Spiked Covariance Model and BBP Phase Transition

Recall: For $Y \sim \mathcal{N}(0, \Sigma)$ with independent rows of mean 0, covariance $\Sigma$, 
\[
\Sigma = \frac{1}{n} Y^T Y \Rightarrow \Sigma \text{ is the sample covariance matrix.}
\]

If $\Sigma = I$ and entries of $Y$ are independent (with bounded moments of all orders), then as $n, p \to \infty$ with $\sqrt{n} \to y \in (0, \infty)$, the ESD of $\Sigma$ converges weakly a.s. to the Marčenko–Pastur law $f_y$. This has density
\[
\rho_y(x) = \frac{1}{2\pi y x} \sqrt{(y^2 - x)(x - \nu)} I_{\nu < x < y^2} \quad \text{where } \nu = \left(1 + \sqrt{y^2 - 1}\right)^2
\]
plus a point mass at 0 if $y > 1$.

If $X \in \mathbb{R}^p$ is Gaussian with $\mathbb{E}[X] = 0, \mathbb{E}[X^T X] = I_p$, then $\mathbb{P} \left[ \lambda^T \mathbf{X} \mathbf{X}^T \lambda < 1 \right] = 1 - \mathbb{E}[\lambda^T \mathbf{X} \mathbf{X}^T \lambda]$

- A fixed number $d$ of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d > 1$ of $\Sigma$ are larger than 1.
- All other eigenvalues $\lambda_{d+1} = \lambda_{d+2} = \cdots = \lambda_p$ of $\Sigma$ equal 1.

This is called the spiked covariance model (Johnstone '01).

At the population level, the covariance $\Sigma$ has $d$ "signal" principal components.

Each sample (row of $Y$) in $\mathbb{R}^p$ has significant variation along these $d$ directions, and isotropic noise in the remaining $p - d$ directions.

What is the behavior of principal components of the sample covariance $\hat{\Sigma}$?

Thus (Bai, Ben Arous, Peché '05): Let $X_{ij}$ be complex Gaussian: $X_{ij} \sim \mathcal{N}(0, \Sigma)$.

- Let $X_{ij} \sim \mathcal{N}(0, \Sigma)$. Define $Y_{ij} = X_{ij}^2$. Suppose $n, p \to \infty$ with $\sqrt{n} \to y \in (0, \infty)$, and $c, d, \nu, \lambda$ are fixed. Then:
(a) If all eigenvalues $\lambda_1, \lambda_2 < 1+\sqrt{8n}$, then the largest eigenvalue $\lambda_{\text{max}}$ satisfies

$$
\lambda_{\text{max}} \geq (1+\sqrt{8})^\frac{n}{2} - \frac{\lambda_1 + \lambda_2}{2}.
$$

(b) If $\lambda_1 = \lambda_2 > 1+\sqrt{8}$ and $\lambda_3, \lambda_4 \leq 1+\sqrt{8}$, then

$$
\lambda_{\text{max}} \geq (1+\sqrt{8})^\frac{n}{2} - \frac{\lambda_3 + \lambda_4}{2}.
$$

(c) If $X$ is real Gaussian, $\lambda_1 \sim N(0,1)$, and $\lambda_2 > (1+\sqrt{8})^\frac{n}{2}$, then

$$
\lambda_{\text{max}} \geq (1+\sqrt{8})^\frac{n}{2} - \frac{\lambda_1 + \lambda_2}{2}.
$$

(1) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(2) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(3) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(4) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(5) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(6) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(7) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(8) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(9) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(10) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(11) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(12) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(13) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(14) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(15) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(16) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(17) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(18) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(19) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(20) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(21) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(22) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(23) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(24) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(25) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(26) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(27) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(28) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(29) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(30) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(31) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(32) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(33) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(34) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(35) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(36) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(37) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(38) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(39) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(40) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(41) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(42) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(43) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(44) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(45) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(46) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(47) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(48) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(49) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(50) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(51) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(52) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(53) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(54) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(55) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(56) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(57) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(58) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(59) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$.

(60) If $X_1, X_2, X_3$ are a certain generalization of $F_{\text{GOE}}$. 

