Supplemental material for "Prior distributions for structured semi-orthogonal matrices"

Michael Jauch[1], Marie-Christine Düker[2], and Peter Hoff[3]

[1]Department of Statistics, Florida State University

[2]Department of Data Science, FAU Erlangen-Nürnberg

[3]Department of Statistical Science, Duke University

1 Notation

For the reader's convenience, notation used throughout the supplemental material is collected here. The maximum and minimum eigenvalues of a symmetric matrix A are denoted by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. Analogously, we write $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ for the largest and smallest singular values. The spectral norm of a matrix A is given by $\|A\| = \sqrt{\lambda_{\max}(A^\top A)}$ and the Frobenius norm as $\|A\|_F = \sqrt{\mathrm{Tr}(A^\top A)}$. The maximum absolute column sum or 1-norm $\|A\|_1 = \max_{j=1,\dots,p} \sum_{i=1}^p |A_{ij}|$. We say that a random variable X is bounded almost surely when there is a $M \geq 0$ such that $\mathrm{P}(|X| > M) = 0$, i.e., the set $\{\omega \mid |X(\omega)| > M\}$ has probability zero. For a sequence of random variables $(\xi_n)_{n\geq 1}$ and a deterministic sequence $(c_n)_{n\geq 1}$, we write $\xi_n = O_{a.s.}(c_n)$ to indicate that ξ_n/c_n is almost surely bounded for large enough n. We write $X \sim \mu$ to indicate that the random variable X was drawn from a distribution μ. Finally, $\overset{d}{=}$ denotes equality in distribution.

2 Proofs of results in Section 3

In this section, we present proofs of the theoretical results in Section 3.

Proof of Theorem 1. Suppose that X is invariant to left multiplication by elements of $\mathcal{L}$ and right multiplication by elements of $\mathcal{R}$, and let $L \in \mathcal{L}$ and $R \in \mathcal{R}$ be given. By assumption, we know that $X \stackrel{d}{=} LXR$. Let $VDV^\top$ be the eigendecomposition of $X^\top X$. Then

$$\begin{aligned} Q_X &\stackrel{d}{=} LXR \left(R^\top X^\top L^\top LXR\right)^{-1/2} \\ &= LXR \left(R^\top X^\top XR\right)^{-1/2} \\ &= LXR \left(R^\top VDV^\top R\right)^{-1/2} \\ &= LXRR^\top VD^{-1/2}V^\top R \\ &= LXVD^{-1/2}V^\top R \\ &= LX \left(X^\top X\right)^{-1/2} R \\ &= LQ_X R. \end{aligned}$$

We conclude that Q_X is likewise invariant to left multiplication by elements of $\mathcal{L}$ and right multiplication by elements of $\mathcal{R}$. $\square$

Proof of Theorem 2. We prove that, under the conditions stated in the theorem, the Wasserstein distance (of order $\ell = 2$ with respect to the Euclidean norm) between the distribution of m entries of $\sqrt{p}Q_{X_k}$ and the distribution of the corresponding entries of X_k is $O\left(\sqrt{\frac{mk}{p}}\right)$ as $k \to \infty$ and $p = p(k) \to \infty$ with $k/p \to 0$.

Suppose that, for each k, the sets $\mathcal{M}_k^{(1)}, \ldots, \mathcal{M}_k^{(k)}$ specify the row indices of the entries to be selected from each of the k columns of $\sqrt{p}Q_{X_k}$ and X_k, as described prior to the statement of Theorem 2 in the main article. From these sets, one can construct the matrices $D_{\mathcal{M}_k^{(1)}}, \ldots, D_{\mathcal{M}_k^{(k)}}$ and the selection matrix $\Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}}$. The $m_k^{(j)}$-dimensional vector $D_{\mathcal{M}_k^{(j)}} X_k e_j$ contains those entries of the jth column of X_k specified by $\mathcal{M}_k^{(j)}$. It will also prove useful to notice that $D_{\mathcal{M}_k^{(j)}} X_k$ is the $m_k^{(j)} \times k$ matrix obtained from X_k by removing row i if $i \notin \mathcal{M}_k^{(j)}$.

The squared Wasserstein distance satisfies

$$
\begin{aligned}
& W_2\left[\text{law}\left\{\Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}} \operatorname{vec}\left(\sqrt{p}Q_{X_k}\right)\right\}, \text{law}\left\{\Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}} \operatorname{vec}\left(X_k\right)\right\}\right]^2 \\
\leq & \mathrm{E}\left\|\Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}} \operatorname{vec}\left(\sqrt{p}Q_{X_k}\right) - \Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}} \operatorname{vec}\left(X_k\right)\right\|_2^2 \qquad (1) \\
= & \mathrm{E}\left\|\Delta_{\mathcal{M}_k^{(1)},\ldots,\mathcal{M}_k^{(k)}} \left\{\operatorname{vec}\left(\sqrt{p}Q_{X_k}\right) - \operatorname{vec}\left(X_k\right)\right\}\right\|_2^2 \\
= & \sum_{j=1}^{k} \mathrm{E}\left\|D_{\mathcal{M}_k^{(j)}}(\sqrt{p}\,Q_{X_k} - X_k)e_j\right\|_2^2 \\
= & \frac{1}{k}\sum_{j=1}^{k} \mathrm{E}\left\|D_{\mathcal{M}_k^{(j)}}(\sqrt{p}\,Q_{X_k} - X_k\right)\|_F^2. \qquad (2)
\end{aligned}
$$

The inequality (1) follows from Definition 3.1 of the Wasserstein distance as an infimum over couplings. The equality (2) holds because, by Lemma 1, the columns of $\sqrt{p}\,Q_{X_k} - X_k$ are exchangeable.

