For $i = 1, \ldots, \widehat{k} = k$, let $\widehat{v}_i$ be an arbitrary unit eigenvector of $\widehat{X}_n$. By the eigenvalue master equation, $\widehat{X}_n \widehat{v}_i = \widehat{\lambda}_i \widehat{v}_i$, it follows that

$$U_{n,k}^H \left(\widehat{\lambda}_i I_n - X_n\right)^{-1} X_n U_{n,k} \Sigma U_{n,k}^H \widehat{v}_i = U_{n,k}^H \widehat{v}_i. \quad (27)$$

Let $X_n = V_n \Lambda_n V_n^H$ be the eigenvalue decomposition of X_n such that $\Lambda_n = \mathbf{diag}(\lambda_1(X_n), \ldots, \lambda_n(X_n))$ and $\lambda_1(X_n) \geq \ldots \geq \lambda_n(X_n)$. Using this decomposition and defining $W_{n,k} = V^H U_{n,k}$, (27) simplifies to

$$W_{n,k}^H \left(\widehat{\lambda}_i I_n - \Lambda_n\right)^{-1} \Lambda_n W_{n,k} \Sigma U_{n,k}^H \widehat{v}_i = U_{n,k}^H \widehat{v}_i. \quad (28)$$

Define the columns of $W_{n,k}$ to be $w_j^{(n)} = [w_{1,j}^{(n)}, \ldots, w_{n,j}^{(n)}]^T$ for $j = 1, \ldots, k$. These columns are orthonormal and isotropically random. We can rewrite (28) as

$$\left[T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^{k} \Sigma U_{n,k}^H \widehat{v}_i = U_{n,k}^H \widehat{v}_i \quad (29)$$

where for $r = 1, \ldots, k$, $j = 1, \ldots, k$, $\mu_{r,j}^{(n)} = \sum_{\ell=1}^{n} \overline{w_{\ell,r}^{(n)}} w_{\ell,j}^{(n)} \delta_{\lambda_\ell(X_n)}$ is a complex measure and $T_{\mu_{r,j}^{(n)}}$ is the T-transform defined by $T_\mu(z) = \int \frac{t}{z-t} d\mu(t)$ for $z \notin$ supp μ. We may rewrite (29) as

$$\left(I_k - \left[\sigma_j^2 T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^{k}\right) U_{n,k}^H \widehat{v}_i = 0.$$

Therefore, $U_{n,k}^H \widehat{v}_i$ must be in the kernel of $M_n\left(\widehat{\lambda}_i\right) = I_k - \left[\sigma_j^2 T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^{k}$. By Proposition 9.3 of [20]

$$\mu_{r,j}^{(n)} \xrightarrow{\text{a.s.}} \begin{cases} \mu_X & \text{for } i = j \\ \delta_0 & \text{o.w.} \end{cases}$$

where μ_X is the limiting eigenvalue distribution of X_n. Therefore,

$$M_n\left(\widehat{\lambda}_i\right) \xrightarrow{\text{a.s.}} \mathbf{diag}\left(1 - \sigma_1^2 T_{\mu_X}\left(\widehat{\lambda}_i\right), \ldots, 1 - \sigma_k^2 T_{\mu_X}\left(\widehat{\lambda}_i\right)\right).$$

As $k_{\text{eff}} = k$, for $i = 1, \ldots, k$, $\sigma_i^2 > 1/T_{\mu_X}(b^+)$, where b is the supremum of the support of μ_X. As $\widehat{\lambda}_i$ is the eigenvalue corresponding to the eigenvector $\widehat{v}_i$, by Theorem 2.6 of [20]$\widehat{\lambda}_i \xrightarrow{\text{a.s.}} T_{\mu_X}^{-1}\left(1/\sigma_i^2\right)$. Therefore,

$$M_n\left(\widehat{\lambda}_i\right) \xrightarrow{\text{a.s.}} \mathbf{diag}\left(1 - \frac{\sigma_1^2}{\sigma_i^2}, \ldots, 1 - \frac{\sigma_{i-1}^2}{\sigma_i^2}, 0, 1 - \frac{\sigma_{i+1}^2}{\sigma_i^2}, \ldots, 1 - \frac{\sigma_k^2}{\sigma_i^2}\right). \quad (30)$$

Recall that $U_{n,k}^H \widehat{v}_i$ must be in the kernel of $M_n\left(\widehat{\lambda}_i\right)$. Therefore, any limit point of $U_{n,k}^H \widehat{v}_i$ is in the kernel of the matrix on the right hand side of (30). Therefore, for $i \neq j$, $i = 1, \ldots, \widehat{k}$, $j = 1, \ldots, k$, we must have that $\left(1 - \frac{\sigma_j^2}{\sigma_i^2}\right) \langle u_j, \widehat{v}_i \rangle = 0$. As $\sigma_i^2 \neq \sigma_j^2$, for this condition to be satisfied we must have that for $j \neq i$, $i = 1, \ldots, \widehat{k}$, $j = 1, \ldots, k$, $\langle u_j, \widehat{v}_i \rangle \xrightarrow{\text{a.s.}} 0$.

