Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models

David Morales-Jimenez[1] Iain M. Johnstone[2] Matthew R. McKay[3] Jeha Yang[2]

[1] *ECIT Institute, Queen's University Belfast, UK*

[2] *Department of Statistics, Stanford University, USA*

[3] *ECE Department, Hong Kong University of Science and Technology, Hong Kong*

Supplementary Material

In this supplementary document, proofs are provided for some of the results presented in the main document. Specifically, proofs for the Gaussian particularizations of our main results (Corollaries 1 and 2) are given in Section S1. The proof for the instrumental tightness properties of Lemma 3 is included in Section S2. Finally, proofs for the key asymptotic properties of normalized bilinear forms (Lemma 1 and Proposition 1) are given in Section S3.

S1 Gaussian particularizations

S1.1 Proof of Corollary 1

Refer to definition (1.4) of $\check{\kappa}$, and introduce also

$$\check{\kappa}_{1,iji'j'} = \mathrm{Cov}(\psi_{ij}, \psi_{i'j'}), \qquad \check{\kappa}_{2,iji'j'} = \mathrm{Cov}(\psi_{ij}, \chi_{i'j'}).$$

Note that the centering terms in covariances $\check{\kappa}_1$ and $\check{\kappa}_2$ are respectively $\mathbb{E}\psi_{ij}\mathbb{E}\psi_{i'j'}$ and $\mathbb{E}\psi_{ij}\mathbb{E}\chi_{i'j'}$, and that both equal $\kappa_{ij}\kappa_{i'j'}$. Observe also that the sum

$$[\mathcal{P}^{\nu}, \check{\kappa}_2] = \sum_{i,j,i',j'} \mathcal{P}^{\nu}_{iji'j'} \mathrm{Cov}(\psi_{ij}, \chi_{i'j'})$$

is unchanged by swapping the indices (ij) with $(i'j')$, so

$$[\mathcal{P}^\nu, \check{\kappa}] = [\mathcal{P}^\nu, \check{\kappa}_1] - 2[\mathcal{P}^\nu, \check{\kappa}_2]\ . \tag{S1.1}$$

Using definition (1.3) of ψ_{ij} and setting $\mu_{iji'j'} = \mathbb{E}[\bar{\xi}_i\bar{\xi}_j\bar{\xi}_{i'}\bar{\xi}_{j'}]$,

$$\sum_{i,j,i',j'} \mathcal{P}^\nu_{iji'j'}\mathbb{E}(\psi_{ij}\psi_{i'j'}) = \frac{1}{4}\sum_{i,j,i',j'} \mathcal{P}^\nu_{iji'j'}\kappa_{ij}\kappa_{i'j'}\mathbb{E}(\bar{\xi}_i^2+\bar{\xi}_j^2)(\bar{\xi}_{i'}^2+\bar{\xi}_{j'}^2) = \sum_{i,j,i',j'} \mathcal{P}^\nu_{iji'j'}\kappa_{ij}\kappa_{i'j'}\mu_{iii'i'},$$

since the latter sum is unaffected by replacing i with j, and i' with j'.

Now we can insert the centering in $\check{\kappa}_1$, and argue in a parallel manner for $\check{\kappa}_2$, to obtain

$$[\mathcal{P}^\nu, \check{\kappa}_1] = \sum_{i,j,i',j'} \mathcal{P}^\nu_{iji'j'}\kappa_{ij}\kappa_{i'j'}(\mu_{iii'i'}-1) = \ell_\nu^2\sum_{i,i'}(p_{\nu,i})^2(p_{\nu,i'})^2(\mu_{iii'i'}-1),$$

$$[\mathcal{P}^\nu, \check{\kappa}_2] = \sum_{i,j,i',j'} \mathcal{P}^\nu_{iji'j'}\kappa_{ij}(\mu_{iii'j'}-\kappa_{i'j'}) = \ell_\nu\sum_{i,i',j'}(p_{\nu,i})^2p_{\nu,i'}p_{\nu,j'}(\mu_{iii'j'}-\kappa_{i'j'}),$$

where in each final equality we used the summation device for indices occurring exactly twice; for example, if j appears twice, $\sum_j \kappa_{ij}p_{\nu,j} = \ell_\nu p_{\nu,i}$.