As $k \to \infty$ and $p = p(k) \to \infty$ with $k/p \to 0$, the Frobenius norm in (2) can be bounded above almost surely as follows (with explanations below):

$$
\begin{aligned}
\left\|D_{\mathcal{M}_k^{(j)}}\left(\sqrt{p}Q_{X_k} - X_k\right)\right\|_F &= \left\|D_{\mathcal{M}_k^{(j)}}\left\{\sqrt{p}X_k\left(X_k^\top X_k\right)^{-\frac{1}{2}} - X_k\right\}\right\|_F \\
&= \left\|D_{\mathcal{M}_k^{(j)}}X_k\left\{\left(\frac{1}{p}X_k^\top X_k\right)^{-\frac{1}{2}} - I_k\right\}\right\|_F \\
&\leq \left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F \left\|\left(\frac{1}{p}X_k^\top X_k\right)^{-\frac{1}{2}} - I_k\right\| \qquad (3) \\
&= \left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F \left\|\sum_{i=1}^{\infty}\binom{-1/2}{i}\left(\frac{1}{p}X_k^\top X_k - I_k\right)^i\right\| \qquad (4) \\
&\leq \left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F \sum_{i=1}^{\infty}\left|\binom{-1/2}{i}\right|\left\|\frac{1}{p}X_k^\top X_k - I_k\right\|^i \qquad (5) \\
&\leq \frac{1}{2}\left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F \sum_{i=1}^{\infty}\left\|\frac{1}{p}X_k^\top X_k - I_k\right\|^i \qquad (6) \\
&= \frac{1}{2}\left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F \frac{\left\|\frac{1}{p}X_k^\top X_k - I_k\right\|}{1 - \left\|\frac{1}{p}X_k^\top X_k - I_k\right\|} \qquad (7)
\end{aligned}
$$

$$= \frac{1}{2} \left\| D_{\mathcal{M}_k^{(j)}} X_k \right\|_F \frac{1}{\frac{1}{\left\| \frac{1}{p} X_k^\top X_k - I_k \right\|} - 1}$$

$$\le \frac{1}{2} \left\| D_{\mathcal{M}_k^{(j)}} X_k \right\|_F \frac{1}{\frac{1}{c\sqrt{\frac{k}{p}}} - 1} \tag{8}$$

$$= \frac{1}{2} \left\| D_{\mathcal{M}_k^{(j)}} X_k \right\|_F \frac{1}{c^{-1}\sqrt{\frac{p}{k}} - 1}. \tag{9}$$

The inequality (3) follows from the fact that $\|AB\|_F \le \|A\|\|B\|_F$ for conformable matrices A, B. The equality (4) comes from the Taylor expansion

$$\left(\frac{1}{p} X_k^\top X_k \right)^{-\frac{1}{2}} = \sum_{i=0}^{\infty} \binom{-1/2}{i} \left(\frac{1}{p} X_k^\top X_k - I_k \right)^i. \tag{10}$$

For a discussion of Taylor expansions of matrix functions, we refer the reader to Section 4.3 of Higham [9] and Theorem 4.7 therein. The Taylor expansion (10) converges provided that the spectral radius of $\frac{1}{p} X_k^\top X_k - I_k$ is less than one. Lemma 2 ensures this condition is satisfied almost surely as $k, p \to \infty$ with $k/p \to 0$. Because the Taylor series converges, we can apply the triangle inequality to get (5). The spectral norm is sub-multiplicative, so we can also exchange the norm and exponent. The inequality (6) holds because $|\binom{-1/2}{i}| \le 1/2$ for all integers $i \ge 1$. We arrive at (7) by evaluating the geometric series. We know the series converges because Lemma 2 guarantees that $\|\frac{1}{p} X_k^\top X_k - I_k\| < 1$ almost surely as $k, p \to \infty$ with $k/p \to 0$. Lemma 2 also ensures that there exists a constant $c > 0$ such that (8) holds almost surely for sufficiently large k and p.

Combining (2) and (9), the Wasserstein distance can be bounded above almost surely for sufficiently large k and p:

$$W_2 \left[\text{law} \left\{ \Delta_{\mathcal{M}_k^{(1)}, \ldots, \mathcal{M}_k^{(k)}} \operatorname{vec} \left(\sqrt{p} Q_{X_k} \right) \right\}, \text{law} \left\{ \Delta_{\mathcal{M}_k^{(1)}, \ldots, \mathcal{M}_k^{(k)}} \operatorname{vec} \left(X_k \right) \right\} \right]^2$$

$$\le \frac{1}{k} \sum_{j=1}^{k} \mathrm{E} \left\| D_{\mathcal{M}_k^{(j)}} \left(\sqrt{p}\, Q_{X_k} - X_k \right) \right\|_F^2$$

$$\le \frac{1}{k} \sum_{j=1}^{k} \mathrm{E} \left\{ \frac{1}{4} \left\| D_{\mathcal{M}_k^{(j)}} X_k \right\|_F^2 \left(c^{-1} \sqrt{\frac{p}{k}} - 1 \right)^{-2} \right\}$$

$$
\begin{aligned}
&= \frac{1}{4k\left(c^{-1}\sqrt{\frac{p}{k}}-1\right)^2}\sum_{j=1}^{k}\mathrm{E}\left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F^2 \\
&= \frac{m}{4\left(c^{-1}\sqrt{\frac{p}{k}}-1\right)^2}.
\end{aligned} \tag{11}
$$

Equation (11) comes from an explicit calculation of the expected value

$$
\begin{aligned}
\sum_{j=1}^{k}\mathrm{E}\left\|D_{\mathcal{M}_k^{(j)}}X_k\right\|_F^2 &= \sum_{j=1}^{k}\mathrm{E}\left[\mathrm{Tr}\left\{X_k^\top D_{\mathcal{M}_k^{(j)}}^\top D_{\mathcal{M}_k^{(j)}}X_k\right\}\right] \\
&= \sum_{j=1}^{k}\mathrm{E}\left[\mathrm{Tr}\left\{Z_k^\top \Omega_p^{\frac{1}{2}} D_{\mathcal{M}_k^{(j)}}^\top D_{\mathcal{M}_k^{(j)}}\Omega_p^{\frac{1}{2}} Z_k\right\}\right] \\
&= \sum_{j=1}^{k}\sum_{i=1}^{k}\mathrm{E}\left[\mathrm{Tr}\left\{z_i^\top \Omega_p^{\frac{1}{2}} D_{\mathcal{M}_k^{(j)}}^\top D_{\mathcal{M}_k^{(j)}}\Omega_p^{\frac{1}{2}} z_i\right\}\right] \qquad (12) \\
&= \sum_{j=1}^{k}\sum_{i=1}^{k}\mathrm{Tr}\left\{\Omega_p^{\frac{1}{2}} D_{\mathcal{M}_k^{(j)}}^\top D_{\mathcal{M}_k^{(j)}}\Omega_p^{\frac{1}{2}}\right\} = k\sum_{j=1}^{k} m_k^{(j)} = km.
\end{aligned}
$$

In (12), z_i indicates the ith column of Z_k. From (11), we see that the Wasserstein distance is $O\left(\sqrt{\frac{mk}{p}}\right)$ almost surely as $k, p \to \infty$ with $k/p \to 0$.

