pr_1^3} + o\left( \frac{1}{p} \right), \\
\mathbb{E} \left( \sum_{i=1}^p \frac{x_i^2 x_j^2}{x_i^2 + p} \right) &= \frac{\sigma_i^2 \sigma_j^2}{r_1^3} + \frac{6\sigma_i^2 \sigma_j^2 r_2 - 4\sigma_i^2 \sigma_j^2 (\sigma_i^2 + \sigma_j^2) r_1}{pr_1^3} + o\left( \frac{1}{p} \right), \\
\mathbb{E} \left( \sum_{i=1}^p \frac{x_i^4}{x_i^2 + p^2} \right) &= \frac{3\sigma_i^4}{r_1^3} + \frac{18\sigma_i^4 r_2 - 24\sigma_i^6 r_1}{pr_1^3} + o\left( \frac{1}{p} \right).
\end{align*}
\]

**Lemma 7.2.** Let \( \mathbf{y} = \sqrt{p} \mathbf{x}/||\mathbf{x}|| \) where \( \mathbf{x} \) is as defined in Lemma 7.1 such that \( \mathbb{E}(\mathbf{y}\mathbf{y}') = \mathbf{\Sigma} \). For any \( p \times p \) complex matrices \( \mathbf{C} \) and \( \hat{\mathbf{C}} \) with bounded spectral norms,

\[
\begin{align*}
\mathbb{E}(\mathbf{y}' \mathbf{C} \mathbf{y} - \text{tr} \, \mathbf{\Sigma} \mathbf{C}) (\mathbf{y}' \hat{\mathbf{C}} \mathbf{y} - \text{tr} \, \mathbf{\Sigma} \hat{\mathbf{C}}) &= \text{tr} \, \mathbf{\Sigma} \mathbf{C} \mathbf{C}' + \text{tr} \, \mathbf{\Sigma} \mathbf{C} \hat{\mathbf{C}} + \frac{2}{p} \left( \gamma_2 \text{tr} \, \mathbf{\Sigma} \mathbf{C} \text{tr} \, \mathbf{\Sigma} \hat{\mathbf{C}} - \text{tr} \, \mathbf{\Sigma}^2 \mathbf{C} \text{tr} \, \mathbf{\Sigma} \hat{\mathbf{C}} \right) + o(p),
\end{align*}
\]

33
where \( \gamma_2 = \text{tr} \Sigma^2 / p \).

Lemmas 7.1 and 7.2 are from elementary calculations, similar to the proofs of the main theorems in Zou et al. (2014).

**Lemma 7.3.** For any \( p \times p \) complex matrix \( C \) and \( y = \sqrt{p}x / ||x|| \) with \( x \sim N(0, \Sigma) \) and \( ||\Sigma|| \leq 1 \),
\[
\mathbb{E} |y'C y - \text{tr} \Sigma C|^q \leq K_q ||C||^q p^{q/2}, \quad q \geq 2,
\]
(7.1)
where \( K_q \) is a positive constant depending only on \( q \).

**Proof.** This lemma follows from Lemma 2.2 in Bai and Silverstein (2004) and similar arguments in the proof of Lemma 5 in Gao et al. (2016).

**Lemma 7.4** (Theorem 35.12 of Billingsley (1995)). Suppose for each \( n Y_{n1}, Y_{n2}, \ldots Y_{nr_n} \) is a real martingale difference sequence with respect to the increasing \( \sigma \)-field \( \{\mathcal{F}_{nj}\} \) having second moments. If for each \( \varepsilon > 0 \),
\[
\sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^2 I(|Y_{nj}| \geq \varepsilon)) \to 0 \quad \text{and} \quad \sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^2 |\mathcal{F}_{nj-1}) \xrightarrow{\text{i.p.}} \sigma^2,
\]
as \( n \to \infty \), where \( \sigma^2 \) is a positive constant, then
\[
\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{D} N(0, \sigma^2).
\]

**Lemma 7.5.** Suppose that Assumptions (a)-(c) hold. Then, for any \( s \) positive,
\[
P( ||B_n|| > \eta_r ) = o(n^{-s}),
\]
whenever \( \eta_r > \limsup_{p \to \infty} ||\Sigma||(1 + \sqrt{c})^2 \). If \( 0 < \liminf_{p \to \infty} \lambda_{\min}^{\Sigma} I_{(0, 1)}(c) \) then,
\[
P(\lambda_{\min}^{B_n} < \eta_l ) = o(n^{-s}),
\]
whenever \( 0 < \eta_l < \liminf_{p \to \infty} \lambda_{\min}^{\Sigma} I_{(0, 1)}(1 - \sqrt{c})^2 \).

