Outliers in signal-plus-noise models

Consider $X = \lambda V + W$, where

$W \sim \text{GOE}(n)$

$v \in \mathbb{R}^n$, $\|v\|_2 = 1$.

$\lambda > 0$.

What is the behavior of eigenvalues/vectors of $X^2$?

More generally, consider $X = \frac{\lambda}{k} V V^T + W$, where

$v_1, v_k \in \mathbb{R}^n$ are orthogonal

$k$ is a small constant.

For simplicity, $\lambda_1, \lambda_k > 0$.

Prop (Weyl's inequalities): Let $\lambda_1(W), \ldots, \lambda_n(W)$ be eigenvalues of $W$, and let $\lambda_1(X), \ldots, \lambda_n(X)$ be those of $X = \frac{\lambda}{k} V V^T + W$, where $\lambda > 0$. Then for each $j, k$, $0 < j < k < n$,

$$\lambda_j(\lambda_k(W)) < \lambda_j(X) < \lambda_k(\lambda_j(W)).$$

Proof: For the lower bound, by Courant-Fischer

$$\lambda_j(X) = \max \left\{ \min_{v \in \text{span}(u)} \sqrt{X v} : V S R^n, \text{Im}(v) = \mathbb{R}^\perp \right\}$$

$$= \max \left\{ \min_{v \in \text{span}(u)} \sqrt{W v} : V S R^n, \text{Im}(v) = \mathbb{R}^\perp \right\} = \lambda_j(W).$$

Since $\sqrt{\lambda_j(W) V V^T} = \frac{\sqrt{\lambda_j(W)}}{\sqrt{k}} X$, $\lambda_j(X) > 0$. This is true for any $X = \lambda W + W'$ when $P$ is positive semi-definite.

For the upper bound, it suffices to show for any $VS R^n$ with $\text{Im}(v) = \mathbb{R}^\perp$,

$$\Rightarrow \lambda_j(X) = \lambda_j(\lambda_k(W)) + \lambda_j(W) = \lambda_j(W).$$

The existence of $v \in \text{span}(v_1, v_k)$ and the top $j-1$ eigenvectors of $W$: This has dimension at most $j-1$. Then the orthogonal complement $U^\perp$ has dimension at least $n-j$, so $U^\perp$ intersects $V$. But for any unit vector $v \in U^\perp$, $\sqrt{X v} = \sqrt{W v} = \lambda_1(W)$ becomes $v$ is orthogonal to $v_1, v_k$, and the top $j-1$ eigenvectors of $W$.

Implication: The empirical distributions of $\frac{\lambda}{k} \mu_j(W)$ and $\frac{\lambda}{k} \mu_j(X)$ both converge to the semicircle laws. Then so does that of $\frac{\lambda}{k} \mu_j(X)$. Furthermore, since $\|W\| \rightarrow 2$, we have $\mu_j(X) \rightarrow 2$ and $\mu_j(X) \rightarrow 2$.

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

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

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

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

Then $\lambda \in \text{span}(v)$. This would imply

$$= \lambda \sqrt{\lambda(W)} = \lambda \sqrt{\lambda(W)}$$

$$= \lambda \sqrt{\lambda} = (W^*) v.$$
Since $\|W\| = 2$, $W \times I$ should be invertible with high probability.

Let $G(x) = (W \times I)^{1/2}$ be the rescaled. Then
\[ G(x) \in (W \times I)^{1/2} \]

Multiply on the left by $V$, and cancel $V^T$ on both sides:
\[ 1 = -\lambda (x_0) \cdot G(x) \cdot v \]

Let $O$ be orthogonal s.t. $Ov = e_1$. Then for any $j \neq 2$,
\[ V^T G(x) = e_1^T O (W \times I)^{1/2} O^T e_1 \]
\[ = e_1^T O W O^T e_1 = e_1^T (W \times I) e_1 = G(x) \]
So we expect $V^T G(x) v \approx \mathbb{E}[V^T G(x) v] \approx \mathbb{E}[G(x) v] = \mathbb{E}[m(x)]$ where $m(x) = \frac{\lambda}{\alpha} \mathbb{E}[\mathcal{E}(x)]$ is the empirical Stirling's transform of $W$.