Remark: (a) Illustrate a phase transition phenomenon: If $\lambda < 1+\sqrt{8}$, the signal is covered by the spectral noise. If $\lambda > 1+\sqrt{8}$, an outlier eigenvalue separates from $\mu$. This is now called the "Bulk-Edge-Across-Peaks" (BEAP) phase transition.

Remark: For real Gaussian $X$, as well as $X \in \mathbb{C}$, the limiting eigenvalue distributions for the first-order (almost sure) limits and second-order (distributional) limits are the same. However, the second-order (distributional) limits are different.

Remark: For real Gaussian $X$, as well as $X \in \mathbb{C}$, the limiting eigenvalue distributions for the first-order (almost sure) limits and second-order (distributional) limits are the same. However, the second-order (distributional) limits are different.

Remark: For real Gaussian $X$, as well as $X \in \mathbb{C}$, the limiting eigenvalue distributions for the first-order (almost sure) limits and second-order (distributional) limits are the same. However, the second-order (distributional) limits are different.

Remark: For real Gaussian $X$, as well as $X \in \mathbb{C}$, the limiting eigenvalue distributions for the first-order (almost sure) limits and second-order (distributional) limits are the same. However, the second-order (distributional) limits are different.
(5) In general, it's not possible to consistently estimate $\lambda_k$ as $n \to \infty$ (if $\lambda_k$ remains bounded). We know $\lambda_k$ should be near a certain curve, controlled around $\tilde{\lambda}_k$, but the direction of error is unknown and (almost uniformly) random.

However, for a fixed unit vector $u \in \mathbb{R}^p$ (not depending on the data), we can consistently estimate $|\lambda_k|^2 u^2$. Note that $|\lambda_k|^2 u^2$ is downward biased. Write $u = \alpha_k v_k + \tilde{w}$, where $\tilde{w}$ is orthogonal to $v_k$. Then $\langle v_k, \tilde{w} \rangle \to 0$, so

$$|\langle \tilde{v}_k, u \rangle| = \alpha_k |K(v_k, v_k)| \to (1)$$

$$\rightarrow \alpha_k \frac{1}{1 + \frac{1}{\lambda_k}} = |K(v_k, v_k)| \frac{1}{1 + \frac{1}{\lambda_k}}$$

Here $\alpha_k \in (0, 1)$ is unknown but may be estimated by plugging $u = \tilde{v}_k$ for $\lambda_k$. Letting this plug-in estimate be $\hat{\lambda}_k$, we see that $\frac{\hat{\lambda}_k}{K(v_k, v_k)}$ is consistent for $|\lambda_k|u^2$.

(3) This is all assuming $EX_k^2 = 1$ for all $i,j$, i.e. $\Sigma = I +$ low-rank perturbation. It is more common to assume $\Sigma = \delta I +$ low-rank, and when $I$ may further be restricted to a $\beta$-dimensional subspace of $\mathbb{R}^p$ for some “effective dimension” $\beta$. The problem $\delta \beta^2$ may be estimated by matching the slope of the eigenvalue distribution to the Marcenko-Pastur law—see Patterson, Price, Reich “Population structure and eigenvalues” (2006) for one implementation of this idea.
Proof of Theorem (a): Recall \( \Sigma = \frac{1}{n} XX^t = \frac{1}{n} \Sigma X_i X_i^t \). Note that
\[
\Sigma \Sigma = \mathbf{V} (\Sigma X_i, X_i)^t \mathbf{V}
\]
for the eigendecomposition \( \Sigma = \mathbf{V} \Sigma \mathbf{V}^t \) of \( \Sigma \), and
\[
\mathbf{V} X_i X_i^t \mathbf{V}^t
\]
is equal in distribution to \( X_i X_i^t \) by rotational invariance of \( N(0, I) \).

So, we may assume \( \Sigma = \begin{pmatrix} \Sigma_1 & \Sigma \Sigma_2 \\ \Sigma \Sigma_1^t & \Sigma_2 \end{pmatrix} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \).