To this point, no Gaussian assumption was used. If the data is Gaussian, $\mu_{iji'j'} = \kappa_{ij}\kappa_{i'j'} + \kappa_{ii'}\kappa_{jj'} + \kappa_{ij'}\kappa_{ji'}$, and in particular $\mu_{iii'i'} - 1 = 2\kappa_{ii'}^2$, and $\mu_{iii'j'} - \kappa_{i'j'} = 2\kappa_{ii'}\kappa_{ij'}$. We then obtain

$$[\mathcal{P}^\nu, \check{\kappa}_1] = 2\ell_\nu^2\sum_{i,i'}(p_{\nu,i}\kappa_{ii'}p_{\nu,i'})^2 = 2\ell_\nu^2\,\mathrm{tr}\,(P_{D,\nu}\Gamma P_{D,\nu})^2,$$

$$[\mathcal{P}^\nu, \check{\kappa}_2] = 2\ell_\nu^3\sum_{i}(p_{\nu,i})^4 = 2\ell_\nu^3\,\mathrm{tr}\,(P_{D,\nu}^4),$$

and inserting these into (S1.1) and then (2.6), we complete the proof.

S1.2 Proof of Corollary 2

If the data is Gaussian, $\kappa_{iji'j'} = 0$ and $\mathbb{E}[\bar{\xi}_i\bar{\xi}_j\bar{\xi}_{i'}\bar{\xi}_{j'}] = \kappa_{ij}\kappa_{i'j'} + \kappa_{ii'}\kappa_{jj'} + \kappa_{ij'}\kappa_{ji'}$. Hence,

$$(\tilde{\Sigma}_\nu)_{kl} = \dot{\rho}_\nu^{-1}\ell_k\ell_\nu\,\delta_{k,l} + [\mathcal{P}^{k\nu l\nu}, \check{\kappa}], \tag{S1.2}$$

and, from the definition of $\breve{\kappa}$ in (1.3)-(1.4), we obtain

$$\breve{\kappa}_{iji'j'} = \frac{1}{2}\kappa_{ij}\kappa_{i'j'}\big(\kappa_{ii'}^2 + \kappa_{jj'}^2 + \kappa_{ij'}^2 + \kappa_{i'j}^2\big) - \kappa_{i'j'}\big(\kappa_{ii'}\kappa_{ji'} + \kappa_{ij'}\kappa_{jj'}\big) - \kappa_{ij}\big(\kappa_{ii'}\kappa_{ij'} + \kappa_{i'j}\kappa_{jj'}\big)\,. \tag{S1.3}$$

With this, we evaluate $[\mathcal{P}^{k\nu l\nu}, \breve{\kappa}]$, where we need to deal with terms of two types: involving products such as $\kappa_{ij}\kappa_{i'j'}\kappa_{ii'}^2$ and such as $\kappa_{i'j'}\kappa_{ii'}\kappa_{ji'}$. For the first type, we use the same summation device as in Section S1.1, i.e., if for example index j occurs exactly twice, $\sum_j \kappa_{ij}p_{\nu,j} = \ell_\nu p_{\nu,i}$, so that

$$\sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{i'j'}\kappa_{ii'}^2 = \ell_\nu^2 \sum_{i,i'} p_{k,i}(p_{\nu,i}\kappa_{ii'}^2 p_{\nu,i'})p_{l,i'} = \ell_\nu^2\, p_k^T P_{D,\nu}(\Gamma \circ \Gamma) P_{D,\nu} p_l = \ell_\nu^2 \mathcal{Z}_{kl},$$

where, recall $P_{D,\nu} = \operatorname{diag}(p_{\nu,1}, \ldots, p_{\nu,m})$. For the second type of terms, by the same device,

$$\sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{i'j'}\kappa_{ii'}\kappa_{ji'} = \ell_k\ell_\nu^2 \sum_{i'} p_{k,i'}p_{\nu,i'}^2 p_{l,i'} = \ell_k\ell_\nu^2\, p_k^T P_{D,\nu}^2 p_l = \ell_k\ell_\nu^2 \mathcal{Y}_{kl}.$$

The rest of terms are evaluated similarly, yielding

$$\sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{i'j'}\kappa_{jj'}^2 = \ell_k\ell_l \mathcal{Z}_{kl}, \qquad \sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{i'j'}\kappa_{ij'}\kappa_{jj'} = \ell_k\ell_l\ell_\nu \mathcal{Y}_{kl},$$

$$\sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{i'j'}\kappa_{ij'}^2 = \ell_\nu\ell_l \mathcal{Z}_{kl}, \qquad \sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{ii'}\kappa_{ij'} = \ell_l\ell_\nu^2 \mathcal{Y}_{kl},$$

$$\sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{i'j'}\kappa_{i'j}^2 = \ell_\nu\ell_k \mathcal{Z}_{kl}, \qquad \sum_{i,j,i',j'} \mathcal{P}^{k\nu l\nu}_{iji'j'}\,\kappa_{ij}\kappa_{i'j}\kappa_{jj'} = \ell_k\ell_l\ell_\nu \mathcal{Y}_{kl}.$$

Combining terms according to (S1.2)-(S1.3) leads to the result of Corollary 2.