Letting $m(x)$ be the Stirling's transform of the semi-circle law,
\[ \sqrt{G(x)} v \approx m(x) \]
So $m(x) \approx -x$. Our $m(x)$ is $(2,0)$-wise independent.

So this is only possible when $\lambda > 1$. If not, applying
\[ m(x) \cdot \sum m(x) + \epsilon \]
and substituting $m(x) \approx -x$ we get for the observed eigenvalue $\hat{\lambda}$
\[ \frac{\hat{\lambda}}{\lambda} \approx \frac{1}{\sqrt{2}} + 1 \approx 1.618 \]
Returning to (i) now take the squared norm on both sides. Then
\[ 1 = \|G(x)\|^2 = \lambda^2 (V^T G(x) v)^2 = \lambda^2 (V^T)^2 V^T G(x) v \]

Note that $\frac{\partial}{\partial x} G(x) = \frac{1}{2} (W \times I)^{1/2}, \quad I + (W \times I)^{1/2} \in G(x)$.

So we expect $V^T G(x)v = \mathbb{E}[V^T G(x)v] \approx \mathbb{E}[m(x)] \approx m(x)$.

Differentiating $m(x)^2 + \sum m(x) = 0$ in $x$,
\[ \sum m(x) m(x) + \sum m(x) = 0 \Rightarrow m(x) = -\frac{m(x)^2}{2m(x)} \Rightarrow \frac{x^2}{2} \approx \frac{1}{\lambda - 1} \]
Then $\frac{V^T}{V^T G(x)} \approx \frac{x^2}{2} \approx 1 - \frac{1}{\lambda - 1}$.

Then let $X = \frac{1}{\lambda - 1} (W \times I) \Rightarrow (W \times I)^{1/2}$ is orthogonal, and $\lambda > 1, \lambda > 0$. Then $a.s.$ as $n \to \infty$

- Corresponding to each $\lambda > 1$ there is an eigenvalue $\hat{\lambda}$ of $X$ which converges to $\lambda + \frac{1}{\lambda}$.
- Letting $\hat{V}$ be the associated unit eigenvector
  \[ (V^T \hat{V})^2 \to 1 - \frac{1}{\lambda} \text{ and } \hat{V} \text{ is unique for all } j \in \{1, \ldots, k\} \]

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

Three qualitative phenomena:
- Phase transition: Only $\lambda$'s in the range $1$ yield an outlier $\hat{\lambda}$'s less than $1$ was caused by the noise.
- Upward bias: $\hat{\lambda} \to \lambda \frac{1}{\lambda}$ and $\hat{\lambda}$. There is an upward bias $\frac{1}{\lambda}$. 
- Eigenvector inconsistency: $(V^T \hat{V})^2 \to 1 - \frac{1}{\lambda}$ and $1$. $\hat{V}$ lies on a cone around the true $V_j$ when the angle is a decreasing function of $\lambda$. 

Lemma: Let $X \sim G \mathcal{E}(n)$, $G(x) = (x - c)^2$, with the semi-circle distribution. For my constants $\varepsilon, C > 0$ and any deterministic unit vectors $v, w \in \mathbb{R}^n$,

$$
\left| v^T (G(w) - (w) m(x)) \right| < \frac{\varepsilon}{\sqrt{n}}
$$

unifromly over $x \in C, \text{dist}(x, [2,2]) \geq C$, $1 \leq C \leq 3$.

(Recall: This seems for any $\varepsilon > 0$ and all large $n$.

$$
P_x \left[ \exists z \in \mathbb{Z}^2 : \left| |G(w_m(x)) - \frac{1}{2} \right| \geq C \right] \leq C n^{-D},$$

and similarly for the second statement.)