Write
\[
\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma \Sigma_2 \\ \Sigma \Sigma_1^t & \Sigma_2 \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \Sigma \Sigma_2 = \frac{1}{n} X_2 X_2^t
\]
then \( \Sigma \Sigma_2 = \frac{1}{n} X_2 X_2^t \) is a white sample covariance matrix. We assume:

Thm (Silverstein '85, Bai '93): All eigenvalues of \( \Sigma \Sigma_2 \) converge a.s.

to \( \lambda_{\Sigma_2} \Sigma_2 (0, I) \).

This may be proven by either the moment method or Stieltjes transform

method as in the Wigner case.

Then for any eigenvalue \( \lambda \) of \( \Sigma \) outside \( (0, I)^2 \), \( 0 = \det (\Sigma - \lambda I) \) \( \Rightarrow \lambda \in \text{roots of } \det (\Sigma_1 - \lambda I - \Sigma \Sigma_2 (\Sigma_2 + \Sigma_1 - 1) \Sigma_2) \)

\[ \Rightarrow 0 = \det (\Sigma_1 - \lambda I - \Sigma_2 (\Sigma_2 - 1) \Sigma_2) \text{ Resolvent formula} \]

Where \( \text{Resolvent formula} \) is the resolvent of \( \Sigma_2 \).

We have
\[
\Sigma_1 - \Sigma_2 (\Sigma_2 - 1) \Sigma_2
\]
\[ = \frac{1}{n} \Sigma_1 X_i X_i^t \Sigma_1^t - \frac{1}{n} \Sigma_1 X_i X_i^t (\Sigma X_i X_i^t - 1) X_i X_i^t \Sigma_1^t \]
\[ = \frac{1}{n} \Sigma_1 X_i X_i^t \left( I - \frac{1}{n} X_i (\Sigma X_i X_i^t - 1) X_i^t \right) X_i X_i^t \Sigma_1^t \]
\[ = \frac{1}{n} \Sigma_1 X_i X_i^t \left( I - \frac{1}{n} X_i (\Sigma X_i X_i^t - 1) X_i^t \right) X_i X_i^t \Sigma_1^t \]
\[ \Rightarrow \lambda \in \text{roots of } (1 - \frac{1}{n} X_i (\Sigma X_i X_i^t - 1) X_i^t) \lambda_1 \]

By same argument as in Wigner case, other eigenvalues \( \lambda \) converge to roots of \( 0 = \det k(\lambda) \Rightarrow 0 = (1 - \frac{1}{n} X_1 (\Sigma X_1 X_1^t - 1)) \lambda_1 \)

Recall \( \lim \) satisfies \( \lambda_2 \lim \frac{1}{n} X_2 X_2^t = 1 \), \( \lambda_2 \lim \frac{1}{n} X_2 X_2^t = 1 \), \( \lambda_2 \lim \frac{1}{n} X_2 X_2^t = 1 \).

Thus we see \( \Sigma \) converges to \( \Sigma_2 \).
satisfying \( \lambda_k = \frac{2m(0)}{M_2 m(0)} \).

We have \( 2m(0) = \int_0^1 \int_0^1 \int_0^{\lambda_1} f(x, y) \, dy \, dx \) and

<table>
<thead>
<tr>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strictly increasing in ( \lambda ) over ((-\infty, 1/\sqrt{2})) and ((1/\sqrt{2}, \infty))</td>
<td></td>
</tr>
<tr>
<td>Positive for ( \lambda \in (-\infty, 1/\sqrt{2}) ) less than 1 for ( \lambda \in (1/\sqrt{2}, \infty) )</td>
<td></td>
</tr>
<tr>
<td>Approaches (-\infty ) as ( \lambda \to 1/\sqrt{2} )</td>
<td></td>
</tr>
<tr>
<td>Also increasing in ( \lambda ) over ((-\infty, 1/\sqrt{2})) and ((1/\sqrt{2}, \infty))</td>
<td></td>
</tr>
<tr>
<td>Less than 1 for ( \lambda \in (-\infty, 1/\sqrt{2}) ) greater than 1 for ( \lambda \in (1/\sqrt{2}, \infty) )</td>
<td></td>
</tr>
<tr>
<td>Approaches (-\infty ) as ( \lambda \to 1/\sqrt{2} )</td>
<td></td>
</tr>
</tbody>
</table>

Thus, \( \lambda_k = \frac{2m(0)}{M_2 m(0)} \) has a root if and only if \( \lambda_k > 1/\sqrt{2} \). In this case, we check that the root is \( \lambda_k = \lambda_k (1 + \sqrt{2}) \) as desired.