**Proof.** Let \( x_j = Az_j \) where \( AA' = T \) and \( z_j \sim N(0, I) \), \( j = 1, \ldots, n \). Also let \( B_n^{(0)} = (1/n) \sum_{j=1}^{n} Az_j z_j' A' \). From Bai and Silverstein (2004), the conclusions of this lemma hold when \((B_n, \Sigma)\) are replaced with \((B_n^{(0)}, T)\). Choose \( \eta_r^{(0)} \) and \( \eta_l^{(0)} \) satisfying
\[
\eta_l < r_1^{-1} \eta_l^{(0)} < \liminf_{p \to \infty} \lambda_{\min}^{\Sigma} I_{(0, 1)}(c)(1 - \sqrt{c})^2,
\]
\[
\limsup_{p \to \infty} ||\Sigma||(1 + \sqrt{c})^2 < r_1^{-1} \eta_r^{(0)} < \eta_r.
\]
where \( r_1 = \text{tr}(T)/p \). From Lemma 7.1, we have
\[
\eta^{(0)}_1 < \liminf_{p \to \infty} \lambda_{\min}^T I(0,1)(c)(1 - \sqrt{c})^2 \quad \text{and} \quad \limsup_{p \to \infty} ||T||(1 + \sqrt{c})^2 < \eta^{(0)}_1.
\]

Using inequalities
\[
\min_{1 \leq j \leq n} \frac{p}{||A z_j||^2} \lambda_{\min}^{B_n} \leq \lambda_{\min} \leq \max_{1 \leq j \leq n} \frac{p}{||A z_j||^2} ||B^{(0)}_n||,
\]
we may get
\[
P(||B_n|| > \eta r) \\
\leq P \left( ||B^{(0)}_n|| > \eta r \right) + P \left( \max_{1 \leq j \leq n} \frac{p}{||A z_j||^2} ||B^{(0)}_n|| > \eta r, ||B^{(0)}_n|| \leq \eta r \right) \\
\leq P \left( \max_{1 \leq j \leq n} \frac{p}{||A z_j||^2} > \frac{\eta r}{\eta r} \right) + o(n^{-s}) \\
\leq n P \left( \left| \frac{||A z_1||^2}{p} - r_1 \right| > r_1 - \frac{\eta r}{\eta r} \right) + o(n^{-s}),
\]
where the last equality is from the Chebyshev inequality and the fact \( r_1 > \eta r / \eta r \). Similarly, \( P(\lambda_{\min} < \eta r) = o(n^{-s}) \).

\[\square\]

References

Bai, Z. D., Chen, J. Q., and Yao, J. F. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. Aust. N. Z. J. Stat. 52, 423–437.

Bai, Z. D., Hu, J., Pan, G. M., and Zhou, W. (2015). Convergence of the empirical spectral distribution function of Beta matrices. Bernoulli, 21, 1538–1574.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab., 32, 553–605.

Bai, Z. D. and Silverstein, J. W. (2010). Spectral analysis of large dimensional random matrices, 2nd ed., Springer, New York.

Bai, Z. D. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. Statist. Sinica, 18, 425–442.
BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed., Wiley, New York.

BOENTE, G. AND FRAIMAN, R. (1999). Comment on “Robust principal component analysis for functional data”. *Test*, 8, 28–35.

CARDOSO, J. F. (1989). Source separation using higher order moments. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-89)*, 2109–2112.

CHAKRABORTY, A. AND CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.*, 45, 771–799.

DONOHOO, D. L. AND GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20, 1803–1827.

DÜRRE, A. AND VOGEL, D. (2016). Asymptotics of the two-stage spatial sign correlation. *J. Multivariate Anal.*, 144, 54–67.

DÜRRE, A., VOGEL, D., AND TYLER D.E. (2014). The spatial sign covariance matrix with unknown location. *J. Multivariate Anal.*, 130, 107–117.

DÜRRE, A., VOGEL, D., AND FRIED, R. (2015). Spatial sign correlation. *J. Multivariate Anal.*, 135, 89–105.