Proof: First consider fixed $z$. Write $X \sim W_{\mathbb{Z}^2}$ when $W_{\mathbb{Z}^2}$ is $n(x)$, and view $v^T G(w)$ as a function of $W$. Recall

$$D_{\mathbb{Z}} G(x) = \frac{1}{(2 \pi)^{n}} G(0) (D_{\mathbb{Z}} G_{\mathbb{Z}} + D_{\mathbb{Z}} G_{\mathbb{Z}^2}) G(0),$$

where $D_{\mathbb{Z}} G_{\mathbb{Z}}$ has the single entry $(z)$ equal to $1$. Then

$$D_{\mathbb{Z}} [v^T G(w)] = \frac{1}{(2 \pi)^{n}} \sum_{z} (v^T G(0) \cdot G(0) w + v^T G(0) \cdot G(0) w).$$

So, we get

$$\|D_{\mathbb{Z}} [v^T G(w)]\| \leq \frac{\varepsilon}{\sqrt{n}}.$$
Recall \( E[M_n(t)] = -\frac{1}{2} \frac{\lambda}{n} E[M_n(t)^2] + \frac{1}{n^2} E[T G(b)^2] \).

On the set \( \{ W \in P^n: \| W \|_n \leq 2 \sqrt{\frac{\lambda}{n}} \} \), for \( z \in \mathbb{D} \),
\[
  \| m_n(z) \| \leq \frac{2}{\sqrt{\lambda}}.
\]

Then a similar Lipschitz extension argument shows
\[
  -\frac{1}{2} E[M_n(t)]^2 - \frac{1}{n} E[T G(b)^2] = -\frac{1}{2} E[M_n(t)]^2 + O(\frac{1}{n}).
\]

So \( E[M_n(t)] = -\frac{1}{2} - \frac{1}{n} E[M_n(t)]^2 + c \) when \( |z| \leq \frac{2}{\sqrt{\lambda}} \).

Then \( |E[M_n(t)] - m(t)| \leq \frac{2}{\sqrt{\lambda}}/n \), and we get
\[
  P\left[ \sup_{z \in \mathbb{D}} \left| \sqrt{\lambda} G(b) w - \sqrt{\lambda} m(t) \right| > n^{-1/2} \right] \leq n^{-D}.
\]

For fixed \( t \).

Now take a grid of values \( z \in \mathbb{D} \) with spacing \( n^{-5} \), say,

and cardinality \( \| z \| \leq C n^0 \). Since \( D \) area is bounded,
\[
  P\left[ \sup_{z \in \mathbb{D}} \left| \sqrt{\lambda} G(b) w - \sqrt{\lambda} m(t) \right| > n^{-1/2} \right] \leq n^{-D}.
\]

When \( |X| \leq 2 \sqrt{\frac{\lambda}{n}} \), \( 2 \sqrt{\lambda} G(b) w + \sqrt{\lambda} m(t) \) is \( \frac{1}{C} \)-Lipschitz in \( \mathbb{D} \). (E.g., \( |G(b) w| \leq |G(b) w| + |G(b) w| \leq \frac{\lambda}{C} \).)

Then
\[
  P\left[ \sup_{z \in \mathbb{D}} \left| \sqrt{\lambda} G(b) w - \sqrt{\lambda} m(t) \right| > n^{-1/2} \right] \leq n^{-D}.
\]

Apply \( P\left[ |X| > 2 \sqrt{\frac{\lambda}{n}} \right] \leq n^{-D} \), we get uniformly over \( z \in \mathbb{D} \)
\[
  \left| \sqrt{\lambda} G(b) w - \sqrt{\lambda} m(t) \right| < \frac{1}{C}.
\]

For the second statement: Let \( f(t) = v^2 G(b) w - (J_w) m(t) \).

Fix \( z \in \mathbb{D} \) and let \( B \) be a ball of radius \( 2 \sqrt{\lambda} \) around \( z \), and \( \partial B \) its boundary. On the event \( |X| \leq 2 \sqrt{\frac{\lambda}{n}} \), \( f \) is analytic on \( B \).

Then \( f(t) \leq \frac{1}{2^n} \int_B f(w) \frac{\partial G(b)}{\partial z_{2n}} dw \), so
\[
  |f(t)| \leq \frac{1}{2^n} \left( \frac{\lambda}{2} \right)^{1/2} \sup_{w \in B} |f(w)|.
\]

If \( |f(w)| \leq n^{-1/2} \) for all \( w \in \mathbb{D} \), \( f(w) \) is bounded in \( B \). In fact
\[
  |f(t)| \leq \frac{1}{2^n} \left( \frac{\lambda}{2} \right)^{1/2} n^{-1/2} \text{ for all } z \in \mathbb{D}.
\]

Applying the first statement to \( f \), we get the second for \( D \).