S2 Proof of Lemma 3, (5.28)–(5.30)

Tightness of $G_n(g_\rho)$ in (5.28), and *a fortiori* that of $n^{-1/2}G_n(g_\rho)$, follows from that of $\widehat{M_n}(z)$ in Gao et al. (2017, Proposition 1), itself an adaptation of Lemma 1.1 of Bai and Silverstein (2004) to the sample correlation setting, and the arguments following that Lemma. Note in particular that, with notation $x_r, \mathcal{C}, \bar{\mathcal{C}}$ from Bai and Silverstein (2004), a complex contour $\mathcal{C} \cup \bar{\mathcal{C}}$ enclosing the support of F_γ can be chosen, by taking $b_\gamma < x_r < b_\gamma + 3\epsilon$, such that $|g_\rho(z)|$ is bounded above by a constant for $z \in \mathcal{C} \cup \bar{\mathcal{C}}$.

To prove (5.29), from (5.27) it suffices to show that the matrix valued process $\{W_n(\rho) \in \mathbb{R}^{m\times m}, \rho \in I\}$ is uniformly tight. Since m stays fixed throughout, we only need to show tightness for each of the scalar processes formed from the matrix entries $e_k^T W_n(\rho) e_l$ on I.

Let $\mathbb{P}_n, \mathbb{E}_n$ denote probability and expectation conditional on the event $E_{n\epsilon} = \{\mu_1 \le b_\gamma + \epsilon\}$. We show tightness of $W_n(\rho)$ on I by establishing the moment criterion of Billingsley (1968, eq. (12.51)): we exhibit C such that for each $k, l \le m$ and $\rho, \rho' \in I$,

$$\mathbb{E}_n |e_k^T [W_n(\rho) - W_n(\rho')] e_l|^2 \le C(\rho - \rho')^2.$$

Write the quadratic form inside the expectation as $\boldsymbol{x}^T \check{B}_n \boldsymbol{y} - \kappa_{kl} \operatorname{tr} \check{B}_n$ with $\boldsymbol{x} = \bar{X}_1^T e_k$ and $\boldsymbol{y} = \bar{X}_1^T e_l$ being the k^{th} and l^{th} rows of $\bar{X}_1$ and $\check{B}_n = n^{-1/2}[B_n(\rho) - B_n(\rho')]$. Lemma 1 with $p = 2$ yields

$$\mathbb{E}_n |e_k^T [W_n(\rho) - W_n(\rho')] e_l|^2 \le 2C_2\nu_4 \mathbb{E}_n [\operatorname{tr} \check{B}_n^2 + \|n^{1/2}\check{B}_n\|^2].$$

Now $n^{1/2}\check{B}_n$ has eigenvalues $(\rho' - \rho)\mu_i(\rho' - \mu_i)^{-1}(\rho - \mu_i)^{-1}$, so that on $E_{n\epsilon}$ we have $\operatorname{tr} \check{B}_n^2 \le \|n^{1/2}\check{B}_n\|^2 \le C(\rho' - \rho)^2$, which establishes the moment condition.

To establish (5.30), we work conditionally on $E_{n\epsilon}$. The tightness just established yields, for given ϵ, a value M for which the event E_n' defined by

$$\sup_{\rho \in I} n^{1/2} \|K(\rho) - K_0(\rho; \gamma_n)\| > \tfrac{1}{2} M$$

has $\mathbb{P}_n$-probability at most ϵ. For all large enough n such that $b_\gamma + 3\epsilon > (1 + \sqrt{\gamma_n})^2$, we combine this with the eigenvalue perturbation bound

$$|\lambda_\nu(\rho) - \lambda_{0\nu}(\rho)| \le \|K(\rho) - K_0(\rho; \gamma_n)\| \tag{S2.1}$$

for $\rho \in I$, where $\lambda_\nu(\rho)$ and $\lambda_{0\nu}(\rho) = -\rho \mathsf{m}(\rho; \gamma_n)\ell_\nu - \rho$ are the ν^{th} eigenvalues of $K(\rho) - \rho I_m$ and $K_0(\rho; \gamma_n) - \rho I_m$ respectively. Observe that $\lambda_{0\nu}(\rho_{\nu n}) = 0$ and

$$\partial_\rho \lambda_{0\nu}(\rho) = -1 - \ell_\nu \int x(\rho - x)^{-2} \mathsf{F}_\gamma(dx) < -1,$$

