Biometrika (2018), **xx**, x, *pp*. 1–5
Printed in Great Britain

Supplementary material for 'Signal-plus-noise matrix models: eigenvector deviations and fluctuations'

BY J. CAPE, M. TANG AND C. E. PRIEBE
Department of Applied Mathematics and Statistics, Johns Hopkins University
3400 North Charles Street; Baltimore, Maryland 21218 U.S.A.
joshua.cape@jhu.edu minh@jhu.edu cep@jhu.edu

SUMMARY

This supplementary file contains a joint proof of the theoretical results in the main paper as well as additional simulation examples.

PROOFS

Proof of Theorems 2, 3 *and* 4. We begin with several important observations, namely that

$$\|(I - UU^{\mathrm{T}})\hat{U}\| = \|\sin\Theta(\hat{U}, U)\| = O(\|E\||\Lambda_{rr}|^{-1}) = O_{\mathbb{P}}\{(n\rho_n)^{-1/2}\} \tag{S1}$$

and that there exists $W \in \mathbb{O}_r$ depending on $\hat{U}$ and U such that

$$\|U^{\mathrm{T}}\hat{U} - W\| \leqslant \|\sin\Theta(\hat{U}, U)\|^2 = O_{\mathbb{P}}\{(n\rho_n)^{-1}\}. \tag{S2}$$

In particular, W can be taken to be the product of the left and right orthogonal factors in the singular value decomposition of $U^{\mathrm{T}}\hat{U}$. Additional details can be found, for example, in Cape et al. (2018).

Importantly, the relation $\hat{U}\hat{\Lambda} = \hat{M}\hat{U} = (M + E)\hat{U}$ yields the matrix equation $\hat{U}\hat{\Lambda} - E\hat{U} = M\hat{U}$. The spectra of $\hat{\Lambda}$ and E are disjoint from one another with high probability as a consequence of Assumptions 2 and 3, so it follows that $\hat{U}$ can be written as the matrix series (Bhatia, 1997, § 7.2)

$$\hat{U} = \sum_{k=0}^{\infty} E^k M \hat{U} \hat{\Lambda}^{-(k+1)} = \sum_{k=0}^{\infty} E^k U \Lambda U^{\mathrm{T}} \hat{U} \hat{\Lambda}^{-(k+1)}, \tag{S3}$$

where the second equality holds since $\mathrm{rank}(M) = r$.

For any choice of $W \in \mathbb{O}_r$, the matrix $\hat{U} - UW$ can be decomposed as

$$\begin{aligned}\hat{U} - UW &= E\hat{U}\hat{\Lambda}^{-1} + U\Lambda(U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-1} - \Lambda^{-1}U^{\mathrm{T}}\hat{U}) + U(U^{\mathrm{T}}\hat{U} - W)\\ &= E\hat{U}\hat{\Lambda}^{-1} + R^{(1)} + R^{(2)}_W.\end{aligned}$$

For $R^{(2)}_W = U(U^{\mathrm{T}}\hat{U} - W)$, it follows that for W satisfying (S2),

$$\|R^{(2)}_W\|_{2\to\infty} \leqslant \|U^{\mathrm{T}}\hat{U} - W\|\|U\|_{2\to\infty} = O_{\mathbb{P}}\{(n\rho_n)^{-1}\|U\|_{2\to\infty}\}.$$

For $R^{(1)} = U\Lambda R^{(3)}$ where $R^{(3)} = (U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-1} - \Lambda^{-1}U^{\mathrm{T}}\hat{U}) \in \mathbb{R}^{r\times r}$, the entries of $R^{(3)}$ satisfy

$$R^{(3)}_{ij} = \langle u_i, \hat{u}_j\rangle\big\{(\hat{\Lambda}_{jj})^{-1} - (\Lambda_{ii})^{-1}\big\} = \langle u_i, \hat{u}_j\rangle(\Lambda_{ii} - \hat{\Lambda}_{jj})(\Lambda_{ii})^{-1}(\hat{\Lambda}_{jj})^{-1}.$$

Define the matrix $H_1 \in \mathbb{R}^{r\times r}$ entrywise according to $(H_1)_{ij} = (\Lambda_{ii})^{-1}(\hat{\Lambda}_{jj})^{-1}$. Then, with $\circ$ denoting the Hadamard matrix product,

$$R^{(3)} = -H_1 \circ (U^{\mathrm{T}}\hat{U}\hat{\Lambda} - \Lambda U^{\mathrm{T}}\hat{U}).$$

