Outliers in signal-plus-noise models

Consider $X = \lambda v v^T + W$, where

- $W \sim GOE(n)$
- $v \in \mathbb{R}^n$, $\|v\|_2 = 1$.
- $\lambda > 0$.

What is the behavior of eigenvalues/vectors of X?

More generally, consider $X = \sum_{i=1}^k \lambda_i v_i v_i^T + W$, where

- $v_1, \dots, v_k \in \mathbb{R}^n$ are orthonormal
- k is a small constant.
- For simplicity, $\lambda_1, \dots, \lambda_k > 0$.

Prop (Weyl's inequalities): Let $\hat\lambda_1(W) \geq \dots \geq \hat\lambda_n(W)$ be eigenvalues of W, and and $\hat\lambda_1(X) \geq \dots \geq \hat\lambda_n(X)$ be those of $X = \sum_{i=1}^k \lambda_i v_i v_i^T + W$ where $\lambda_i \geq 0$ for each i. Then for each $j \in \{k+1, \dots, n\}$,

$$\hat\lambda_j(W) \leq \hat\lambda_j(X) \leq \hat\lambda_{j-k}(W).$$

Proof: For the lower bound, by Courant-Fischer

$$\hat\lambda_j(X) = \max\left\{\min_{v \in V, \|v\|=1} v^T X v : V \subseteq \mathbb{R}^n, \dim(V) = j\right\}$$

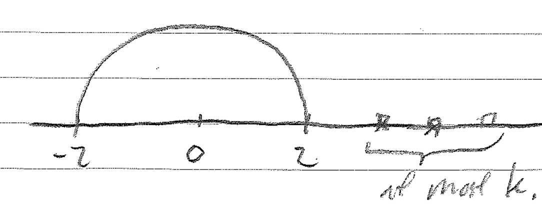
$$\geq \max\left\{\min_{v \in V, \|v\|=1} v^T W v : V \subseteq \mathbb{R}^n, \dim(V) = j\right\} = \hat\lambda_j(W),$$

since $v^T\left(\sum_{i=1}^k \lambda_i v_i v_i^T\right)v = \sum_{i=1}^k \lambda_i (v_i^T v)^2 \geq 0$. (This is true for any $X = W + P$ where P is positive semi-definite.)

For the upper bound, it suffices to show for any $V \subseteq \mathbb{R}^n$ with $\dim(V) = j$, there exists $v \in V$ where $v^T X v \leq \hat\lambda_{j-k}(W)$. Let U be the span of $v_1, \dots, v_k$ and the top $j-k-1$ eigenvectors of W: This has dimension at most $j-1$. Then the orthogonal complement $U^\perp$ has dimension at least $n-j+1$, so $U^\perp$ intersects V. But for any unit vector $v \in U^\perp \cap V$, $v^T X v = v^T W v \leq \hat\lambda_{j-k}(W)$, because v is orthogonal to $v_1, \dots, v_k$ and the top $j-k-1$ eigenvectors of W.

Implication: The empirical distributions of $\{\hat\lambda_1(W), \dots, \hat\lambda_{n-k}(W)\}$ and of $\{\hat\lambda_{k+1}(W), \dots, \hat\lambda_n(W)\}$ both converge to the semicircle law. Then so does that of $\{\hat\lambda_{k+1}(X), \dots, \hat\lambda_n(X)\}$. Furthermore, since $\|W\| \to 2$, we have $\hat\lambda_{k+1}(X) \to 2$ and $\hat\lambda_n(X) \to -2$.

So all but possibly the top k eigenvalues of X belong to the semicircle "bulk," and there are at most k "outliers."



If $\lambda_1, \dots, \lambda_k$ are fixed, what is the behavior of the outliers as $n \to \infty$?

Heuristic calculation: Consider $X = \lambda v v^T + W$.

Suppose $\hat\lambda > 2 + \varepsilon$ is an outlier eigenvalue of X, with unit eigenvector $\hat v$.

Then $\hat\lambda \hat v = X \hat v$

$= \lambda v (v^T \hat v) + W \hat v$

$\Rightarrow -\lambda (v^T \hat v) \cdot v = (W - \hat\lambda I)\hat v.$

Since $\|W\| \approx 2$, $W - \hat{\lambda} I$ should be invertible with high probability.