hence for $\rho_{n\pm} = \rho_{\nu n} \pm Mn^{-1/2}$, we have $\lambda_{0\nu}(\rho_{n-}) \geq Mn^{-1/2}$ and $\lambda_{0\nu}(\rho_{n+}) \leq -Mn^{-1/2}$. Now (S2.1) shows that on event $E_n'^c$, $\lambda_\nu(\rho_{n-}) \geq \frac{1}{2}Mn^{-1/2}$ and $\lambda_\nu(\rho_{n+}) \leq -\frac{1}{2}Mn^{-1/2}$. Since $\lambda_\nu(\rho)$ is continuous in ρ, there exists $\rho_{\nu *} \in (\rho_{\nu -}, \rho_{\nu +})$ such that $\lambda_\nu(\rho_{\nu *}) = 0$; note from the Schur complement decomposition

$$\det(R - \rho I_{m+p}) = \det(R_{22} - \rho I_p) \det(K(\rho) - \rho I_m)$$

that $\rho_{\nu *}$ is an eigenvalue of R. This is almost surely $\hat{\ell}_\nu$, since $\hat{\ell}_\nu, \rho_{\nu n} \xrightarrow{\text{a.s.}} \rho_\nu$, and $\rho_\nu = \rho(\ell_\nu, \gamma)$ is different from the almost sure limit of any eigenvalue of R adjacent to $\hat{\ell}_\nu$ (given by (5.20)), because ℓ_ν is simple and supercritical. Therefore, we have $\hat{\ell}_\nu \in (\rho_{n-}, \rho_{n+})$, and thus $|\hat{\ell}_\nu - \rho_{\nu n}| \leq Mn^{-1/2}$, which proves (5.30).

S3 Proofs of asymptotic properties of normalized bilinear forms

S3.1 Proof of Lemma 1 (Trace Lemma)

Lemma 1 is established by using truncation arguments, similar to Gao et al. (2017, Lemma 5), but adapted to bilinear forms instead of quadratic forms. Also, in contrast to that result, we do not consider data that is centered with the sample mean.

Let $\mathcal{C}_s$ denote a constant depending only on s, with different instances not necessarily identical. Define the events $\mathcal{E}_n^x \triangleq \{|n^{-1}\|x\|^2 - 1| \leq \epsilon\}$ and $\mathcal{E}_n^y \triangleq \{|n^{-1}\|y\|^2 - 1| \leq \epsilon\}$, for some $\epsilon \in (0, 1/2)$, and use $\bar{\mathcal{E}}_n^x$, $\bar{\mathcal{E}}_n^y$ to denote their complements. Using Markov's inequality and Burkholder inequalities for sums of martingale difference sequences (Bai and Silverstein, 2009, Lemmas 2.13), we have, for any $s \geq 1$,

$$\mathbb{P}\,[\bar{\mathcal{E}}_n^x] \leq \epsilon^{-s}\,\mathbb{E}\left|n^{-1}\|x\|^2 - 1\right|^s = (n\epsilon)^{-s}\,\mathbb{E}\left|\sum_{i=1}^n (x_i^2 - 1)\right|^s$$
$$\leq \mathcal{C}_s (n\epsilon)^{-s}\left[\left(\sum_{i=1}^n \mathbb{E}|x_i^2 - 1|^2\right)^{s/2} + \sum_{i=1}^n \mathbb{E}|x_i^2 - 1|^s\right] = O\left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right), \tag{S3.1}$$

and a bound of the same order for $\mathbb{P}\,[\bar{\mathcal{E}}_n^y]$, for the same reason. Now define $\mathcal{E}_n = \mathcal{E}_n^x \cap \mathcal{E}_n^y$ and its complement $\bar{\mathcal{E}}_n$. Then $\mathbb{P}\,[\bar{\mathcal{E}}_n] \leq \mathbb{P}\,[\bar{\mathcal{E}}_n^x] + \mathbb{P}\,[\bar{\mathcal{E}}_n^y] = O\left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right)$ by (S3.1). Also, since $1 = \mathbf{1}_{\mathcal{E}_n} + \mathbf{1}_{\bar{\mathcal{E}}_n}$(recall that $\mathbf{1}_{\mathrm{A}}$ denotes the indicator function on set A), we have

$$\mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B\right|^s = \mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B\right|^s \mathbf{1}_{\mathcal{E}_n} + \mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B\right|^s \mathbf{1}_{\bar{\mathcal{E}}_n}. \tag{S3.2}$$

We now bound the two terms on the right hand side of (S3.2). For the second term, from $|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B| \leq 2\|B\|$ and (S3.1), we have

$$\mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B\right|^s \mathbf{1}_{\bar{\mathcal{E}}_n} \leq 2^s\|B\|^s\mathbb{P}\,[\bar{\mathcal{E}}_n] = \|B\|^s O\left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right).$$