The rightmost matrix factor can be expanded as

$$(U^{\mathrm{T}}\hat{U}\hat{\Lambda} - \Lambda U^{\mathrm{T}}\hat{U}) = U^{\mathrm{T}}E\hat{U} = U^{\mathrm{T}}EUU^{\mathrm{T}}\hat{U} + U^{\mathrm{T}}E(I - UU^{\mathrm{T}})\hat{U}$$

and is therefore bounded in spectral norm using (S1) in the manner

$$\|U^{\mathrm{T}}\hat{U}\hat{\Lambda} - \Lambda U^{\mathrm{T}}\hat{U}\| \leqslant \|U^{\mathrm{T}}EU\| + O_{\mathbb{P}}(1).$$

Combining the above observations with properties of matrix norms yields the following two-to-infinity norm bound on $R^{(1)}$:

$$\begin{aligned}\|R^{(1)}\|_{2\to\infty} = \|U\Lambda R^{(3)}\|_{2\to\infty} &\leqslant r\|U\|_{2\to\infty}\|\Lambda\|\|H_1\|_{\max}\|U^{\mathrm{T}}\hat{U}\hat{\Lambda} - \Lambda U^{\mathrm{T}}\hat{U}\| \\ &= O_{\mathbb{P}}\big\{r(n\rho_n)^{-1}(\|U^{\mathrm{T}}EU\| + 1)\|U\|_{2\to\infty}\big\}.\end{aligned}$$

Assumptions 2 and 3 together with an application of Weyl's inequality (Bhatia, 1997, Corollary 3.2.6) guarantee that there exist constants $C_1, C_2 > 0$ such that $\|E\| \leqslant C_1(n\rho_n)^{1/2}$ and $\|\hat{\Lambda}^{-1}\| \leqslant C_2(n\rho_n)^{-1}$ with high probability for n sufficiently large. Therefore, by applying the earlier matrix series expansion,

$$\begin{aligned}\|E\hat{U}\hat{\Lambda}^{-1}\|_{2\to\infty} &= \left\|\sum_{k=1}^{\infty} E^k U\Lambda U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-(k+1)}\right\|_{2\to\infty} \\ &\leqslant \sum_{k=1}^{k(n)} \|E^k U\|_{2\to\infty}\|\Lambda\|\|\hat{\Lambda}^{-1}\|^{k+1} + \sum_{k=k(n)+1}^{\infty} \|E\|^k\|\Lambda\|\|\hat{\Lambda}^{-1}\|^{k+1} \\ &= O_{\mathbb{P}}\big\{r^{1/2}(n\rho_n)^{-1/2}(\log n)^{\xi}\|U\|_{2\to\infty} + (n\rho_n)^{-1/2}\|U\|_{2\to\infty}\big\},\end{aligned}$$

where we have used the facts that $n\rho_n = \omega\{(\log n)^{2\xi}\}$, that $(n\rho_n)^{-k(n)/2} \leqslant n^{-1/2} \leqslant \|U\|_{2\to\infty}$ for n sufficiently large, and that by Assumption 4, for each $k \leqslant k(n)$, with high probability

$$\|E^k U\|_{2\to\infty} \leqslant r^{1/2} \max_{i\in[n], j\in[r]} |\langle E^k u_j, e_i\rangle| \leqslant r^{1/2}(C_E n\rho_n)^{k/2}(\log n)^{k\xi}\|U\|_{2\to\infty}.$$

Since $\|U^{\mathrm{T}}EU\| \leqslant \|E\|$ and $r^{1/2} \leqslant (\log n)^{\xi}$ with $n\rho_n = \omega\{(\log n)^{2\xi}\}$, we have

$$\begin{aligned}\|\hat{U} - UW\|_{2\to\infty} &\leqslant \|E\hat{U}\hat{\Lambda}^{-1}\|_{2\to\infty} + \|R^{(1)}\|_{2\to\infty} + \|R_W^{(2)}\|_{2\to\infty} \\ &= O_{\mathbb{P}}\big\{r^{1/2}(n\rho_n)^{-1/2}(\log n)^{\xi}\|U\|_{2\to\infty}\big\}.\end{aligned}$$

This completes the proof of Theorem 2.