Let $G(\hat{\lambda}) = (W - \hat{\lambda} I)^{-1}$ be the resolvent. Then

(*) $\quad \hat{v} = -\lambda (v^T \hat{v}) \cdot G(\hat{\lambda}) \cdot v.$

Multiply on the left by v^T, and cancel $v^T \hat{v}$ on both sides:

$$1 = -\lambda\, v^T G(\hat{\lambda}) v.$$

Let O be orthogonal s.t. $Ov = e_1$. Then for any fixed z,

$$v^T G(z) v = e_1^T O (W - zI)^{-1} O^T e_1$$

$$= e_1^T (OWO^T - zI)^{-1} e_1 \overset{d}{=} e_1^T (W - zI) e_1 = G(z)_{11}.$$

So we expect $v^T G(z) v \approx \mathbb{E}[v^T G(z) v]$ $\mathbb{E}[G(z)_{11}] = \mathbb{E}[m_n(z)]$,

where $m_n(z) = \frac{1}{n} \mathrm{Tr}\, G(z)$ is the empirical Stieltjes transform of W.

Letting $m(z)$ be the Stieltjes transform of the semi-circle law,

$$v^T G(\hat{\lambda}) v \approx m(\hat{\lambda}).$$

So $m(\hat{\lambda}) \approx -\frac{1}{\lambda}$. Over $z \in (2, \infty)$, $m(z)$ increases from -1 to 0.

So this is only possible when $\lambda > 1$. If $\lambda > 1$, applying

$$m(\hat{\lambda})^2 + \hat{\lambda} m(\hat{\lambda}) + 1 = 0$$

and substituting $m(\hat{\lambda}) \approx -\frac{1}{\lambda}$, we get for the observed eigenvalue $\hat{\lambda}$

$$\frac{1}{\lambda^2} - \frac{\hat{\lambda}}{\lambda} + 1 \approx 0 \Rightarrow \hat{\lambda} \approx \lambda + \frac{1}{\lambda}.$$

Returning to (*), now take the squared-norm on both sides. Then

$$1 = \|\hat{v}\|^2 = \lambda^2 (v^T \hat{v})^2 \cdot \|G(\hat{\lambda}) v\|^2 = \lambda^2 (v^T \hat{v})^2 \cdot v^T G(\hat{\lambda})^2 v.$$

Note that $\frac{d}{d\lambda} G(\lambda) = \frac{d}{d\lambda} (W - \lambda I)^{-1} = (W - \lambda I)^{-1} \cdot I \cdot (W - \lambda I)^{-1} = G(\lambda)^2$.

So we expect $v^T G(\hat{\lambda})^2 v = v^T G'(\hat{\lambda}) v \approx \mathbb{E}[m_n'(\hat{\lambda})] \approx m'(\hat{\lambda})$.

Differentiating $m(z)^2 + z m(z) + 1 = 0$ in z,

$$2 m(\hat{\lambda}) \cdot m'(\hat{\lambda}) + m(\hat{\lambda}) + \hat{\lambda} m'(\hat{\lambda}) = 0$$

$$\Rightarrow m'(\hat{\lambda}) = -\frac{m(\hat{\lambda})}{2m(\hat{\lambda}) + \hat{\lambda}} \approx \frac{1/\lambda}{-2/\lambda + \lambda + 1/\lambda} = \frac{1}{\lambda^2 - 1}.$$

Then $(v^T \hat{v})^2 = \frac{1}{\lambda^2 v^T G(\hat{\lambda})^2 v} \approx \frac{\lambda^2 - 1}{\lambda^2} = 1 - \frac{1}{\lambda^2}$.