For the first term in (S3.2), use the decomposition

$$n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B = \frac{1}{\|x\|\|y\|}\left[x^T By - \rho\operatorname{tr} B\right] + \frac{\rho\operatorname{tr} B}{\|x\|\|y\|}\left[1 - n^{-1}\|x\|\|y\|\right] \triangleq a_1 + a_2,$$

and the triangle inequality to write

$$\mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\operatorname{tr} B\right|^s \mathbf{1}_{\mathcal{E}_n} \leq \mathcal{C}_s\left(\mathbb{E}|a_1|^s\mathbf{1}_{\mathcal{E}_n} + \mathbb{E}|a_2|^s\mathbf{1}_{\mathcal{E}_n}\right).$$

Noting that $\epsilon \in (0, 1/2)$, $\|x\|^2 \geq n/2$ and $\|y\|^2 \geq n/2$ on $\mathcal{E}_n$, so that

$$\mathbb{E}|a_1|^s\mathbf{1}_{\mathcal{E}_n} \leq 2^s n^{-s}\mathbb{E}\left|x^T By - \rho\operatorname{tr} B\right|^s \leq \mathcal{C}_s n^{-s}\left[\nu_{2s}\operatorname{tr} B^s + (\nu_4\operatorname{tr} B^2)^{s/2}\right],$$

where the last inequality follows from [JY, Lemma 4]. For a_2, for the same reasons and

$|\rho| \leq 1$,

$$\mathbb{E}|a_2|^s \mathbf{1}_{\mathcal{E}_n} \leq 2^s \left(n^{-1} \operatorname{tr} B\right)^s \mathbb{E}\left|1 - n^{-1}\|x\|\|y\|\right|^s \mathbf{1}_{\mathcal{E}_n} \ \leq 2^s \|B\|^s \mathbb{E}\left|1 - n^{-1}\|x\|\|y\|\right|^s \mathbf{1}_{\mathcal{E}_n}. \quad \text{(S3.3)}$$

We now show that $\mathbb{E}\left|1 - n^{-1}\|x\|\|y\|\right|^s \mathbf{1}_{\mathcal{E}_n} = O(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s})$. Note that

$$\left|1 - n^{-1}\|x\|\|y\|\right| \leq n^{-1/2}\|y\| \left|n^{-1/2}\|x\| - 1\right| + \left|n^{-1/2}\|y\| - 1\right|,$$

and that, on $\mathcal{E}_n$ and with $\epsilon \in (0, 1/2)$, we have $n^{-1/2}\|y\| \leq \sqrt{3/2}$. Therefore,

$$\begin{aligned}\mathbb{E}\left|1 - n^{-1}\|x\|\|y\|\right|^s \mathbf{1}_{\mathcal{E}_n} &\leq \mathcal{C}_s \left[\mathbb{E}\left|n^{-1/2}\|x\| - 1\right|^s + \mathbb{E}\left|n^{-1/2}\|y\| - 1\right|^s\right] \\ &\leq \mathcal{C}_s \left[\mathbb{E}\left|n^{-1}\|x\|^2 - 1\right|^s + \mathbb{E}\left|n^{-1}\|y\|^2 - 1\right|^s\right] = O\left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right),\end{aligned}$$

by the fact that $|a - 1| \leq |a^2 - 1|$ for $a \geq 0$, and (S3.1). Combining this bound with (S3.3), we obtain

$$\mathbb{E}|a_2|^s \mathbf{1}_{\mathcal{E}_n} \leq \mathcal{C}_s \|B\|^s \left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right).$$

The proof is complete after combining the different bounds and using them back in (S3.2).

S3.2 Proof of Proposition 1 (CLT)

We use the Cramer-Wold device and show for each $c \in \mathbb{R}^M$ that $c^T Z_n \xrightarrow{\mathcal{D}} N_M(0, c^T Dc)$. The proof follows a martingale CLT approach of Baik and Silverstein presented in the Appendix of Capitaine et al. (2009). While here normalized data vectors are considered, a parallel treatment for bilinear forms with un-normalized data is presented in the companion manuscript [JY, Theorem 10].