□

3 Technical lemmas and their proofs

Lemma 1. *In the setting of Theorem 2, the columns of $\sqrt{p}Q_{X_k} - X_k$ are exchangeable for all k.*

Proof. The columns of $\sqrt{p}Q_{X_k} - X_k$ are exchangeable if and only if $\sqrt{p}Q_{X_k} - X_k$ is invariant to right multiplication by permutation matrices. We know from the construction of Section 3 that the columns of X_k are exchangeable or, equivalently, X_k is invariant to right multiplication by permutation matrices. To prove the lemma, it is sufficient to establish that invariance of X_k to right multiplication by permutation matrices implies invariance of $\sqrt{p}Q_{X_k} - X_k$ to right multiplication by permutation matrices. The proof of this fact is similar to the proof of Theorem 1. Let $\mathcal{R}_k \subset \mathcal{O}(k)$ denote the set of $k \times k$ permutation matrices and let $R \in \mathcal{R}_k$ be given. We know that

$X_k \stackrel{d}{=} X_k R$. Let $V_k D_k V_k^\top$ be the eigendecomposition of $X_k^\top X_k$. Then

$$\begin{aligned}
\sqrt{p} Q_{X_k} - X_k &= \sqrt{p} X_k \left(X_k^\top X_k\right)^{-1/2} - X_k \\
&\stackrel{d}{=} \sqrt{p} X_k R \left\{(X_k R)^\top X_k R\right\}^{-1/2} - X_k R \\
&= \sqrt{p} X_k R \left(R^\top X_k^\top X_k R\right)^{-1/2} - X_k R \\
&= \sqrt{p} X_k R \left(R^\top V_k D_k V_k^\top R\right)^{-1/2} - X_k R \\
&= \sqrt{p} X_k R R^\top V_k D_k^{-1/2} V_k^\top R - X_k R \\
&= \sqrt{p} X_k V_k D_k^{-1/2} V_k^\top R - X_k R \\
&= \sqrt{p} X_k \left(X_k^\top X_k\right)^{-1/2} R - X_k R \\
&= \sqrt{p} Q_{X_k} R - X_k R \\
&= (\sqrt{p} Q_{X_k} - X_k) R.
\end{aligned}$$

We conclude that $\sqrt{p} Q_{X_k} - X_k$ is invariant to right multiplication by elements of $\mathcal{R}_k$. Thus, its columns are exchangeable. □

The following lemma is based on Lemma 6.1 in Qiu et al. [15]. Theorem 4 in Chen and Pan [5] gives a more general statement for the case $\Omega_p = I_p$. In our case, it is sufficient to require $\mathrm{E}\,|Z_{ij}|^{4+\delta} < \infty$ for some $\delta > 0$. In particular, it is not necessary to assume $\mathrm{E}\,|Z_{ij}|^{6+\varepsilon_0} < \infty$. All conditions on Ω_p in Theorem 2 are naturally satisfied for $\Omega_p = I_p$. We note that the statement involves almost surely boundedness, indicated by "a.s.", and with respect to the underlying probability measure P generating the infinite i.i.d. array $(Z_{ij})_{i,j\geq 1}$. For each k and $p = p(k)$, the matrix Z_k is formed from the first $p \times k$ block of this array. Thus, the almost sure statement is relative to this single probability space, and does not depend on k or p.

Lemma 2. *Suppose the conditions of Theorem 2 are satisfied. Then,*

$$\left\| \frac{1}{p} X_k^\top X_k - I_k \right\| = O_{a.s.}\left(\sqrt{\frac{k}{p}}\right)$$

as $k \to \infty$ and $p = p(k) \to \infty$ with $k/p \to 0$.

Proof. Recall $c_\Omega(p) := p^{-1} \operatorname{Tr}\left(\Omega_p^2\right)$ from Condition 2. We first prove a high probability bound for $\sqrt{\frac{p}{c_\Omega(p)k}} \left\| \frac{1}{p} X_k^\top X_k - I_k \right\|$ with subsequent application of the Borel-Cantelli lemma to infer almost sure convergence.

Recall that $X_k = \Omega_p^{\frac{1}{2}} Z_k$ with Z_k having i.i.d. entries. With $c_\Omega(p) := p^{-1} \operatorname{Tr}\left(\Omega_p^2\right)$, we set

$$\eta = 2 \limsup_{p\to\infty} \|\Omega_p\| \frac{1}{\sqrt{c_\Omega(p)}}$$

and

$$\begin{aligned} O_k &= \sqrt{\frac{p}{c_\Omega(p)k}} \left(\frac{1}{p} X_k^\top X_k - I_k - \operatorname{diag}\left(\frac{1}{p} X_k^\top X_k - I_k \right) \right) \\ &= \sqrt{\frac{p}{c_\Omega(p)k}} \left(\frac{1}{p} Z_k^\top \Omega_p Z_k - \operatorname{diag}\left(\frac{1}{p} Z_k^\top \Omega_p Z_k \right) \right), \end{aligned} \tag{13}$$

where $\operatorname{diag}(A)$ for a matrix A refers to the matrix A with all off-diagonal entries set to zero. With (13), we can write

$$\sqrt{\frac{p}{c_\Omega(p)k}} \left(\frac{1}{p} X_k^\top X_k - I_k \right) = O_k + \sqrt{\frac{p}{c_\Omega(p)k}} \operatorname{diag}\left(\frac{1}{p} X_k^\top X_k - I_k \right).$$

We follow the proof of Lemma 6.1 in [15]. Note that their results suffice for us to prove that for any $\varepsilon > 0$,

$$\mathrm{P}\left(\sqrt{\frac{p}{c_\Omega(p)k}} \left\| \frac{1}{p} X_k^\top X_k - I_k \right\| > \varepsilon \right) = o(k^{-1}).$$

In contrast, [5] prove that, for any $\varepsilon > 0$ and $l > 0$,

$$\mathrm{P}\left(\sqrt{\frac{p}{k}} \left\| \frac{1}{p} Z_k^\top Z_k - I_k \right\| > \sqrt{\frac{k}{p}} \right) = o(k^{-l}).$$