Thm: Let $X = \sum_{i=1}^{k} \lambda_i v_i v_i^T + W$ where $W \sim GOE(n)$, $v_1, \dots, v_k \in \mathbb{R}^n$ are orthonormal, and $\lambda_1 > \dots > \lambda_k > 0$. Then a.s. as $n \to \infty$:

- Corresponding to each $\lambda_i > 1$, there is an eigenvalue $\hat{\lambda}_i$ of X which converges to $\lambda_i + \frac{1}{\lambda_i}$. Letting $\hat{v}_i$ be the corresponding unit eigenvector,

$$(v_i^T \hat{v}_i)^2 \to 1 - \frac{1}{\lambda_i^2} \text{ and } (v_j^T \hat{v}_i)^2 \to 0 \text{ for all } j \in \{1, \dots, k\} \setminus \{i\}.$$

- All other eigenvalues of X converge to the interval $[-2, 2]$.

Three qualitative phenoma:

- Phase transition: Only λ_i's larger than 1 yield an outlier. λ_i's less than 1 are "covered by the noise".
- Upward bias: $\hat{\lambda}_i \to \lambda_i + \frac{1}{\lambda_i}$, not λ_i. There is an upward bias $\frac{1}{\lambda_i}$.
- Eigenvector inconsistency: $(v_i^T \hat{v}_i)^2 \to 1 - \frac{1}{\lambda_i^2}$, not 1. $\hat{v}_i$ lies on a cone around the true v_i, where the angle is a decreasing function of λ_i.

Lemma: Let $X \sim GOE(n)$, $G(z) = (X - zI)^{-1}$, $m(z)$ the semicircle Stieltjes transform.

For any constants $\varepsilon, C > 0$ and any deterministic unit vectors $v, w \in \mathbb{R}^n$,

$$|v^T G(z) w - (v^T w) m(z)| \prec \frac{1}{\sqrt{n}}$$

$$|v^T G'(z) w - (v^T w) m'(z)| \prec \frac{1}{\sqrt{n}}$$

uniformly over $\mathbb{D} = \{z \in \mathbb{C} : dist(z, [-2,2]) \geq \varepsilon, |z| \leq C\}$.

(Recall: This means for any $\delta, D > 0$ and all large n

$$\mathbb{P}[\exists z \in \mathbb{D} : |v^T G(z) v - m(z)| > n^{-\frac{1}{2}+\delta}] < n^{-D},$$

and similarly for the second statement.)

Proof: First consider fixed z. Write $X = \frac{W + W^T}{\sqrt{2n}}$ where $W_{ij} \overset{iid}{\sim} N(0,1)$, and view $v^T G(z) w$ as a function of W. Recall

$$\partial_{W_{ij}} G(z) = -\frac{1}{\sqrt{2n}} G(z) (\Delta_{ij} + \Delta_{ji}) G(z),$$

where $\Delta_{ij} = e_i e_j^T$ has the single entry (i,j) equal to 1. Then

$$\partial_{W_{ij}} [v^T G(z) w] = -\frac{1}{\sqrt{2n}} (v^T G(z) e_i \cdot e_j^T G(z) w + v^T G(z) e_j \cdot e_i^T G(z) w).$$

$$\| \nabla_W [v^T G(z) w] \|^2 = \frac{1}{2n} \sum_{i,j} \left(v^T G(z) e_i \cdot e_j^T G(z) w + v^T G(z) e_j \cdot e_i^T G(z) w \right)^2$$

$$= \frac{1}{n} \sum_{i,j} \Big(v^T G(z) e_i \cdot e_i^T G(z) v \cdot w^T G(z) e_j \cdot e_j^T G(z) w$$
$$+ v^T G(z) e_i \cdot e_i^T G(z) w \cdot v^T G(z) e_j \cdot e_j^T G(z) w \Big)$$

$$= \frac{1}{n} (v^T G(z)^2 v) \cdot (w^T G(z)^2 w) + \frac{1}{n} (v^T G(z)^2 w)^2.$$

(We've used $\sum_{i=1}^n e_i e_i^T = I$.)

If $\|X\| \leq 2 + \varepsilon/2$, then $\|G(z)\| = \sup_i \frac{1}{|\lambda_i(X) - z|} \leq \frac{2}{\varepsilon}$ for $z \in \mathbb{D}$.

So we get $\| \nabla_W [v^T G(z) w] \|^2 \leq \frac{C}{n}$ for a constant $C = C(\varepsilon) > 0$.