Recall that our observed vectors $y_i \in \mathbb{C}^{n \times 1}$ have covariance matrix $U_{n,k} \Sigma U_{n,k}^H + I_n = P_n + I_n$. Therefore, our observation matrix, Y_n which is a $n \times m$ matrix, may be written $Y_n = (P_n + I_n)^{1/2} Z_n$. The sample covariance matrix, $S_n = \frac{1}{m} Y_n Y_n^H$, may be written $S_n = (I_n + P_n)^{1/2} X_n (I_n + P_n)^{1/2}$. By similarity transform, if $\widehat{v}_i$ is a unit-norm eigenvector of $\widehat{X}_n$ then $\widehat{s}_i = (I_n + P_n)^{1/2} \widehat{v}_i$ is an eigenvector of S_n. If $\widehat{u}_i = \widehat{s}_i / \|\widehat{s}_i\|$ is a unit-norm eigenvector of S_n, it follows that

$$\langle u_j, \widehat{u}_i \rangle = \frac{\sqrt{\sigma_i^2 + 1} \langle u_j, \widehat{v}_i \rangle}{\sqrt{\sigma_i^2 |\langle u_j, \widehat{v}_i \rangle|^2 + 1}}.$$

As $\langle u_j, \widehat{v}_i \rangle \xrightarrow{\text{a.s.}} 0$ for all $i \neq j$, $i = 1, \ldots, \widehat{k}$, $j = 1, \ldots, k$, it follows that $\langle u_j, \widehat{u}_i \rangle \xrightarrow{\text{a.s.}} 0$ for all $i \neq j$ $i = 1, \ldots, \widehat{k}$, $j = 1, \ldots, k$.

Claim 5.1: We conjecture that this result holds for the general case of $i \neq j$, $i = 1, \ldots, \widehat{k}$, $j = 1, \ldots, k$, not just when $\widehat{k} = k_{\text{eff}} = k$. Consider the case when $k = 1$. For $i > 2$, if $\widehat{\lambda}_i$ is an eigenvalue of $\widehat{X}_n = X_n(I_n + \sigma^2 u u^H)$, then it satisfies $\det(\widehat{\lambda}_i I_n - X_n(I_n + \sigma^2 u u^H)) = \det(\widehat{\lambda}_i I_n - X_n) \det(I_n - (\widehat{\lambda}_i I_n - X_n)^{-1} X_n \sigma^2 u u^H) = 0$. Therefore, if $\widehat{\lambda}_i$ is not an eigenvalue of X_n, the corresponding unit norm eigenvector $\widehat{v}_i$ is in the kernel of $I_n - (\widehat{\lambda}_i I_n - X_n)^{-1} X_n \sigma^2 u u^H$. Therefore

$$|\langle \widehat{v}_i, u \rangle|^2 = \frac{1}{\sigma^4 u^H X_n \left(\widehat{\lambda}_i I_n - X_n\right)^{-2} X_n u}.$$

Recall that Weyl's interlacing lemma for eigenvalues gives $\lambda_i(X_n) \leq \widehat{\lambda}_i \leq \lambda_{i-1}(X_n)$. Letting $X_n = V_n \Lambda_n V_n^H$ and $w = V_n^H u$, we see the importance of the asymptotic spacing of eigenvalues of X_n in

$$u^H X_n (\widehat{\lambda}_i I_n - X_n)^{-2} X_n u = \sum_{\ell=1}^{n} \frac{|w_\ell|^2 \lambda_\ell^2(X_n)}{\left(\widehat{\lambda}_i - \lambda_\ell\right)^2}$$

$|w_i|^2$

$$\geq \frac{\min_j \lambda_j^2(X_n) \cancel{\min_j} |w_j|^2}{\max_j |\lambda_{j-1} - \lambda_j|^2}. \quad |w_i|^2$$

In [32] it is shown that $\min_j \lambda_j^2(X_n) = \lambda_n^2(X_n) \xrightarrow{\text{a.s.}} (1 - \sqrt{c})^2$. The typical spacing between eigenvalues is $O(1/n)$ while the typical magnitude of w_j^2 is $O(1/n)$[33]. Therefore, the right hand side of the above inequality will typically be $O(n)$ and we get the desired result of $|\langle \widehat{v}_i, u \rangle|^2 \xrightarrow{\text{a.s.}} 0$. More generally, it is the behavior of the largest eigenvalue gap ~~and the smallest element of w_i~~ that drives this convergence. Thus, so long as the eigenvector whose elements are w_i are delocalized (i.e., having elements of $O(1/\sqrt{n})$) and the largest gap between k successive eigenvalues is at most $O(1/(n^{(0.5+\epsilon)})$, the right hand side of the inequality will be unbounded with n. The claim follows after applying a similarity transform as in the proof of Theorem 5.1. ■

REFERENCES

[1] L. L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991, vol. 1.

[2] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer, 2001, vol. 1.