We aim to combine the arguments in [5] and [15] to get a slightly stronger statement needed to also infer almost sure convergence. To be more precise, for any $\varepsilon > 0$, $\ell > 1$,

$$\mathrm{P}\left(\sqrt{\frac{p}{c_\Omega(p)k}} \left\| \frac{1}{p} X_k^\top X_k - I_k \right\| > \varepsilon \right) = o(k^{-\ell}).$$

As in the proof of Lemma 6.1 in equations (S2.42), (S2.43), and (S2.44) in [15], it suffices to show that for any $\varepsilon > 0, \ell > 1$,

$$\mathrm{P}\left(\|O_k\| > \eta + \varepsilon\right) = o(k^{-\ell}), \tag{14}$$

$$\mathrm{P}\left(\frac{1}{\sqrt{kpc_\Omega(p)}} \max_{1\le r\le k}\left|\sum_{s=1}^{p}(Z_{sr}^2 - 1)\right| > \varepsilon\right) = o(k^{-\ell}), \tag{15}$$

$$\mathrm{P}\left(\frac{1}{\sqrt{kpc_\Omega(p)}} \max_{1\le r\le k}\left|\sum_{s\neq t}^{p}(\Omega_p)_{st} Z_{sr} Z_{ts}\right| > \varepsilon\right) = o(k^{-\ell}), \tag{16}$$

where $\Omega_p = ((\Omega_p)_{st})_{s,t=1,\dots,p}$ and noting that $(\Omega_p)_{ss} = 1$. We consider the three relations separately. First, (14) is inferred from equation (8) in [5]. Furthermore, (15) follows by equation (9) in [5]. Finally, (16) can be inferred by following (S2.47) in [15]. To be more precise, we rephrase (S2.47) here which gives, for any $\delta > 0$,

$$\begin{aligned}
&\mathrm{P}\left(\frac{1}{\sqrt{kpc_\Omega(p)}} \max_{1\le r\le k}\left|\sum_{s\neq t}^{p}(\Omega_p)_{st} Z_{sr} Z_{ts}\right| > \varepsilon\right) \\
&\le k(\varepsilon\sqrt{kpc_\Omega(p)})^{-(4+\delta)}\left(\mathrm{E}\,|Z_{11}|^{4+\delta}\right)^2 (\mathrm{Tr}(\Omega_p^2))^{2+\delta/2} \\
&\le kp^{2+\delta/2}(\varepsilon kpc_\Omega(p))^{-(2+\delta/2)}\left(\mathrm{E}\,|Z_{11}|^{4+\delta}\right)^2 (c_\Omega(p))^{2+\delta/2} \\
&= \varepsilon^{-(2+\delta/2)} k^{-(1+\delta/2)}\left(\mathrm{E}\,|Z_{11}|^{4+\delta}\right)^2.
\end{aligned}$$

Finally, we can infer, for some constant c,

$$\sum_{k=1}^{\infty}\mathrm{P}\left(\left\|\frac{1}{p}X_k^\top X_k - I_k\right\| > \sqrt{\frac{kc_\Omega(p)}{p}}\varepsilon\right) \le c\sum_{k=1}^{\infty}k^{-\ell} < \infty.$$

Therefore, by the Borel–Cantelli lemma,

$$\left\|\frac{1}{p}X_k^\top X_k - I_k\right\| = O_{a.s.}\left(\sqrt{\frac{kc_\Omega(p)}{p}}\right).$$

Finally, Condition 2 ensures that $c_\Omega(p)$ remains bounded away from zero and infinity, which yields

$$\left\|\frac{1}{p}X_k^\top X_k - I_k\right\| = O_{a.s.}\left(\sqrt{\frac{k}{p}}\right).$$

□

4 Proofs of results in Section 4

Proof of Proposition 1. Proposition 1 presents a Gaussian scale mixture representation of the density (3) from the main text for $\ell \in (0,1)$. To arrive at such a representation, we follow Andrews and Mallows [1].

The first theorem of Andrews and Mallows [1] provides a necessary and sufficient condition for the symmetric density (3) to admit a Gaussian scale mixture representation, i.e. for there to exist independent random variables U_{ij} and V_{ij} with U_{ij} standard normal such that $Z_{ij} = U_{ij}/V_{ij}$ has density (3). We must have

$$\left(-\frac{d}{dy}\right)^k f\left(y^{1/2} \mid \ell\right) \geq 0$$

for $y > 0$. Substituting in the explicit definition of $f(\cdot \mid \ell)$, the condition becomes

$$\left(-\frac{d}{dy}\right)^k \frac{(\ell/2)^{\ell/2}}{\Gamma(\ell/2)} y^{(\ell-1)/2} \exp\left(-\ell y/2\right) \geq 0$$

for $y > 0$. In other words, we need to show that the function $y \mapsto y^{(\ell-1)/2}\exp\left(-\ell y/2\right)$ is completely monotonic [14]. From the examples in (1.2) of Miller and Samko [14], we can conclude that the functions $y \mapsto y^{(\ell-1)/2}$ and $y \mapsto \exp\left(-\ell y/2\right)$ are completely monotonic for $\ell \in (0,1)$. Theorem 1 of Miller and Samko [14] allows us to conclude that the function $y \mapsto y^{(\ell-1)/2}\exp\left(-\ell y/2\right)$ is completely monotonic when $\ell \in (0,1)$ because it is a product of completely monotonic functions.