Start with a single bilinear form $\bar{x}^T B \bar{y} = \sum_{i,j} \bar{x}_i b_{ij} \bar{y}_j$ built from n vectors $(i = 1, \ldots, n)$

$$(\bar{x}_i, \bar{y}_i) = \left(\hat{\sigma}_x^{-1} x_i, \hat{\sigma}_y^{-1} y_i\right) \in \mathbb{R}^2,$$

where the zero mean i.i.d. vectors (x_i, y_i) have covariance

$$\mathrm{Cov}\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and $\hat{\sigma}_x^2 = n^{-1}\sum_{i=1}^n x_i^2$ and $\hat{\sigma}_y^2 = n^{-1}\sum_{i=1}^n y_i^2$ are the sample variances. Rewrite $\hat{\sigma}_x^2 = 1 + v_x$ with $v_x = n^{-1}\sum_{i=1}^n x_i^2 - 1 = O_p(n^{-1/2})$, and use the Taylor expansion of $f(a) = 1/\sqrt{1+a}$ around $a = 0$ to obtain

$$\hat{\sigma}_x^{-1} = 1 - \frac{1}{2}v_x + o_p(n^{-1/2}), \qquad \hat{\sigma}_y^{-1} = 1 - \frac{1}{2}v_y + o_p(n^{-1/2}). \tag{S3.4}$$

The symmetry of B allows the decomposition

$$n^{-1}\left(\bar{x}^T B\bar{y} - \rho\,\mathrm{tr}\,B\right) = n^{-1}\sum_i (\bar{x}_i\bar{y}_i - \rho)b_{ii} + \bar{x}_i S_i(\bar{y}) + \bar{y}_i S_i(\bar{x}), \tag{S3.5}$$

where $S_i(\bar{y}) = \sum_{j=1}^{i-1} b_{ij}\bar{y}_j$. The terms in the sum above are not martingale differences, since the data vectors $\bar{x}$, $\bar{y}$ are normalized to unit length. In order to apply the Baik-Silverstein argument, we aim at finding an alternative decomposition in terms of the unnormalized data vectors x, y; let us see this, term by term. For the first term, using (S3.4),

$$\begin{aligned} n^{-1}\sum_i (\bar{x}_i\bar{y}_i - \rho)b_{ii} &= (\hat{\sigma}_x\hat{\sigma}_y\, n)^{-1}\sum_i x_i y_i b_{ii} - n^{-1}\sum_i \rho\, b_{ii} \\ &= \left[1 - \frac{1}{2}(v_x + v_y) + o_p(n^{-1/2})\right] n^{-1}\sum_i x_i y_i b_{ii} - n^{-1}\sum_i \rho\, b_{ii} \\ &= n^{-1}\sum_i (x_i y_i - \rho)b_{ii} - \frac{1}{2}(v_x + v_y)\, n^{-1}\sum_i x_i y_i b_{ii} + o_p(n^{-1/2}). \end{aligned} \tag{S3.6}$$

Note that

$$n^{-1}\sum_i x_i y_i b_{ii} = n^{-1}\sum_i \rho\, b_{ii} + n^{-1}\sum_i (x_i y_i - \rho)b_{ii} = n^{-1}\sum_i \rho\, b_{ii} + O_p(n^{-1/2})$$

and recall that v_x and v_y are $O_p(n^{-1/2})$ so that, from (S3.6),

$$n^{-1}\sum_i(\bar{x}_i\bar{y}_i-\rho)b_{ii}=n^{-1}\sum_i(x_iy_i-\rho)b_{ii}-\frac{1}{2}\rho\,(n^{-1}\operatorname{tr}B)\,(x_i^2+y_i^2-2)+o_p(n^{-1/2}).$$

For the second term in (S3.5),

$$n^{-1}\sum_i\bar{x}_iS_i(\bar{y})=(\hat{\sigma}_x\hat{\sigma}_y)^{-1}\,n^{-1}\sum_i x_iS_i(y)$$

where, from the independence of x_iy_j and b_{ij} and the spectral norm bound of B,

$$n^{-1}\sum_i x_iS_i(y)=n^{-1}\sum_i\sum_{j=1}^{i-1}x_iy_jb_{ij}=O_p(n^{-1/2}).$$

This, along with the fact that v_x and v_y are $O_p(n^{-1/2})$, yield

$$\begin{aligned}n^{-1}\sum_i\bar{x}_iS_i(\bar{y})&=\left[1-\frac{1}{2}(v_x+v_y)+o_p(n^{-1/2})\right]n^{-1}\sum_i x_iS_i(y)\\&=n^{-1}\sum_i x_iS_i(y)+o_p(n^{-1/2}).\end{aligned}$$