Proof of Thm (c): Again, it suffices to consider \( \Sigma = (\lambda_k, \theta_k) \). Then the population eigenvalues \( \nu_k \), \( \nu_k \) are the standard basis vectors \( e_{\nu_k} \).

For \( \lambda_k > 1/\sqrt{2} \), write the eigenvector \( \nu \) for \( \lambda_k = \lambda_k (1 + \sqrt{2}) \) as \( (\nu, \nu) \), and

\[
\lambda_k \nu = \left( \begin{array}{c} \nu_k \\ \nu_k \end{array} \right),
\]

\[
\Rightarrow 0 = (S_{11} - \tilde{S}_{11}) \nu_k + \tilde{S}_{11} \nu_k,
\]

\[
0 = \tilde{S}_{11} \nu_k + (S_{11} - \tilde{S}_{11}) \nu_k.
\]

Since \( S_{11} - \tilde{S}_{11} \) is invertible as for all large, letting \( \tilde{R}(t) = (\tilde{S}_{11} - \tilde{S}_{11})(1 + \tilde{S}_{11}) \),

\[
\tilde{S}_{11} = \tilde{R}(t) : \tilde{S}_{11} \nu_k,
\]

We deduce from this:

\[
(1 + \tilde{S}_{11}) \tilde{R}(t) \tilde{S}_{11} = \tilde{S}_{11} (I + \tilde{S}_{11} R(t) \tilde{S}_{11}) \tilde{S}_{11},
\]

\[
0 = \tilde{S}_{11} (\tilde{S}_{11} - \tilde{S}_{11} R(t) \tilde{S}_{11}) \tilde{S}_{11}.
\]

Recall \( \tilde{S}_{11} (\tilde{S}_{11} - \tilde{S}_{11} R(t) \tilde{S}_{11}) \tilde{S}_{11} \to \tilde{R}(t) \equiv (1 - \gamma - \gamma \tilde{R} \gamma(t)) \tilde{S}_{11} = \tilde{S}_{11} \to \tilde{I} \). Hence

\[
\tilde{S}_{11} (\tilde{S}_{11} - \tilde{S}_{11} R(t) \tilde{S}_{11}) \tilde{S}_{11} \to \tilde{R}(t) \equiv (1 - \gamma - \gamma \tilde{R} \gamma(t)) \tilde{S}_{11} = \tilde{S}_{11} \to \tilde{I}.
\]

This and (c) imply \( \tilde{e}_k \to 0 \) for all \( k \), so \( \tilde{e}_k = 0 \).

Note that in (c), \( \tilde{e}_k \tilde{R}(t) \tilde{S}_{11} = -\tilde{R}(t) \). By a continuity argument (as in the Wigner case), \( \tilde{R}(t) \to \tilde{I} \) entrywise.

Let \( \lambda_k(t) = \frac{2m(0)}{1 + \tilde{S}_{11}(t)} \Rightarrow 2m(t) = \frac{\lambda_k(t)}{1 + \lambda_k(t)} \).

Thus \( \frac{d}{dt} [2m(t)] = \lambda_k(t) \frac{1}{(1 - \lambda_k(t)) \tilde{R}(t)} \tilde{S}_{11} \). Also differentiating \( 2m(t) \lambda_k(t) \)

\[
= (1 + \frac{\lambda_k(t)}{\tilde{S}_{11}(t)} - \frac{\lambda_k(t)}{\tilde{S}_{11}(t)} \lambda_k(t)) \tilde{R}(t) = (1 - \tilde{S}_{11}(t) \tilde{S}_{11} - \lambda_k(t)).
\]