Now that we know a Gaussian scale mixture representation exists, we turn to finding the density of V_{ij}. The existence of the Gaussian scale mixture representation implies, for all non-zero z, that

$$f(z \mid \ell) = (2\pi)^{-1/2} \int_0^\infty v \exp\left(-v^2 z^2/2\right) \, dG(v)$$

where G is the distribution function of V_{ij}. This representation and the relation (2.2) in Andrews and Mallows [1] imply that the Laplace transform of G is

$$\begin{aligned}
\int_0^\infty \exp\left(-sv\right) \, dG(v) &= 2\int_0^\infty f(z \mid \ell) \exp\left(-s^2 z^{-2}/2\right) \, dz \\
&= 2\frac{(\ell/2)^{\ell/2}}{\Gamma(\ell/2)} \int_0^\infty z^{\ell-1} \exp\left[-\left(\ell z^2 + s^2/z^2\right)/2\right] \, dz \\
&= \frac{(\ell/2)^{\ell/2}}{\Gamma(\ell/2)} \int_0^\infty y^{\ell/2-1} \exp\left[-\left(\ell y + s^2/y\right)/2\right] \, dy && (17) \\
&= \frac{(\ell/2)^{\ell/2}}{\Gamma(\ell/2)} \frac{2K_{\ell/2}\left(\sqrt{\ell s^2}\right)}{\left(\ell/s^2\right)^{\ell/4}} && (18) \\
&= \frac{2^{1-\ell/2}\ell^{\ell/4}}{\Gamma\left(\ell/2\right)} s^{\ell/2} K_{\ell/2}\left(\sqrt{\ell}s\right) \\
&= \frac{2^{1-\ell/2}\ell^{\ell/4}}{\Gamma\left(\ell/2\right)} s^{\ell/2} K_{-\ell/2}\left(\sqrt{\ell}s\right) && (19) \\
&= \int_1^\infty \exp\left(-\sqrt{\ell}st\right) \cdot \frac{2\left(t^2-1\right)^{-\frac{1+\ell}{2}}}{B\left(\frac{1-\ell}{2}, \frac{\ell}{2}\right)} \, dt && (20) \\
&= \int_{\sqrt{\ell}}^\infty \exp\left(-s\tilde{v}\right) \cdot \frac{2\ell^{-1/2}\left(\tilde{v}^2/\ell - 1\right)^{-\frac{1+\ell}{2}}}{B\left(\frac{1-\ell}{2}, \frac{\ell}{2}\right)} \, d\tilde{v}. && (21)
\end{aligned}$$

The equality (17) comes from changing variables to $y = z^2$, while (18) comes from recognizing that the integrand in (17) is proportional to a generalized inverse Gaussian density [12]. The equality (19) is an application of 10.27.3 from DLMF [7], while (20) is an application of 10.32.8 from DLMF [7]. Finally, (21) is the result of changing variables to $\tilde{v} = \sqrt{\ell}t$. We can conclude from this derivation that the random variable V_{ij} has density

$$v \mapsto \frac{2\ell^{-1/2}\left(v^2/\ell - 1\right)^{-\frac{1+\ell}{2}}}{B\left(\frac{1-\ell}{2}, \frac{\ell}{2}\right)} \mathbb{1}_{\left(\sqrt{\ell},\infty\right)}(v). \tag{22}$$

We now define a random variable $W_{ij} = V_{ij}^2/\ell - 1$. One use can use the density (22) and a change of variables to show that $W_{ij} \sim \text{BetaPrime}\left(\frac{1-\ell}{2}, \frac{\ell}{2}\right)$; see [11]. By well-known properties of the beta prime and beta distributions, we have that that $W_{ij} \overset{d}{=} (1-\theta_{ij})/\theta_{ij}$ where $\theta_{ij} \sim \text{Beta}\left(\frac{\ell}{2}, \frac{1-\ell}{2}\right)$. Proposition 1 follows from observing that

$$Z_{ij} = \frac{U_{ij}}{V_{ij}} = \frac{U_{ij}}{\sqrt{\ell\,(W_{ij}+1)}} \overset{d}{=} \frac{U_{ij}}{\sqrt{\ell\left(\frac{1-\theta_{ij}}{\theta_{ij}}+1\right)}} = U_{ij}\sqrt{\frac{\theta_{ij}}{\ell}}.$$

□

Proof of Proposition 2. **Part I.** Both of the claims in Part I are immediate.

Part II. The claim follows from the definition of the density (3) from the main text and basic properties of limits.

Part III. That the entries of Z are i.i.d. follows by assumption. We need to verify that a random variable Z_{ij} with density (3) has a mean of zero, a variance of one, and a finite fourth moment. There are two cases: $\ell = 1$ and $\ell \in (0,1)$. In the former case, Z_{ij} is a standard normal random variable and we can immediately conclude that Z_{ij} has a mean of zero, a variance of one, and a finite fourth moment.

In the latter case, in which $\ell \in (0,1)$, Proposition 1 establishes that

$$Z_{ij}\,|\,\theta_{ij} \overset{\text{ind.}}{\sim} \text{Normal}(0, \theta_{ij}/\ell) \qquad \text{where} \qquad \theta_{ij} \overset{\text{i.i.d.}}{\sim} \text{Beta}[\ell/2, (1-\ell)/2]\,.$$

By the law of total expectation,

$$\mathrm{E}(Z_{ij}) = \mathrm{E}[\mathrm{E}(Z_{ij} \mid \theta_{ij})] = 0.$$

Again by the law of total expectation, we may compute for any $r > 0$:

$$\mathrm{E}\left(|Z_{ij}|^r\right) = \mathrm{E}\left[\mathrm{E}\left(|Z_{ij}|^r \mid \theta_{ij}\right)\right]$$

$$
\begin{aligned}
&= \int \left[\int \frac{|z|^r}{\sqrt{2\pi\theta/\ell}} \exp\left(-\frac{z^2}{2\theta/\ell}\right)\, dz \right] \frac{\theta^{\ell/2-1}(1-\theta)^{(1-\ell)/2-1}}{B[\ell/2,(1-\ell)/2]}\, d\theta \\
&= \int C_r\, (\theta/\ell)^{r/2}\, \frac{\theta^{\ell/2-1}(1-\theta)^{(1-\ell)/2-1}}{B[\ell/2,(1-\ell)/2]}\, d\theta \\
&= C_r\, \ell^{-r/2} \int \theta^{r/2}\, \frac{\theta^{\ell/2-1}(1-\theta)^{(1-\ell)/2-1}}{B[\ell/2,(1-\ell)/2]}\, d\theta \\
&= C_r\, \ell^{-r/2}\, \mathrm{E}\left(\theta_{ij}^{r/2}\right), \qquad (23)
\end{aligned}
$$

where $C_r = \mathrm{E}(|N(0,1)|^r)$ is the rth absolute moment of the standard normal, which is finite for all $r > 0$. Since $\theta_{ij} \in (0,1)$ almost surely, its $r/2$th moment is finite for every $r > 0$.