The set $\{W \in \mathbb{R}^{n \times n} : \|\frac{W + W^T}{\sqrt{2n}}\| \leq 2 + \frac{\varepsilon}{2}\}$ is convex, so $v^T G(z) w$ is $\sqrt{\frac{C}{n}}$-Lipschitz on this set.

Let $F(W)$ be the Lipschitz extension of $W \mapsto v^T G(z) w$. Then

$$\mathbb{P}[|F(W) - \mathbb{E} F(W)| \geq n^{-\frac{1}{2}+\delta}] \leq 2e^{-c \cdot (n^{-\frac{1}{2}+\delta})^2 / \frac{C}{n}} < n^{-D}$$ for any $\delta, D > 0$

and large n. Using $\mathbb{P}[\|X\| > 2 + \frac{\varepsilon}{2}] \leq n^{-D}$ for any $D > 0$ and large n, and the standard Lipschitz extension argument, we get

$$|v^T G(z) w - \mathbb{E}[v^T G(z) w]| \prec \frac{1}{\sqrt{n}}.$$

Let O be orthogonal, s.t. $Ov = e_1$, $Ow = (v^T w) e_1 + \sqrt{1 - (v^T w)^2} e_2$.

Then by rotation invariance,

$$\mathbb{E}[v^T G(z) w] = (v^T w) \mathbb{E}[e_1^T G(z) e_1] + \sqrt{1 - (v^T w)^2} \, \mathbb{E}[e_1^T G(z) e_2].$$

By symmetry of coordinates,

$$\mathbb{E}[e_1^T G(z) e_1] = \frac{1}{n} \sum_i \mathbb{E}[e_i^T G(z) e_i] = \mathbb{E}[m_n(z)],$$

$$n(n-1) \mathbb{E}[e_1^T G(z) e_2] + n \mathbb{E}[e_1^T G(z) e_1]$$

$$= \sum_{i,j} \mathbb{E}[e_i^T G(z) e_j] = \mathbb{E}[1^T G(z) 1] = n \cdot \mathbb{E}[e_1^T G(z) e_1],$$

the last step using rotation invariance again. Then $\mathbb{E}[e_1^T G(z) e_2] = 0$, and

$$|v^T G(z) w - (v^T w) \mathbb{E}[m_n(z)]| \prec \frac{1}{\sqrt{n}}.$$

Recall $\mathbb{E}[m_n(z)] = -\frac{1}{z} - \frac{1}{z}\mathbb{E}[m_n(z)^2] - \frac{1}{n^2 z}\mathbb{E}[\operatorname{Tr} G(z)^2]$.

On the set $\{W \in \mathbb{R}^{n\times n} : \|\frac{W+W^T}{\sqrt{2n}}\| \le 2+\frac{\varepsilon}{2}\}$, for $z\in\mathbb{D}$:

- $W \mapsto m_n(z)$ is $\frac{C}{n}$-Lipschitz
- $|m_n(z)|, \|G(z)\| \le \frac{2}{\varepsilon}$.

Then a similar Lipschitz extension argument shows

$$-\frac{1}{z}\mathbb{E}[m_n(z)^2] - \frac{1}{n^2 z}\mathbb{E}[\operatorname{Tr} G(z)^2] = -\frac{1}{z}\mathbb{E}[m_n(z)]^2 + O\left(\frac{1}{n}\right).$$

So $\mathbb{E}[m_n(z)] = -\frac{1}{z} - \frac{1}{z}\mathbb{E}[m_n(z)]^2 + r_n$ where $|r_n| \le \frac{C}{n}$.

Then $|\mathbb{E}[m_n(z)] - m(z)| \le \frac{C'}{n}$, and we get

$$\mathbb{P}\left[|v^T G(z) w - (v^T w) m(z)| > n^{-\frac{1}{2}+\delta}\right] < n^{-D}$$

for fixed z.