The third term in (S3.5), $n^{-1}\sum_i\bar{y}_iS_i(\bar{x})$, is handled similarly. Altogether, we have the decomposition

$$n^{-1}\left(\bar{x}^TB\bar{y}-\rho\operatorname{tr}B\right)=n^{-1}\sum_i(x_iy_i-\rho)b_{ii}-\frac{1}{2}\rho\,(n^{-1}\operatorname{tr}B)\,(x_i^2+y_i^2-2)+x_iS_i(y)+y_iS_i(x)+o_p(n^{-1/2}),$$

where we can now apply the Baik-Silverstein argument. Specifically, in the setting of the theorem,

$$c^TZ_n=n^{-1/2}\sum_l c_l\left(\bar{x}_{l\cdot}^TB_n\bar{y}_{l\cdot}-\rho_l\operatorname{tr}B_n\right)=\sum_{i=1}^n Z_{di}+Z_{yi}+Z_{xi}+o_p(1)=\sum_{i=1}^n Z_{ni}+o_p(1),$$

where

$$\begin{aligned}\sqrt{n}\,Z_{di}&=\sum_l c_l\left[(x_{li}y_{li}-\rho_l)b_{ii}-\tfrac{1}{2}\rho_l\,(n^{-1}\operatorname{tr}B_n)\,(x_{li}^2+y_{li}^2-2)\right]=\sum_l c_l\,\mathsf{b}_i^T\mathsf{z}_{li}\\\sqrt{n}\,Z_{yi}&=\sum_l c_lx_{li}S_i(y_{l\cdot})\\\sqrt{n}\,Z_{xi}&=\sum_l c_ly_{li}S_i(x_{l\cdot})\end{aligned}$$

are martingale differences w.r.t. $\mathcal{F}_{n,i}$, the σ-field generated by B_n and $\{(x_{lj}, y_{lj}), 1 \leq l \leq M, 1 \leq j \leq i\}$. In the case of $Z_{d,i}$ we have introduced notation

$$\bar{b} = n^{-1} \operatorname{tr} B_n, \qquad \mathsf{b}_i = \begin{pmatrix} b_{ii} \\ -\bar{b} \end{pmatrix}, \qquad \mathsf{z}_{li} = \begin{pmatrix} z_{li} - \rho_l \\ w_{li} - \rho_l \end{pmatrix},$$

recalling that $z_{li} = x_{li} y_{li}$ and $w_{li} = \rho_l (x_{li}^2 + y_{li}^2)/2$.

Let $\mathbb{E}_{i-1}$ denote conditional expectation w.r.t. $\mathcal{F}_{n,i-1}$ and apply the martingale CLT. The limiting variance is found from $v^2 = \operatorname{plim} V_n^2$ with

$$V_n^2 = \sum_{i=1}^n \mathbb{E}_{i-1}[Z_{ni}^2] = V_{n,dd} + 2(V_{n,dy} + V_{n,dx}) + V_{n,yy} + V_{n,xx} + 2V_{n,xy}, \tag{S3.7}$$

where $V_{n,ab} = \sum_{i=1}^n \mathbb{E}_{i-1}[Z_{ai} Z_{bi}]$ for indices $a, b \in \{d, y, x\}$. The terms Z_{yi} and Z_{xi} are exactly as in [JY] and, therefore

$$\begin{aligned} V_{n,yy} = V_{n,xx} &\xrightarrow{p} \frac{1}{2}(\theta - \omega) c^T C^{xx} \circ C^{yy} c \\ V_{n,xy} &\xrightarrow{p} \frac{1}{2}(\theta - \omega) c^T C^{xy} \circ C^{yx} c. \end{aligned}$$

We only need to compute $V_{n,dd}$, $V_{n,dx}$ and $V_{n,dy}$. Start with $V_{n,dd} = \sum_{i=1}^n \mathbb{E}_{i-1}[Z_{di}^2]$, where

$$n\mathbb{E}_{i-1} Z_{di}^2 = \sum_{l,l'} c_l c_{l'} \mathsf{b}_i^T \mathbb{E}_{i-1}(\mathsf{z}_{li} \mathsf{z}_{li}^T) \mathsf{b}_i, \qquad \text{and} \qquad \mathbb{E}_{i-1}(\mathsf{z}_{li} \mathsf{z}_{li}^T) = \begin{pmatrix} C_{ll'}^{zz} & C_{ll'}^{zw} \\ C_{ll'}^{wz} & C_{ll'}^{ww} \end{pmatrix}$$

does not depend on i. Consequently

$$V_{n,dd} = \sum_{l,l'} c_l c_{l'} \left[(n^{-1} \sum_i b_{ii}^2) C_{ll'}^{zz} + \bar{b}^2 (C_{ll'}^{ww} - C_{ll'}^{wz} - C_{ll'}^{zw}) \right]$$

and $\operatorname{plim} V_{n,dd} = c^T (\omega K_1 + \phi K_2) c$, with K_1, K_2 given by (4.15).