Applying \( \lambda_k(t) = \lambda_k \)

\[
K'(\tilde{e}_k) = -\frac{\lambda_k(t)}{\tilde{S}_{11}(t) - \lambda_k(t)} (1 + \tilde{S}_{11}(t) - \lambda_k(t)) = -\frac{1 + \lambda_k(t)}{(1 + \lambda_k(t))}.
\]

From \( \tilde{e}_k = \frac{1}{i} \tilde{K}(t) \tilde{e}_k \to 0 \) and \( \tilde{S}_{11} \)

\[
1 = ||\tilde{S}_{11}(t) - \tilde{S}_{11}(t)\tilde{S}_{11}(t)|| = ||\tilde{S}_{11}(t)\tilde{S}_{11}(t) - \tilde{S}_{11}(t)|| / (1 + \lambda_k(t)) \to 0.
\]
\[ \Rightarrow \mathbb{C}^N, x^2 \rightarrow (1 - \mathbb{C}^N) / (1 + \mathbb{C}^N). \]

Thus:
\[ \left\{ \begin{array}{l}
q_j \rightarrow 0 \\
q_j \rightarrow 1 - \frac{n}{2} \mathbb{C}^N \end{array} \right. \quad j = k. \]

Finally, observe that the distributions of \( \hat{Z}_n \) is rotationally invariant on \( \mathbb{R}^p \), because the law of \( \hat{Z}_n \) does not change under the transformation \( \mathbb{R}^p \rightarrow \mathbb{R}^p \).

To get \( \text{span} (\hat{E}_n, \ldots, \hat{E}_p) \), we can use:
\[ \frac{1}{\mathbb{E}^n} \rightarrow \frac{1}{\mathbb{E}^p} \quad \text{as} \quad n \rightarrow \infty. \]

By a concentration of measure argument, \( 1 - \mathbb{E}^p \rightarrow 0 \).

Combining with \( 1 - \mathbb{E}^n \rightarrow 0 \), we get \( 1 - \mathbb{E}^n \rightarrow 0 \) for any deterministic unit vector \( u \) orthogonal to \( \hat{Z}_n \).

Proof sketch of Theorem (6): We find the preceding arguments more quantitatively.

Let \( \hat{K}(x) = \hat{E}_n \rightarrow \hat{K}(x) = \hat{E}_n. \)

\[ \hat{K}(x) = \frac{1}{n} \sum_{i=1}^{n} x_i^T \left( I - \frac{1}{n} \mu(x) x_i x_i^T \right) x_i - 2I. \]

Thus \( \hat{K}(x) = \frac{1}{n} \sum_{i=1}^{n} x_i^T M x_i - 2I. \)

\[ \text{Then} \quad \hat{K}(x) = (\frac{1}{n} \sum_{i=1}^{n} x_i^T M x_i) - \frac{1}{n} \hat{M}. \]

Finally, 2 such that \( \lambda_2, \lambda_3 \geq \lambda_2 \), a more detailed analysis of the Stiefel transformation shows \( \text{m}(n) \rightarrow \text{m}(n) \rightarrow \frac{1}{n} \), so \( \mathbb{E}^N \rightarrow \mathbb{C}^N \).

Hence, Wigner's theorem gives \( \mathbb{C}^N_k \).

For \( z \) such that \( \text{dist} (z, \text{span}(\hat{E}_n, \ldots, \hat{E}_p)) > \delta \), a more detailed analysis of the Stiefel transformation shows \( \text{m}(n) \rightarrow \text{m}(n) \rightarrow \frac{1}{n} \), so \( \mathbb{E}^N \rightarrow \mathbb{C}^N \).

Thus \( \mathbb{C}^N_k \rightarrow \mathbb{C}^N_k \).\]

\[ \text{Note: } \frac{1}{n} \hat{K}(x) \rightarrow 0, \text{ and } \frac{1}{n} \hat{K}(x) \rightarrow 0. \text{ Then on the one hand,} \]

\[ \text{on the other hand,} \]

\[ \mathbb{E}^N \hat{K}(x) \rightarrow (\frac{1}{n} \hat{K}(x) \mathbb{E}^N) - (\frac{1}{n} \hat{K}(x) \mathbb{E}^N) \rightarrow \frac{1}{n}. \]

\[ \text{So } (\mathbb{E}^N \hat{K}(x)) \hat{K}(x) + A(\hat{K}(x)) \hat{K}(x) \rightarrow \frac{1}{n}. \]