Now take a grid of values $L \subset \mathbb{D}$ with spacing n^{-5}, of cardinality $|L| \le Cn^{10}$. Since D above is arbitrary,

$$\mathbb{P}\left[\sup_{z\in L} |v^T G(z) w - (v^T w) m(z)| > n^{-\frac{1}{2}+\delta}\right] < n^{-D}.$$

When $\|X\| \le 2+\frac{\varepsilon}{2}$, $z \mapsto v^T G(z) w$ and $z \mapsto m(z)$ are $\frac{4}{\varepsilon^2}$-Lipschitz in z. (E.g. $|\partial_z[v^T G(z) w]| = |v^T G(z)^2 w| \le \|G(z)\|^2 \le \frac{4}{\varepsilon^2}$.) Then

$$\mathbb{P}\left[\|X\| \le 2+\frac{\varepsilon}{2}, \sup_{z\in\mathbb{D}} |v^T G(z) w - (v^T w) m(z)| > n^{-\frac{1}{2}+\delta} + \frac{4}{\varepsilon^2} n^{-5}\right] < n^{-D}.$$

Applying $\mathbb{P}[\|X\| > 2+\frac{\varepsilon}{2}] < n^{-D}$, we get uniformly over $z\in\mathbb{D}$

$$|v^T G(z) w - (v^T w) m(z)| \prec \frac{1}{\sqrt{n}}.$$

For the second statement: Let $f(z) = v^T G(z) w - (v^T w) m(z)$.

Fix $z \in \mathbb{D}$ and let B be a ball of radius $\frac{\varepsilon}{4}$ around z, and ∂B its boundary. On the event $\|X\| \le 2+\frac{\varepsilon}{2}$, f is analytic on B.

Then $f'(z) = \frac{1}{2\pi i}\oint_{\partial B} \frac{f(w)}{(z-w)^2}\,dw$, so

$$|f'(z)| \le \frac{1}{2\pi}\cdot\left(\frac{4}{\varepsilon}\right)^2 \cdot \sup_{w\in B}|f(w)|.$$

If $|f(w)| \le n^{-\frac{1}{2}+\delta}$ for all $w \in \mathbb{D}' = \{w : \operatorname{dist}(w,[-2,2]) \ge \frac{3\varepsilon}{4}, |w| \le C+\frac{\varepsilon}{4}\}$,

then $|f'(z)| \le \frac{1}{2\pi}\left(\frac{4}{\varepsilon}\right)^2 n^{-\frac{1}{2}+\delta}$ for all $z\in\mathbb{D}$.

Applying the first statement to $\mathbb{D}'$, we get the second for $\mathbb{D}$.

Remark: This lemma holds also for a general (non-Gaussian) Wigner matrix X. This is called an <u>isotropic local law</u>, and may be proved using the Schur complement approach of Lecture 3.

Prop: For each $z>1$, $m(z+\frac{1}{z}) = -\frac{1}{z}$. In particular, $m(z)$ is increasing over $z\in(2,\infty)$ with $\lim_{z\to 2} m(z) = -1$ and $\lim_{z\to\infty} m(z) = 0$.

Proof: Recall $m(z+\frac{1}{z})$ satisfies

$$m\left(z+\tfrac{1}{z}\right)^2 + \left(z+\tfrac{1}{z}\right) m\left(z+\tfrac{1}{z}\right) + 1 = 0.$$

Solving yields $m(z+\frac{1}{z}) \in \{-\frac{1}{z}, -z\}$. m is continuous on $(2,\infty)$ and satisfies $m(z) \sim -\frac{1}{z}$ as $z\to\infty$. Then the correct root must be $m(z+\frac{1}{z}) = -\frac{1}{z}$ for all $z>1$. As z increases from 1 to ∞, $z+\frac{1}{z}$ increases from 2 to ∞, and $-\frac{1}{z}$ increases from -1 to 0.

Proof of Thm: Fix $\varepsilon > 0$. Let $\mathcal{E}$ be the event where $\|W\| \le 2 + \frac{\varepsilon}{2}$.

$\hat\lambda$ is eigenvalue of $X \iff 0 = \det(X - \hat\lambda I) = \det(W - \hat\lambda I + V \Lambda V^T)$,

where $V = (v_1 \dots v_k) \in \mathbb{R}^{n \times k}$, $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^{k \times k}$. Two facts about det:

- $\det(AB) = \det A \cdot \det B$ for A, B square.
- $\det(I + AB) = \det(I + BA)$, for $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times n}$.