Next, we further decompose the matrix $E\hat{U}\hat{\Lambda}^{-1}$ by extending the above proof techniques in order to obtain second-order fluctuations. Using the matrix series form in (S3) yields

$$\begin{aligned} E\hat{U}\hat{\Lambda}^{-1} &= EU\Lambda U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-2} + \sum_{k=2}^{\infty} E^{k}U\Lambda U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-(k+1)} \\ &= EU\Lambda^{-1}W + EU\Lambda(U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-2} - \Lambda^{-2}U^{\mathrm{T}}\hat{U}) + EU\Lambda^{-1}(U^{\mathrm{T}}\hat{U} - W) \\ &\quad + \sum_{k=2}^{\infty} E^{k}U\Lambda U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-(k+1)} \\ &= EU\Lambda^{-1}W + R_{2}^{(1)} + R_{2,W}^{(2)} + R_{2}^{(\infty)}. \end{aligned}$$

The final term satisfies the bound

$$\|R_{2}^{(\infty)}\|_{2\to\infty} = O_{\mathbb{P}}\big\{r^{1/2}(n\rho_n)^{-1}(\log n)^{2\xi}\|U\|_{2\to\infty}\big\},$$

which follows from Assumption 4 holding up to $k(n)+1$, i.e.,

$$\begin{aligned} \|R_{2}^{(\infty)}\|_{2\to\infty} &\leqslant \sum_{k=2}^{k(n)+1} \|E^{k}U\|_{2\to\infty}\|\Lambda\|\|\hat{\Lambda}^{-1}\|^{k+1} + \sum_{k=k(n)+2}^{\infty} \|E\|^{k}\|\Lambda\|\|\hat{\Lambda}^{-1}\|^{k+1} \\ &= O_{\mathbb{P}}\big\{r^{1/2}(n\rho_n)^{-1}(\log n)^{2\xi}\|U\|_{2\to\infty} + (n\rho_n)^{-1}\|U\|_{2\to\infty}\big\}. \end{aligned}$$

On the other hand, modifying the above argument used to bound $R_{W}^{(2)}$ gives

$$\|R_{2,W}^{(2)}\|_{2\to\infty} \leqslant \|EU\|_{2\to\infty}\|\Lambda^{-1}\|\|U^{\mathrm{T}}\hat{U} - W\| = O_{\mathbb{P}}\big\{r^{1/2}(n\rho_n)^{-3/2}(\log n)^{\xi}\|U\|_{2\to\infty}\big\}.$$

We now bound $R_{2}^{(1)} = EU\Lambda(U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-2} - \Lambda^{-2}U^{\mathrm{T}}\hat{U})$ by extending the previous argument used to bound $R^{(1)}$. For $R_{2}^{(1)} = EU\Lambda R_{2}^{(3)}$ where $R_{2}^{(3)} = (U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{-2} - \Lambda^{-2}U^{\mathrm{T}}\hat{U}) \in \mathbb{R}^{r\times r}$, the entries of $R_{2}^{(3)}$ satisfy

$$R_{ij}^{(3)} = \langle u_i, \hat{u}_j\rangle\big\{(\hat{\Lambda}_{jj})^{-2} - (\Lambda_{ii})^{-2}\big\} = \langle u_i, \hat{u}_j\rangle(\Lambda_{ii}^{2} - \hat{\Lambda}_{jj}^{2})(\Lambda_{ii})^{-2}(\hat{\Lambda}_{jj})^{-2}.$$

Define the matrix $H_2 \in \mathbb{R}^{r\times r}$ entrywise according to $(H_2)_{ij} = (\Lambda_{ii})^{-2}(\hat{\Lambda}_{jj})^{-2}$. Then, with $\circ$ denoting the Hadamard matrix product,

$$R_{2}^{(3)} = -H_2 \circ (U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{2} - \Lambda^{2}U^{\mathrm{T}}\hat{U}).$$

The rightmost matrix factor can be written as

$$(U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{2} - \Lambda^{2}U^{\mathrm{T}}\hat{U}) = U^{\mathrm{T}}(\hat{M})^{2}\hat{U} - U^{\mathrm{T}}M^{2}\hat{U} = U^{\mathrm{T}}(ME + EM)\hat{U}$$

and has spectral norm of the order of $O_{\mathbb{P}}\{(n\rho_n)^{3/2}\}$. Hence,

$$\begin{aligned} \|R_{2}^{(1)}\|_{2\to\infty} = \|EU\Lambda R_{2}^{(3)}\|_{2\to\infty} &\leqslant r\|EU\|_{2\to\infty}\|\Lambda\|\|H_2\|_{\max}\|U^{\mathrm{T}}\hat{U}\hat{\Lambda}^{2} - \Lambda^{2}U^{\mathrm{T}}\hat{U}\| \\ &= O_{\mathbb{P}}\big\{r^{3/2}(n\rho_n)^{-1}(\log n)^{\xi}\|U\|_{2\to\infty}\big\}. \end{aligned}$$