Turn now to

$$\begin{aligned} V_{n,dy} &= \sum_{i=1}^n \mathbb{E}_{i-1}[Z_{di} Z_{yi}] \\ &= \sum_{l,l'} c_l c_{l'} M_{l,l'}^{(1)} \left[n^{-1} \sum_i b_{ii} S_i(y_{l\cdot}) \right] - \sum_{l,l'} c_l c_{l'} M_{l,l'}^{(2)} \left[n^{-1} \sum_i S_i(y_{l\cdot}) \right], \end{aligned} \tag{S3.8}$$

where

$$M^{(1)}_{l,l'} = \mathbb{E}[(x_l y_l - \rho_l)x_{l'}], \quad M^{(2)}_{l,l'} = \frac{1}{2}\rho_l(n^{-1}\operatorname{tr} B_n)\mathbb{E}[(x_l^2 + y_l^2 - 2)x_{l'}].$$

By [JY, Lemma 12], the two quantities between brackets in (S3.8) converge to zero in probability and, therefore, $V_{n,dy} \xrightarrow{p} 0$. Similarly, $V_{n,dx} \xrightarrow{p} 0$. Combining terms according to (S3.7) and the previous limits, we finally get $v^2 = c^T Dc$, with D as in the theorem.

Finally, we verify the Lindeberg condition. An important closure property, shown in Capitaine et al. (2009, Appendix), called [A] below, states that, for random variables X_1, X_2 and positive ϵ,

$$\mathbb{E}[|X_1 + X_2|^2 \mathbf{1}_{|X_1+X_2|\geq\epsilon}] \leq 4\left(\mathbb{E}[|X_1|^2\mathbf{1}_{|X_1|\geq\epsilon/2}] + \mathbb{E}[|X_2|^2\mathbf{1}_{|X_2|\geq\epsilon/2}]\right).$$

It suffices to establish the Lindeberg condition for the martingale difference sequences

$$Z^{(1)}_{li} = \frac{1}{\sqrt{n}}(x_{li}y_{li} - \rho_l)b_{ii}, \quad Z^{(2)}_{li} = \frac{\rho_l}{2\sqrt{n}}(n^{-1}\operatorname{tr} B_n)\,(x_{li}^2 - 1), \quad Z^{(3)}_{li} = \frac{\rho_l}{2\sqrt{n}}(n^{-1}\operatorname{tr} B_n)\,(y_{li}^2 - 1),$$
$$Z^{(4)}_{li} = \frac{1}{\sqrt{n}}x_{li}S_i(y_{l\cdot}), \quad Z^{(5)}_{li} = \frac{1}{\sqrt{n}}y_{li}S_i(x_{l\cdot}).$$

This follows just as in [A]; recalling that $\|B\| \leq \beta$ we have, for $\epsilon > 0$,

$$\sum_{i=1}^{n}\mathbb{E}[|Z^{(1)}_{li}|^2\mathbf{1}_{|Z^{(1)}_{li}|\geq\epsilon}] \leq \beta^2\mathbb{E}[(x_{li}y_{li} - \rho_l)^2\mathbf{1}_{|x_{li}y_{li}-\rho_l|\geq\sqrt{n}\epsilon/\beta}] \to 0$$

as $n \to \infty$, by the dominated convergence theorem. The sequences $Z^{(2)}_{li}$ and $Z^{(3)}_{li}$ are handled analogously, with $(x_{li}^2 - 1)$ and $(y_{li}^2 - 1)$ in place of $(x_{li}y_{li} - \rho_l)$. For $Z^{(4)}_{li}$, it can be easily shown, as in [A], that $\mathbb{E}[|S_i(y_l)|^4] = O(1)$, so that $\mathbb{E}[|Z^{(4)}_{li}|^4] = O(n^{-2})$ and

$$\sum_{i=1}^{n}\mathbb{E}[|Z^{(4)}_{li}|^2\mathbf{1}_{|Z^{(4)}_{li}|\geq\epsilon}] \leq (1/\epsilon^2)\sum_{i=1}^{n}\mathbb{E}|Z^{(4)}_{li}|^4 \to 0$$

as $n \to \infty$. The same reasoning applies to the last sequence $Z^{(5)}_{li}$, and the proof is complete.

References

Bai, Z. D. and J. Silverstein (2009). *Spectral Analysis of Large Dimensional Random Matrices* (2nd ed.). New York: Springer.

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

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York-London-Sydney.

Capitaine, M., C. Donati-Martin, and D. Féral (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Annals of Probability 37*(1), 1–47.

Gao, J., X. Han, G. Pan, and Y. Yang (2017). High dimensional correlation matrices: The central limit theorem and its applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(3), 677–693.