Then on $\mathcal{E}$:

$\hat\lambda$ is eigenvalue of X in $(2+\varepsilon, \infty) \iff 0 = \det(W - \hat\lambda I) \cdot \det(I + G(\hat\lambda) V \Lambda V^T)$

$$\iff 0 = \det(I + V^T G(\hat\lambda) V \cdot \Lambda),$$

where $G(z) = (W - zI)^{-1}$.

Applying the lemma to each $v_i^T G(z) v_j$,

$$\|V^T G(z) V - m(z) \cdot I\|_\infty \prec \frac{1}{\sqrt{n}}$$

uniformly over $z \in \mathbb{D} = \{z \in \mathbb{C} : \mathrm{dist}(z, [-2, 2]) > \varepsilon, |z| < C\}$.

On $\mathcal{E}$, the entries of $I + V^T G(z) V \Lambda$ and $I + m(z) \Lambda$ are uniformly bounded by a constant over all $z \in \mathbb{D}$. Then

$$(*) \qquad |\det(I + V^T G(z) V \Lambda) - \det(I + m(z)\Lambda)| \prec \frac{1}{\sqrt{n}}$$

uniformly over $z \in \mathbb{D}$. Let $k' = |\{i : \lambda_i > 1\}|$. Then the function

$$z \mapsto \det(I + m(z)\Lambda) = \prod_{i=1}^k (1 + m(z)\lambda_i)$$

is analytic over $z \notin [-2, 2]$, with exactly k' roots,

$$\{\lambda_1 + \tfrac{1}{\lambda_1}, \dots, \lambda_{k'} + \tfrac{1}{\lambda_{k'}}\}.$$

Letting ε, C be s.t. these k' roots are contained in $\mathbb{D}$, $(*)$ and Rouché's Thm imply that for any $D > 0$ and all sufficiently large n, w/ probability $1 - n^{-D}$, the analytic function

$$z \mapsto \det(I + V^T G(z) V \Lambda)$$

also has exactly k' roots in $\mathbb{D}$, which converge to $\{\lambda_1 + \frac{1}{\lambda_1}, \dots, \lambda_{k'} + \frac{1}{\lambda_{k'}}\}$.

Choosing C large enough ensures $\|X\| \le \lambda_1 + \|W\| \le C$ with probability $1 - n^{-D}$. Then, as $\varepsilon > 0$ is arbitrary, this shows

$$\hat\lambda_i \to \lambda_i + \frac{1}{\lambda_i} \quad \text{for } i = 1, \dots, k', \text{ where } \hat\lambda_i \text{ is the } i^{\text{th}} \text{ largest eig. of } X,$$

and the remaining eigenvalues of X converge to $[-2, 2]$, a.s. as $n \to \infty$.

For the eigenvectors: Fix $i \in \{1, \dots, k'\}$, and suppose

$$\hat\lambda_i \hat v_i = X \hat v_i = W \hat v_i + \sum_{j=1}^k \lambda_j (v_j^T \hat v_i) v_j.$$

When $\|X\| \le 2 + \frac{\varepsilon}{2}$ and $\hat\lambda_i > 2 + \varepsilon$, $W - \hat\lambda_i I$ is invertible and so

$$\hat v_i = -\sum_{j=1}^k \lambda_j (v_j^T \hat v_i) G(\hat\lambda_i) v_j.$$

Then for $i' \ne i$, applying the lemma,

$$v_{i'}^T \hat v_i = -\sum_{j=1}^k \lambda_j (v_j^T \hat v_i) v_{i'}^T G(\hat\lambda_i) v_j$$
$$= -\lambda_{i'} (v_{i'}^T \hat v_i) m(\hat\lambda_i) + O_\prec\left(\frac{1}{\sqrt{n}}\right)$$

Note that $m(\hat\lambda_i) \to \frac{-1}{\lambda_i}$, and $\frac{\lambda_{i'}}{\lambda_i} \ne 1$. Then $v_{i'}^T \hat v_i \to 0$ a.s.