For $R = R^{(1)} + R_{W}^{(2)} + R_{2}^{(1)} + R_{2,W}^{(2)} + R_{2}^{(\infty)}$, we have therefore shown that

$$\hat{U} - UW = EU\Lambda^{-1}W + R \tag{S4}$$

where, since $r^{1/2} \leqslant (\log n)^{\xi}$, the residual matrix R satisfies

$$\|R\|_{2\to\infty} = O_{\mathbb{P}}\Big[(n\rho_n)^{-1} \times r \times \max\big\{(\log n)^{2\xi}, \|U^{\mathrm{T}}EU\| + 1\big\} \times \|U\|_{2\to\infty}\Big].$$

The leading term agrees with the order of the bound in Theorem 2, namely

$$\|EU\Lambda^{-1}W\|_{2\to\infty} = O_{\mathbb{P}}\big\{(n\rho_n)^{-1/2} \times r^{1/2}(\log n)^{\xi}\|U\|_{2\to\infty}\big\}.$$

This establishes Theorem 3 en route to proving Theorem 4, which we now proceed to finish.

Since $M = \rho_n XX^{\mathrm{T}} \equiv U\Lambda U^{\mathrm{T}}$, there exists an orthogonal matrix W_X (depending on n) such that $\rho_n^{1/2}X = U\Lambda^{1/2}W_X$; hence $\rho_n X^{\mathrm{T}}X = W_X^{\mathrm{T}}\Lambda W_X$. Following some algebraic manipulations, the matrix $EU\Lambda^{-1}W$ can therefore be written as

$$EU\Lambda^{-1}W = \rho_n^{-1}EX(X^{\mathrm{T}}X)^{-3/2}(W_X^{\mathrm{T}}W).$$

Upon plugging this observation into (S4), and subsequent matrix multiplication, we obtain the relation

$$\hat{U}W^{\mathrm{T}}W_X - UW_X = \rho_n^{-1}EX(X^{\mathrm{T}}X)^{-3/2} + RW^{\mathrm{T}}W_X.$$

For fixed i, let $\hat{U}_i$, U_i and R_i be column vectors corresponding to the ith rows of $\hat{U}$, U and R, respectively. Equation (4) in the main paper implies that $n\rho_n^{1/2}\|R_i\| \to 0$ in probability. In addition, $(n^{-1}X^{\mathrm{T}}X)^{-3/2} \to \Xi^{-3/2}$ by Assumption 5 together with the continuous mapping theorem. The scaled ith row of EX converges in distribution to $Y_i \sim N_r(0, \Gamma_i)$ by Assumption 5, so combining the above observations with Slutsky's theorem yields that there exist sequences of orthogonal matrices (W) and (W_X) such that

$$\begin{aligned} n\rho_n^{1/2}W_X^{\mathrm{T}}\big(W\hat{U}_i - U_i\big) &= \big(n^{-1}X^{\mathrm{T}}X\big)^{-3/2}\big\{(n\rho_n)^{-1/2}(EX)_i\big\} + n\rho_n^{1/2}W_X^{\mathrm{T}}WR_i \\ &\Rightarrow \Xi^{-3/2}Y_i + 0. \end{aligned}$$

In particular, we have the row-wise convergence in distribution

$$n\rho_n^{1/2}W_X^{\mathrm{T}}\big(W\hat{U}_i - U_i\big) \Rightarrow N_r(0, \Sigma_i),$$

where $\Sigma_i = \Xi^{-3/2}\Gamma_i\Xi^{-3/2}$. This completes the proof of Theorem 4. □

ADDITIONAL SIMULATION EXAMPLES

Two-block stochastic block model (continued)

Consider n-vertex graphs arising from the two-block stochastic block model with 40% of the vertices belonging to the first block and where the block edge probability matrix B has entries $B_{11} = 0.5$, $B_{12} = B_{21} = 0.3$ and $B_{22} = 0.3$. This model corresponds to Fig. 1(b) in the main paper. Here, Table S1 shows block-conditional sample covariance matrix estimates for the centred random vectors $n\rho_n^{1/2}W_X^{\mathrm{T}}(W\hat{U}_i - U_i)$. Also shown are the corresponding theoretical covariance matrices.