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

Prop: For each \( t \in [0, 1), m(t + \frac{1}{2}) = \frac{1}{2} \). In particular, \( m(t) \) is increasing on \( [0, 1) \) and \( \lim_{t \to 1} m(t) = 1 \) and \( \lim_{t \to 0} m(t) = 0 \).

Proof: Recall \( m(t + \frac{1}{2}) \) satisfies
\[
  m(t + \frac{1}{2}) + (t + \frac{1}{2}) m(t + \frac{1}{2}) = 1.
\]

Solving yields \( m(t + \frac{1}{2}) \equiv \frac{1}{2} \). \( m \) is continuous on \( [0, 1] \)

and \( m(t + \frac{1}{2}) \equiv \frac{1}{2} \) as \( t \to 0 \). Then the correct rate must be
\[
  m(t + \frac{1}{2}) = \frac{1}{2} \text{ for all } t \in [0, 1].
\]

As \( t \) increases from \( 1 \) to \( 0 \), \( m(t + \frac{1}{2}) \) increases from \( 1 \) to \( 0 \) and \( -\frac{1}{2} \) increases from \( -1 \) to \( 0 \).
Proof of Theorem 1: Fix $\delta > 0$. Let $E$ be the event where $\|W\| \leq 2 + \frac{\delta}{n}$. 

$\hat{S}$ is eigenvalue of $X \iff 0 \in \text{det}(X - \hat{S}I) = \text{det}(W, \hat{S}I + VA^*)$,  
where $V \in \{v_1, \ldots, v_N\} \subset R^{m \times K}$, $A \in \text{Sym}(K \times K) \subset R^{K \times K}$. Two facts about $\text{det}$:

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

Then on $E$:

$\hat{S}$ is eigenvalue of $X$ in $(0, \infty) \implies 0 \in \text{det}(W, \hat{S}I) = \text{det}(I + VA^*)$.  

$\implies 0 \in \text{det}(I + V^*G(\hat{S}))A^*)$.  

where $G(\hat{S}) = (W - \hat{S}I)^{-1}$.

Applying the lemma to each $v_i^* G(\hat{S}) v_j$, we have  

$$\|v_i^* G(\hat{S}) v_j\| = m(\hat{S}) \|v_i\| \|v_j\| \leq \frac{\delta}{\sqrt{n}}$$

uniformly over $z \in D = \{z \in C : \text{det}(z, \mathbb{C}^2) > 0, \|z\| < 1\}$.

On $E$, the entries of $I + V^* G(\hat{S}) A^* V$, and $I + VA^*$ are uniformly bounded by a constant over all $z \in \mathbb{C}^2$. Then

$$|\text{det}(I + V^* G(\hat{S}) A^*)| \leq \frac{1}{m(\hat{S})}$$

uniformly over $z \in D$. Let $k' = \lfloor \frac{1}{\delta^2} \log(1 + 13\delta) \rfloor$ Then the function $z \mapsto \text{det}(I + \lambda G(\hat{S}) A^*)$ is analytic over $-z \in [-2, 2]$, with exactly $k'$ roots:

$$\lambda \in \{\lambda \in Z : -2 < \lambda < 2\}.$$  

Letting $\epsilon, \delta > 0$, these $k'$ roots are contained in $D$.

$(\epsilon)$ and Rouché's Theorem imply that for any $D > 0$ and all sufficiently large $n$,  

$W$ is probability $1 - n^{-D}$ the analytic function

$$z \mapsto \text{det}(I + V^* G(\hat{S}) A^*)$$

also has exactly $k'$ roots in $D$, which converge to $\lambda_k^* - \lambda_k$ as $n \to \infty$.

Choosing $D$ large enough ensures $\|W\| \leq \lambda_1^* = \lambda_k^* + k'$.

Then $\|W\| \leq 2 + \frac{\delta}{n}$, which proves the result.