Then $\hat v_i = -\lambda_i (v_i^T \hat v_i) G(\hat\lambda_i) v_i + o(1)$.

Taking the squared norm, and applying the lemma and $G'(z) = G(z)^2$,

$$1 = \|\hat v_i\|^2 = \lambda_i^2 (v_i^T \hat v_i)^2 v_i^T G(\hat\lambda_i)^2 v_i + o(1)$$
$$= \lambda_i^2 (v_i^T \hat v_i)^2 \cdot m'(\lambda_i + \tfrac{1}{\lambda_i}) + o(1).$$

Differentiating $m(z)^2 + zm(z) + 1 = 0$ and applying $m(\lambda_i + \frac{1}{\lambda_i}) = -\frac{1}{\lambda_i}$, we get $m'(\lambda_i + \frac{1}{\lambda_i}) = \frac{1}{\lambda_i^2 - 1}$. Then

$$(v_i^T \hat{v}_i)^2 \to \frac{\lambda_i^2 - 1}{\lambda_i^2} = 1 - \frac{1}{\lambda_i^2}.$$

This concludes the proof.

Problem 5: You may assume (without proof) the following isotropic local law: Let $W \in \mathbb{R}^{n \times n}$ be symmetric, $(W_{ij} : i \leq j)$ independent, $\mathbb{E}[W_{ij}] = 0$, $n \cdot \mathbb{E}[W_{ij}^2] = 1 + o(1)$, and $\mathbb{E}[|\sqrt{n} W_{ij}|^k] \leq C_k$ for all k and constants $C_1, C_2, \ldots > 0$. Then for any fixed $v, w \in \mathbb{R}^n$,

$$|v^T (W - zI)^{-1} w - (v^T w) m(z)| < \frac{1}{\sqrt{n}}$$

uniformly over $z \in \mathbb{D} = \{z \in \mathbb{C} : \mathrm{dist}(z, [-2,2]) > \varepsilon, |z| \leq C\}$.

(a) Consider a stochastic block model with two communities: $\{1, \ldots, n\} = S_1 \cup S_2$ where $|S_1| = |S_2| = \frac{n}{2}$. For two vertices $i, j \in \{1, \ldots, n\}$,

$$\mathbb{P}[i \sim j] = \begin{cases} \frac{1}{2} + \frac{c}{\sqrt{n}} & \text{if } i,j \in S_1 \text{ or } i,j \in S_2 \\ \frac{1}{2} - \frac{c}{\sqrt{n}} & \text{if } i \in S_1, j \in S_2 \text{ or } i \in S_2, j \in S_1, \end{cases}$$

where $c > 0$ is a constant. (This includes the self-loop case $i = j$.) Let $A \in \mathbb{R}^{n \times n}$ be the (symmetric) adjacency matrix of this graph, and $\check{A} = A - \frac{1}{2}$ entrywise. Describe the outlier eigenvalues and eigenvectors of $\check{A}$. For which c is there an outlier? What does the eigenvector reveal about S_1 and S_2?

(b) More generally, let $\alpha_1, \ldots, \alpha_k \in (0,1)$ be s.t. $\alpha_1 + \cdots + \alpha_k = 1$. Suppose there are k communities, $\{1, \ldots, n\} = S_1 \cup \ldots \cup S_k$, where $|S_a| = \alpha_a n$. For a function $c : \{1, \ldots, k\} \times \{1, \ldots, k\}$ satisfying $c(a,b) = c(b,a)$ and a constant $p \in (0,1)$, suppose

$$\mathbb{P}[i \sim j] = p + \frac{1}{\sqrt{n}} \cdot c(a,b) \text{ if } i \in S_a, \text{ and } b \in S_b.$$

Let A be the adjacency matrix and $\check{A} = A - p$.

Under what condition does $\check{A}$ have at least one outlier, as $n \to \infty$ and $\alpha_1, \ldots, \alpha_k$, p, and c are fixed?

Under what condition does $\check{A}$ have k outliers?

When would you expect to be able to get good recovery of all k communities $S_1, \ldots, S_k$ based on the spectral decomposition of $\check{A}$?