In particular, plugging in $r = 2$ yields

$$
\mathrm{Var}(Z_{ij}) = \mathrm{E}(|Z_{ij}|^2) = 1,
$$

and $r = 4$ yields

$$
\mathrm{E}(|Z_{ij}|^4) = 1 + 2/\ell < \infty.
$$

More generally, (23) shows that $\mathrm{E}(|Z_{ij}|^{4+\delta}) < \infty$ for every $\delta > 0$.

Part IV. The claim follows from a change of variables.

□

5 Discussion of Condition 2

In this section, we verify Condition 2 for different examples of correlation matrices given in Proposition 3.

Proof of Proposition 3. Recall that we omit the dependence of Ω on p in the statement of Proposition 3. In particular, we set $\Omega = (\Omega_{ij})_{i,j=1,\dots,p}$.

Part I For the first property in Condition 2, note that

$$
\begin{aligned}
\frac{1}{p}\operatorname{Tr}(\Omega^2) &= \frac{1}{p}\sum_{i,j=1}^{p}\Omega_{ij}^2 = \frac{1}{p}\sum_{i,j=1}^{p}\rho^{2|i-j|} \\
&= \sum_{|k|<p}\Big(1-\frac{k}{p}\Big)\rho^{2|k|} \to \sum_{k=-\infty}^{\infty}\rho^{2|k|}
\end{aligned}
$$

as $p \to \infty$. For the second property in Condition 2, we get, with explanations given below,

$$
\begin{aligned}
\|\Omega\| \le \|\Omega\|_1 &= \max_{j=1,\dots,p}\sum_{i=1}^{p}|\rho^{|i-j|}| \\
&= \max_{j=1,\dots,p}\sum_{k=j-p}^{j-1}|\rho^{|k|}| \\
&\le \max_{j=1,\dots,p}\int_{j-p-1}^{j-1}\rho^{|x|}dx \\
&= 2b\max_{i=1,\dots,p}\int_{j-p-1}^{j-1}\frac{1}{2b}\exp\left[-\frac{|x|}{b}\right]dx && (24)\\
&\le 2b = 2\frac{1}{\log(1/\rho)}, && (25)
\end{aligned}
$$

where (24) follows since $\rho^{|x|} = \exp(|x|\log(\rho)) = \exp(-|x|\log(1/\rho))$ and setting $b = 1/\log(1/\rho)$. Finally, (25) is due to integrating the density of the Laplace distribution over the real line.

Part II For the first property in Condition 2, note that

$$
\begin{aligned}
\frac{1}{p}\operatorname{Tr}(\Omega^2) &= \frac{1}{p}\sum_{i,j=1}^{p}\Omega_{ij}^2 = \frac{1}{p}\sum_{i,j=1}^{p}\exp\left[-(i-j)^2/\rho^2\right] \\
&= \frac{1}{p}\sum_{|k|<p}\Big(1-|k|\Big)\exp\left[-k^2/\rho^2\right] \\
&= \sum_{|k|<p}\Big(1-\frac{|k|}{p}\Big)\exp\left[-k^2/\rho^2\right] \to \sum_{k=-\infty}^{\infty}\exp\left[-k^2/\rho^2\right], && (26)
\end{aligned}
$$

as $p \to \infty$. The infinite series in (26) is also known as the theta function and is finite.

For the second property in Condition 2, we get

$$\|\Omega\| \leq \|\Omega\|_1 = \max_{j=1,\ldots,p} \sum_{i=1}^{p} \left| \exp\left[-\frac{1}{2}(i-j)^2/\rho^2\right]\right| \tag{27}$$

$$= \max_{j=1,\ldots,p} \sum_{i=j-p}^{j-1} \left| \exp\left[-\frac{1}{2}k^2/\rho^2\right]\right|$$

$$\leq \max_{j=1,\ldots,p} \int_{j-p-1}^{j-1} \exp\left[-\frac{1}{2}x^2/\rho^2\right] dx$$

$$= \sqrt{2\pi\rho^2} \max_{j=1,\ldots,p} \int_{j-p-1}^{j-1} \frac{1}{\sqrt{2\pi\rho^2}} \exp\left[-\frac{1}{2}x^2/\rho^2\right] dx \leq \sqrt{2\pi\rho^2}, \tag{28}$$

where (27) follows since the operator norm can be bounded by the maximum absolute column sum. In (28), we bounded the integral by the density of a Gaussian random variable with zero mean and variance ρ^2 over the real line.

Part III For the first property in Condition 2, note that

$$\frac{1}{p}\operatorname{Tr}(\Omega^2) = \frac{1}{p}\sum_{i,j=1}^{p} \Omega_{ij}^2 = \frac{1}{p}\sum_{i,j=1}^{p} C_\nu^2(|i-j|)$$

$$= 2\sum_{k=1}^{p-1}\Big(1-\frac{k}{p}\Big)C_\nu^2(k)+1$$

$$= \frac{2^{2-2\nu}}{\Gamma^2(\nu)} 2\sum_{k=1}^{p-1}\Big(1-\frac{k}{p}\Big)\left(\sqrt{2\nu}\frac{|k|}{\rho}\right)^{2\nu} K_\nu^2\left(\sqrt{2\nu}\frac{|k|}{\rho}\right)+1$$

$$\to \frac{2^{2-2\nu}}{\Gamma^2(\nu)} 2\sum_{k=1}^{\infty}\left(\sqrt{2\nu}\frac{k}{\rho}\right)^{2\nu} K_\nu^2\left(\sqrt{2\nu}\frac{k}{\rho}\right)+1<\infty,$$

where the infinite series is finite since $K_\nu(x) \sim \sqrt{\frac{\pi}{2}}x^{-\frac{1}{2}}e^{-x}$ for $\nu > 0$ and as $x \to \infty$ by (A.5) in Gaunt [8]; and since

$$\frac{\pi}{2}\frac{2^{2-2\nu}}{\Gamma^2(\nu)}\sum_{k=1}^{\infty}\left(\sqrt{2\nu}\frac{k}{\rho}\right)\exp\left(-2\sqrt{2\nu}\frac{k}{\rho}\right) = -\frac{\pi}{2}\frac{2^{2-2\nu}}{\Gamma^2(\nu)}e^2/(-1+e^2)^2 < \infty.$$