For the eigenvectors: Fix $z \in \mathbb{C}^2$, $k', \epsilon > 0$, and suppose

$$\lambda \in \mathbb{R} \subset (0, \infty) \implies \text{det}(I + \lambda G(\hat{S}) A^*) = 0.$$  

Then for $\lambda = \frac{\lambda}{\lambda_k}$, applying the lemma,

$$v_i^* V = \frac{1}{\lambda_k} \lambda (v_i^* G(\hat{S})) v_j$$

Note that $m(\lambda_k) \to \frac{1}{m(\hat{S})}$ as $n \to \infty$.

Then $\|V\| \leq \frac{1}{m(\hat{S})}$, and $\frac{1}{m(\hat{S})} \|V\| \leq 1$. Then $v_i^* V \rightarrow 0$ a.s. as $n \to \infty$.

Taking the square norm and applying the lemma and $G(\hat{S}) = G(\hat{S})^*$,

$$1 = \|V\|^2 |\lambda_k^* (v_i^* G(\hat{S})) v_j^* G(\hat{S}) v_i^* G(\hat{S}) v_j v_i^* G(\hat{S}) v_j^* G(\hat{S}) v_i^* G(\hat{S}) v_j + o(1)|$$

$$= \lambda^2 (v_i^* G(\hat{S}))^2 m(\lambda) \frac{1}{m(\lambda)} + o(1).$$
Differentiating \( m(x)^2 + a(x^2) + b = 0 \) and applying \( m(x)^2 + b = -\frac{1}{a^2} \), we get \( m(x)^2 + b = \frac{1}{a^2} \). Then

\[
\left( \frac{v_i v_j}{x_i x_j} \right) \rightarrow \frac{a^2}{x_i^2} - 1 - \frac{b}{x_i^2}.
\]

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, \( E(W_{ij} \neq 0) \) independent, \( \mathbb{E}[W_{ij}^2] = 0 \), \( \mathbb{E}[W_{ij}] = 0 \), and \( \mathbb{E}[\text{tr}(W)^k] \leq C_k \). For all \( k \) and constants \( C_k \), \( C_{k,2} \), \( C_{k,3} \), \( C_{k,4} \), \( C_{k,5} \) uniformly over \( x \) and \( \epsilon \in D = \{ z \in C : \text{dist}(z, E(x)) \leq \epsilon \} \), \( 1 \in C(\alpha) \).

(a) Consider a stochastic block model with two communities:

\[
\{ u_1, u_2 \} = S_1 S_2 \quad \text{where} \quad |S_1| = |S_2| = \frac{n}{2}.
\]

For two vertices \( u \in S_1, v \in S_2 \),

\[
\mathbb{P}(x_{uv}^2) = \begin{cases} \frac{1}{2} + \frac{c}{n} & \text{if} \quad v \in S_1, \quad u \in S_2 \\ \frac{1}{2} - \frac{c}{n} & \text{if} \quad u \in S_1, \quad v \in S_2 \end{cases}
\]

where \( c > 0 \) is a constant. (This includes the self-loop case \( x_{ii} \).)

Let \( A \in \mathbb{R}^{n \times n} \) be the (symmetric) adjacency matrix of this graph, and \( \hat{A} = A - \frac{1}{n} \) entrywise. Describe the outlier eigenvalues and eigenvectors of \( \hat{A} \). For which \( c \) is there an outlier? What does the eigenvector reveal about \( S_1 \) and \( S_2 \)?

(b) More generally, let \( \alpha_{1,2} \ldots \alpha_k \in \mathbb{C} \) be real, \( \alpha_1, \ldots, \alpha_k = 1 \).

Suppose there are \( k \) communities, \( S_1, u_1, u_2, \ldots, u_n \), where \( k \leq n \).

For a function \( c : \{ 1, 2 \} \times \{ 1, \ldots, k \} \) satisfying \( c(1, b) = c(2, b) \) and a constant \( p \in (0, 1) \), suppose

\[
\mathbb{P}(x_{uv}^2) = p + \frac{c}{n} c(1, b) \quad \text{if} \quad v \in S_1, \quad u \in S_2, \quad \text{and} \quad b \leq 5.
\]

Let \( A \) be the adjacency matrix and \( \hat{A} = A - \frac{1}{n} \).

Under what conditions does \( \hat{A} \) have at least one outlier, as \( n \to \infty \) and any \( p \) and \( c \) are fixed?

Under what conditions does \( \hat{A} \) 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 \( \hat{A} \)?