Apply Taylor expansion: \( \hat{K}(\hat{K}(x)) \hat{K}(x) + A(\hat{K}(x)) \hat{K}(x) \rightarrow \frac{1}{n}. \)

\[ \text{From part (c), } -\hat{K}(x) \hat{K}(x) \hat{K}(x) = (\frac{1}{n} \hat{K}(x) \mathbb{E}^N) - A(\hat{K}(x)) \hat{K}(x) \rightarrow \frac{1}{n}. \]

We apply a CLT for \( A(\hat{K}(x)) \).

**Lemma:** If \( \mathbb{N}(0, I) \) entries and \( A \in \mathbb{R}^{n \times n} \) satisfies

\[ \mathbb{N}(0, I) \rightarrow \mathbb{N}(0, I), \text{ then} \]

\[ \mathbb{W} \mathbb{A} W^T A \mathbb{W} \rightarrow \mathbb{N}(0, I), \text{ as } \mathbb{W} \rightarrow \mathbb{W} \rightarrow 0. \]

**Proof:** By diagonalizing \( \mathbb{A} \) and rotational invariance, we may assume \( \mathbb{A} \) is diagonal. Then \( \mathbb{W} \mathbb{A} W^T A = \sum_{i=1}^{n} \lambda_i \mathbb{W} \mathbb{A}_i \mathbb{W}^T \).

This has variance \( \sum_{i=1}^{n} \mathbb{A}^2 \mathbb{V}_i \mathbb{W}_i = 2 \mathbb{W} \mathbb{A}^2 \mathbb{W}^T \).

Thus the result follows from the Lindeberg/Lyapunov CLT.
Note that $A_k = \lambda_k \cdot \frac{1}{\sqrt{n}} (w^T M w - T M)$, where $w$ is column $k$ of $X$.

Applying the Lemma to $A = \frac{A_k}{\sqrt{n}}$, we get

$$\sqrt{n} A_k \cdot \frac{1}{\sqrt{n}} (w^T M w - T M) \rightarrow N(0, 1).$$

Finally, $\frac{1}{n} (w^T M w - T M)^2 \rightarrow \frac{(1 + \frac{z^2}{\delta^2})}{\delta^2}$ by a similar computation as for $\frac{1}{n} Tr M$. The additional ingredient is

$$\frac{1}{n} Tr (\frac{1}{n} (\beta_0 - \beta_0^2)).$$

Thus,

$$\frac{1}{n} \int \frac{x^2}{(x - \beta_0)^2} dx = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} (1 + \frac{z^2}{\delta^2}) \cdot \sigma^2 dx$$

$$= \sigma^2 \left[ 1 + 2\sigma m(\beta_0) + 2m^2(\beta_0) \right].$$

with $m(\beta_0) = \frac{\lambda_{\beta_0}}{1 - \lambda_{\beta_0}}$ and a computation for $m(\beta_0)$ as before.

Combining these, $\sqrt{n} (\beta_0 - \beta_0^2) 
\rightarrow N(0, \sigma^2)$ where

$$\varphi^2 = 2 \cdot \lambda_1 \cdot (1 + \frac{z^2}{\delta^2}) \cdot \left( 1 - \frac{\lambda_{\beta_0}}{1 - \lambda_{\beta_0}} \right) = 2\lambda_1 \left( 1 - \frac{\lambda_{\beta_0}}{1 - \lambda_{\beta_0}} \right).$$

Problem 7. (a) Implement a method that estimates principal eigenvalues of $\Sigma$ in the spiked covariance model, based on observing $X$, and also provides 95% confidence intervals in the setting of real Gaussian data.

Note that $m(\beta)$ may be computed from its analytic expression, or estimated as $\frac{1}{n} \sum \frac{1}{\sqrt{n}}$ when the sum is over the bulk eigenvalues $\lambda_i$ of $\Sigma$.

Are these the same? Which leads to more accurate estimates of confidence intervals in practice?