For the second property in Condition 2, we get, with explanations given below,

$$
\begin{aligned}
\|\Omega\| \leq \|\Omega\|_1 &= \max_{j=1,\ldots,p} \sum_{i=1}^{p} C_\nu(|i-j|) \\
&= \max_{j=1,\ldots,p} \sum_{k=j-p}^{j-1} C_\nu(|k|) \\
&\leq \max_{j=1,\ldots,p} \int_{j-p-1}^{j-1} C_\nu(|x|)dx && (29) \\
&\leq \int_{-\infty}^{\infty} \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{|x|}{\rho}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|x|}{\rho}\right) dx \\
&= \frac{2^{1-\nu}}{\Gamma(\nu)} \frac{1}{\sqrt{2\nu}} \int_{-\infty}^{\infty} \left(\frac{|x|}{\rho}\right)^{\nu} K_\nu\left(\frac{|x|}{\rho}\right) dx \\
&= \sqrt{2}\rho\sqrt{\frac{\pi}{2}} 2^{1-\nu} \frac{\Gamma(\nu+\frac{1}{2})}{\Gamma(\nu)} \frac{1}{\sqrt{2\nu}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}\rho} \frac{1}{\Gamma(\nu+\frac{1}{2})} \sqrt{\frac{2}{\pi}} \left(\frac{|x|}{\rho}\right)^{\nu} K_\nu\left(\frac{|x|}{\rho}\right) dx && (30) \\
&= \sqrt{2}\rho\sqrt{\frac{\pi}{2}} 2^{1-\nu} \frac{\Gamma(\nu+\frac{1}{2})}{\Gamma(\nu)} \frac{1}{\sqrt{2\nu}}. && (31)
\end{aligned}
$$

The inequality (29) is due to the function $x \mapsto C_\nu(|x|)$ being monotonically decreasing. The monotonicity follows since $\frac{d}{dx}x^\nu K_\nu(x) = -x^\nu K_{\nu-1}(x)$; see for example Baricz [3]. Furthermore, $K_\nu(x) > 0$ for all $x > 0$ and $\nu \in \mathbb{R}$ as stated in Appendix A.1. in Gaunt [8]. Due to equation (4.1.32) in Kotz et al. [13] with the parameters therein chosen as $\sigma = \sqrt{2}\rho$, $\tau = \nu + \frac{1}{2}$, the expression in the integral in (30) is a density. We therefore infer (31).

Part IV The banded correlation matrix trivially satisfies Condition 2 since the respective correlation functions satisfy Condition 2. □

6 The Wasserstein distance between $\sqrt{p}Q_X$ and X

Proposition 1 below provides a simple, non-asymptotic expression for the squared Wasserstein distance between the distributions of $\sqrt{p}Q_X$ and X. This distance can be written in terms of the eigenvalues of a sample covariance matrix that has been the subject of several recent articles in the random matrix theory literature.

Proposition 1. *The squared Wasserstein distance between the law of X and the law of $\sqrt{p}Q_X$ can be written as*

$$W_2\left[\operatorname{law}\left(\sqrt{p}\,Q_X\right), \operatorname{law}\left(X\right)\right]^2 = 2p\,\mathrm{E}\left[\sum_{j=1}^{k}(1-\lambda_j^{1/2}(S))\right]$$

where $S = p^{-1}X^\top X = p^{-1}Z^\top \Omega Z$ and $\lambda_j^{1/2}(S)$ denotes the square root of the jth largest eigenvalue of S.

A number of articles have studied a re-normalized version of the sample covariance matrix S as $p, k \to \infty$ with $k/p \to 0$. In the case where Ω is the identity, Bai and Yin [2] proved that the spectral distribution of the re-normalized sample covariance matrix converges to the semicircle law, while Chen and Pan [5] investigated the behavior of the largest eigenvalue. For a general correlation matrix Ω, Wang and Paul [16] derived conditions under which the spectral distribution of the re-normalized sample covariance matrix converges to the semicircle law, while Chen and Pan [6] and Qiu et al. [15] established central limit theorems for the linear spectral statistics.

Proof of Proposition 1. By Lemma 3 below, we have

$$W_2\left[\operatorname{law}\left(\sqrt{p}\,Q_X\right), \operatorname{law}\left(X\right)\right]^2 = \mathrm{E}\,\|\sqrt{p}\,Q_X - X\|_F^2 \,.$$

The squared Frobenius norm can be written as

$$\begin{aligned}\|\sqrt{p}\,Q_X - X\|_F^2 &= \operatorname{Tr}\left[\left(\sqrt{p}\,Q_X - X\right)^\top\left(\sqrt{p}\,Q_X - X\right)\right] \\ &= \operatorname{Tr}\left[p\,Q_X^\top Q_X\right] - 2\operatorname{Tr}\left[\sqrt{p}\,Q_X^\top X\right] + \operatorname{Tr}\left[X^\top X\right].\end{aligned}$$

We evaluate the three terms and their expectations separately and then combine the results. In terms of Z and Ω, the matrix Q_X is given as

$$Q_X = X(X^\top X)^{-1/2}$$

$$= \Omega^{1/2} Z (Z^\top \Omega Z)^{-1/2}.$$

Thus, the first term is

$$\begin{aligned}
\mathrm{Tr}\left[p\, Q_X^\top Q_X\right] &= p\,\mathrm{Tr}\left[(Z^\top \Omega Z)^{-1/2} Z^\top \Omega^{1/2}\Omega^{1/2} Z (Z^\top \Omega Z)^{-1/2}\right] \\
&= p\,\mathrm{Tr}\left[(Z^\top \Omega Z)^{-1} Z^\top \Omega Z\right] \\
&= p\,\mathrm{Tr}\left[I_k\right] \\
&= pk.
\end{aligned}$$

Then, we have $\mathrm{E}\,\mathrm{Tr}\left[p\, Q_X^\top Q_X\right] = pk$. The third term is $\mathrm{Tr}\left[X^\top X\right] = \mathrm{Tr}\left[Z^\top \Omega Z\right]$. Let Z_j denote the jth column of Z. Then,

$$\begin{aligned}
\mathrm{E}\,\mathrm{Tr}\left[Z^\top \Omega Z\right] &= \mathrm{E}\left[\sum_{j=1}^k Z_j^\top \Omega Z_j\right] \\
&= k\,\mathrm{E}\left[\sum_{j=1}^k Z_1^\top \Omega Z_1\right] \\
&= k\,\mathrm{Tr}\,\Omega \\
&= pk.
\end{aligned}$$