Spike matrix models

Figure S1 presents two additional examples illustrating Theorem 4 in the main paper for one- and two-dimensional spike matrix models, written in the rescaled form $\hat{M} = \lambda UU^{\mathrm{T}} + E$ with $\rho_n \equiv 1$. In the left plot, $\lambda = n$, $U = n^{-1/2}e \in \mathbb{R}^n$ and $E_{ij} \sim \text{Laplace}(0, 2^{-1/2})$ independently for $i \leqslant j$ with $E_{ij} = E_{ji}$. Here Ξ is the one-dimensional identity matrix, i.e., $\Xi = I_1$, and

Table S1. *Empirical and theoretical covariance matrices for the two-block model*

n	1000	2000	∞
$\hat{\Sigma}_1$	$\begin{bmatrix} 14.11 & -36.08 \\ -36.08 & 110.13 \end{bmatrix}$	$\begin{bmatrix} 14.94 & -36.85 \\ -36.85 & 108.55 \end{bmatrix}$	$\begin{bmatrix} 15.14 & -38.05 \\ -38.05 & 112.34 \end{bmatrix}$
$\hat{\Sigma}_2$	$\begin{bmatrix} 11.76 & -30.09 \\ -30.09 & 93.07 \end{bmatrix}$	$\begin{bmatrix} 12.91 & -33.04 \\ -33.04 & 101.64 \end{bmatrix}$	$\begin{bmatrix} 13.12 & -33.93 \\ -33.93 & 103.94 \end{bmatrix}$

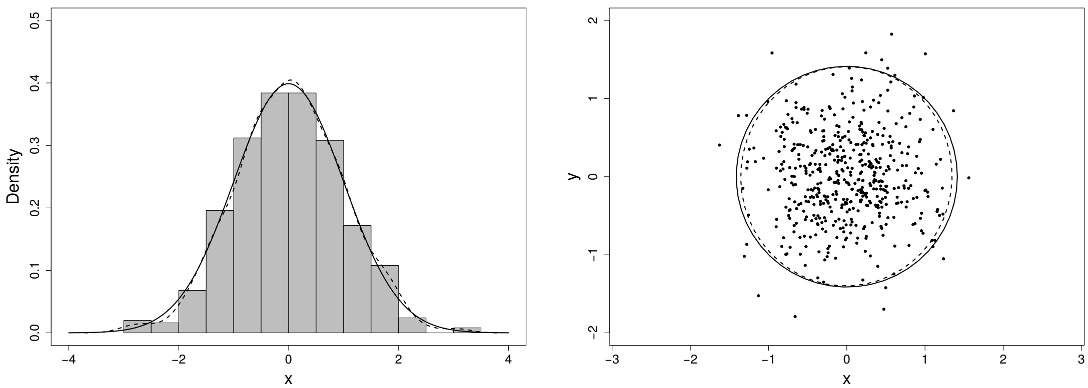

Fig. S1. The left panel shows a one-dimensional simulation for $n = 500$ with the empirical (dashed) and theoretical (solid) eigenvector fluctuation densities. The right panel shows a two-dimensional simulation for $n = 500$, where the dashed ellipse gives the 95% level curve for the empirical distribution and the solid ellipse gives the 95% level curve for the row-wise theoretical distribution.

$(n\rho_n)^{-1/2}(EX)_i \Rightarrow N_1(0, 1)$ by the central limit theorem, so that for each fixed row i Theorem 4 yields convergence in distribution to $N_1(0, 1)$. In the right plot, $\lambda = n$, and $U_{ij} = n^{-1/2}$ for $1 \leqslant i \leqslant n$ and $j = 1$ and for $1 \leqslant i \leqslant n/2$ and $j = 2$, while $U_{ij} = -n^{-1/2}$ otherwise. In addition, $E_{ij} \sim \text{Un}[-1, 1]$ independently for $i \leqslant j$ with $E_{ij} = E_{ji}$, so $\text{var}(E_{ij}) = 1/3$. Here $(n\rho_n)^{-1/2}(EX)_i$ converges in distribution to a centred multivariate normal random variable with covariance matrix $\Gamma_i = (1/3)I_2 \in \mathbb{R}^{2\times 2}$ by the multivariate central limit theorem, while the second moment matrix for the rows X_i in Assumption 5 is simply $\Xi = I_2$. Theorem 4 therefore yields $n\rho_n^{1/2} W_X^{\mathrm{T}}(W\hat{U}_i - U_i) \Rightarrow N_2(\mu, \Sigma_i)$, where $\mu = (0, 0)^{\mathrm{T}} \in \mathbb{R}^2$ and $\Sigma_i = (1/3)I_2 \in \mathbb{R}^{2\times 2}$. The plots depict all vectors computed from a single simulated adjacency matrix.

References

BHATIA, R. (1997). *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*. New York: Springer.

CAPE, J., TANG, M. & PRIEBE, C. E. (2018). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* to appear; *arXiv:* 1705.10735.