DÜRRE, A., R. FRIED, AND VOGEL, D. (2017). The spatial sign covariance matrix and its application for robust correlation estimation. *Austrian Journal of Statistics*, 46, 13–22.

EL KAROU, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.*, 36, 2757–2790.

EL KAROU, N. (2009). Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.*, 19, 2362–2405.

FENG, L. AND SUN, F. (2016). Spatial-sign based high-dimensional location test. *Electronic Journal of Statistics*, 10, 2420–2434.

GAO, J. T., HAN, X., PAN, G. M., AND YANG, Y. R. (2017). Large-dimensional correlation matrices: CLT and its applications. *J. Roy. Stat. Soc. Ser. B*, 79, 677-693.

GERVINI, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95, 587–600.

GUPTA, A. K., VARGA, T., BODNAR, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*, 2nd ed., Springer, New York.
Hu, J., Li, W.M., Liu, Z. and Zhou W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. Ann. Statist., 47, 527-555.

Huber, P. J. (1977). Robust Statistical Procedures. Philadelphia: Society for Industrial and Applied Mathematics.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). Robust Statistics: The Approach Based on Influence Functions. New York: Wiley.

I. Johnstone, I. (2007). High dimensional statistical inference and random matrices. In International Congress of Mathematicians. Vol. I, pages 307–333. Eur. Math. Soc., Zürich.

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. J. Multivariate Anal., 12, 1–38.

Li Y., Wang, Z. and Zou, C. (2016). A simpler spatial-sign-based two-sample test for high-dimensional data. J. Multivariate Anal., 149, 192–198.

Li, W. M. and Yao, J. F. (2014). A local moment estimator of the spectrum of a large dimensional covariance matrix. Statist. Sinica, 24, 919–936.

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., and Cohen, K. L. (1999). Robust principal component for functional data. Test, 8, 1–73.

Li, W. M. and Yao, J. F. (2018). On structure testing for component covariance matrices of a high-dimensional mixture. J. Roy. Stat. Soc. Ser. B, 80, 293–318.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. Mat. Sb. (N.S.), 72, 507–536.

Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. Ann. Statist. 4, 51–67.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). Robust Statistics: Theory and Methods. John Wiley & Sons Ltd., Chichester.

Magyar A. F., Tyler D. E. (2014). The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. Biometrika, 101, 673–688.

Nica, A. and Speicher, R. (2006). Lectures on the Combinatorics of Free Probability. Cambridge University Press, New York.

Oja, H. (2010). Multivariate Nonparametric Methods with R. An approach based on spatial signs and ranks. Springer-Verlag, New York.

Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. Ann. Appl. Probab., 18, 1232–1270.

37
Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli*, **22**, 1745–1769.

Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, **150**, 1–29.

Qin, Y. L. and Li, W. M. (2016). Testing the order of a population spectral distribution for high-dimensional data. *Computational Statistics & Data Analysis*, **95**, 75–82.

Qin, Y. L. and Li, W. M. (2017). “Bias-reduced estimators of moments of a population spectral distribution and their applications”, In *Big and Complex Data Analysis: Statistical Methodologies and Applications* (Ejaz Ahmed ed.), Springer.

Rao, N. R., Mingo, J. A., Speicher, R., and Edelman, A. (2008). Statistical eigen-inference from large Wishart matrices. *Ann. Statist.*, **36**, 2850–2885.

Rousseeuw, P. (1985). *Multivariate estimation with high breakdown point*. In: *Mathematical Statistics and Applications*, Volume B (Proc. 4th Pannonian Symp. Math. Statist., Bad Tatzmannsdorf, 1983). Dordrecht: Reidel, pp. 283–97.

Sirkia, S., Taskinen, S., Oja, H. and Tyler, D.E. (2009). Tests and estimates of shape based on spatial signs and ranks. *Journal of Nonparametric Statistics*, **21**, 155–176.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, **55**, 331–339.

Taskinen, S., Sirkiä, S. and Oja, H. (2010). K-step shape estimators based on spatial signs and ranks. *Journal of Statistical Planning and Inference*, **140**, 3376–3388.

Taskinen, S., Koch, I. and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statistics and Probability Letters*, **82**, 765–774.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, **15**, 234–51.

Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, **91**, 557–575.

Zou, C. L., Peng, L. H., Feng, L., and Wang, Z. J. (2014). Multivariate-sign-based high-dimensional tests for sphericity. *Biometrika*, **101**, 229–236.