The second term is

$$\begin{aligned}
-2\,\mathrm{Tr}\left[\sqrt{p}\, Q_X^\top X\right] &= -2\,\mathrm{Tr}\left[\sqrt{p}\left(Z^\top \Omega Z\right)^{-1/2} Z^\top \Omega^{1/2}\Omega^{1/2} Z\right] \\
&= -2\,\mathrm{Tr}\left[\left(p^{-1} Z^\top \Omega Z\right)^{-1/2} Z^\top \Omega Z\right] \\
&= -2p\,\mathrm{Tr}\left[\left(p^{-1} Z^\top \Omega Z\right)^{-1/2} p^{-1} Z^\top \Omega Z\right] \\
&= -2p\,\mathrm{Tr}\left[\left(p^{-1} Z^\top \Omega Z\right)^{1/2}\right] \\
&= -2p\,\mathrm{Tr}\left[S^{1/2}\right]
\end{aligned}$$

where $S = p^{-1}Z^\top \Omega Z$. Let $\lambda_1 \geq \cdots \geq \lambda_k \geq 0$ be the eigenvalues of S. Putting this all together, we have that

$$\begin{aligned}
\mathrm{E}\,\|\sqrt{p}\,Q_X - X\|_F^2 &= 2pk - 2p\,\mathrm{E}\,\mathrm{Tr}\left[S^{1/2}\right] \\
&= 2p\,\mathrm{E}\,\mathrm{Tr}\left[I_k - S^{1/2}\right] \\
&= 2p\,\mathrm{E}\left[\sum_{j=1}^{k}(1 - \lambda_j^{1/2})\right].
\end{aligned}$$

□

As observed in Higham and Schreiber [10], the polar decomposition yields the nearest unitary matrix in the Frobenius norm. Lemma 3 provides a formal argument that the squared Wasserstein distance (defined in Section 3) between the distributions of X and $\sqrt{p}\,Q_X$ (also defined in Section 3) is equal to the expected value of their squared Frobenius distance.

Lemma 3. *Recall that $X \in \mathbb{R}^{p\times k}$ and $Q_X = X(X^\top X)^{-1/2}$. The squared Wasserstein distance satisfies*

$$W_2\left[\mathrm{law}\left(\sqrt{p}\,Q_X\right), \mathrm{law}\left(X\right)\right]^2 = \mathrm{E}\,\|\sqrt{p}\,Q_X - X\|_F^2\,. \tag{32}$$

Proof. In order to show that (32) is indeed satisfied, we use Brenier's theorem (see Brenier [4]) to derive that the optimal transport map between X and $\sqrt{p}X(X^\top X)^{-1/2}$ is given by

$$T(X) = \sqrt{p}X(X^\top X)^{-1/2}.$$

By Brenier's theorem, for the two random matrices X (source) and $Y = \sqrt{p}X(X^\top X)^{-1/2}$ (target), the squared Wasserstein distance is equal to

$$W_2^2(\mu,\nu) = \inf_{T:T_{\#}\mu=\nu}\int_{\mathbb{R}^{p\times k}} \|X - T(X)\|_F^2\,d\mu(X),$$

where $T_{\#}\mu = \nu$ means that the map T pushes the distribution μ (of X) to the distribution ν (of Y). Therefore, since by Higham and Schreiber [10], the polar decomposition yields the nearest unitary matrix in the Frobenius norm, (32) is satisfied.

While not necessary for the proof, we note as an additional remark that Brenier's Theorem states further that if μ and ν are absolutely continuous measures with finite second moments, then the optimal transport map T exists and is unique, given by the partial derivative of a convex function ϕ through

$$T(X) = \nabla\phi(X).$$

For the transport map $T(X) = \sqrt{p}X(X^\top X)^{-1/2}$ we have

$$\phi(X) = \frac{\sqrt{p}}{2}\operatorname{Tr}(X(X^\top X)^{-1/2}X^\top).$$

In particular, the gradient of ϕ with respect to X satisfies,

$$\nabla\phi(X) = \sqrt{p}X(X^\top X)^{-1/2} = T(X).$$

Since the function $(X^\top X)^{-1/2}$ involves an inverse matrix square root, which is smooth and convex over positive semidefinite matrices and since the trace operation and scaling preserve convexity, ϕ is a valid convex potential function, and $T(X) = \nabla\phi(X)$ satisfies Brenier's conditions. $\square$

References

[1] Andrews, D. F. and Mallows, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 36(1):99–102, 1974.

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

[3] Baricz, Á. Bounds for modified Bessel functions of the first and second kinds. *Proceedings of the Edinburgh Mathematical Society*, 53(3):575–599, 2010.

[4] Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[5] Chen, B. and Pan, G. Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli*, 18(4):1405–1420, 2012.

[6] Chen, B. and Pan, G. CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli*, 21(2):1089–1133, 2015.

[7] DLMF. *NIST Digital Library of Mathematical Functions*. `https://dlmf.nist.gov/`, Release 1.2.3 of 2024-12-15. URL `https://dlmf.nist.gov/`. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[8] Gaunt, R. E. Inequalities for modified Bessel functions and their integrals. *Journal of Mathematical Analysis and Applications*, 420(1):373–386, 2014.

[9] Higham, N. J. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[10] Higham, N. J. and Schreiber, R. S. Fast polar decomposition of an arbitrary matrix. *SIAM Journal on Scientific and Statistical Computing*, 11(4):648–655, 1990.

[11] Johnson, N. L., Kotz, S., and Balakrishnan, N. *Continuous univariate distributions, volume 2*, volume 289. John Wiley & Sons, 1995.

[12] Jørgensen, B. *Statistical properties of the generalized inverse Gaussian distribution*, volume 9. Springer Science & Business Media, 1982.

[13] Kotz, S., Kozubowski, T., and Podgórski, K. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Number 183. Springer Science & Business Media, 2001.

[14] Miller, K. S. and Samko, S. G. Completely monotonic functions. *Integral Transforms and Special Functions*, 12(4):389–402, 2001.

[15] Qiu, J., Li, Z., and Yao, J. Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when and applications. *The Annals of Statistics*, 51(3):1427–1451, 2023.

[16] Wang, L. and Paul, D. Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *Journal of Multivariate Analysis*, 126(C):25–52, 2014.