



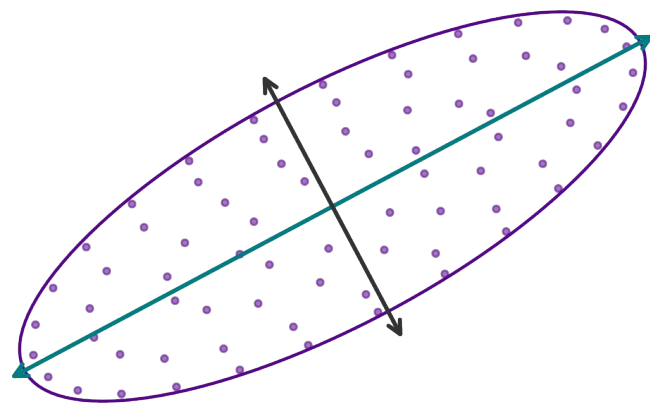
Eigenvalue Exclusion for Sample Covariance Matrices with Dependent Entries

Alex Hassett

Faculty Mentor: Dr. Simona Diaconu

SURE 2026

Covariance and eigenvalues



Centered observations

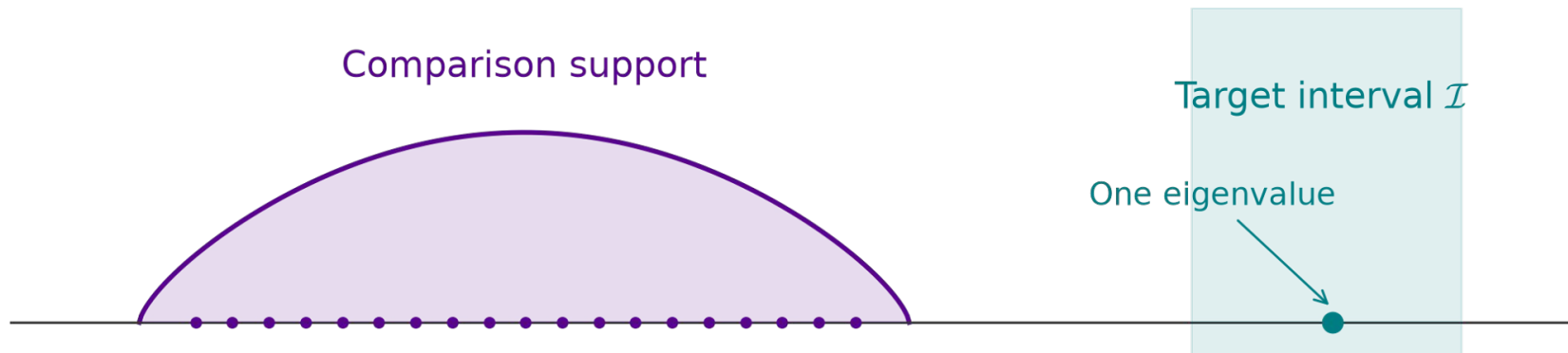
Eigenvectors: principal directions

Eigenvalues: variance in those directions

$$T = \mathbb{E}[yy^T] \quad (\mathbb{E}[y] = 0)$$

A global limit can miss an eigenvalue

$$\nu_n = \frac{1}{p_n} \sum_{k=1}^{p_n} \delta_{\lambda_k(B_n)}$$

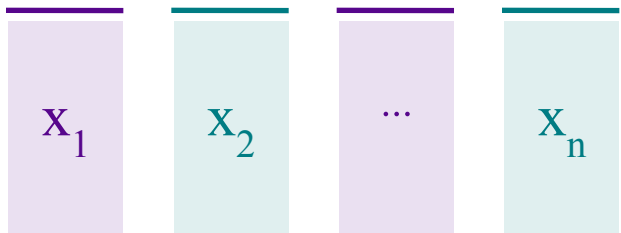


One eigenvalue has mass $1/p$.
That mass tends to zero as p grows.

The sample covariance model

$$B_n = \frac{1}{n} T_n^{1/2} X_n X_n^T T_n^{1/2}, \quad X_n = (x_1, \dots, x_n)$$

p coordinates; n samples; deterministic covariance T_n



Columns are independent.
Their coordinates need not be.

Within each column:

$$\mathbb{E}[X_{ij} \mid \text{past coordinates}] = 0$$

$$\mathbb{E}[X_{ij}^2 \mid \text{past coordinates}] = 1$$

Main result: eventual exclusion

p and n grow at comparable rates.

The largest eigenvalue of T_n stays bounded.

2626th absolute moments have a uniform bound.

A neighborhood of I is outside the comparison support.

The defining equation is stable there.

Fixed compact $I \subset (0, \infty)$.

$$\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) \leq Cn^{-379/272}$$

**Almost surely, I contains no eigenvalues
for all sufficiently large n .**

Quadratic-form concentration

x is a column from our model. A is a fixed $p \times p$ matrix.

$$\mathbb{E} \left[|x^T A x - \text{Tr } A|^q \right] \leq C_q \|A\|_F^q, \quad 2 \leq q \leq 1313$$

$$\|A\|_F^2 = \sum_{i,k} |a_{ik}|^2$$

For real symmetric A , the proof starts with

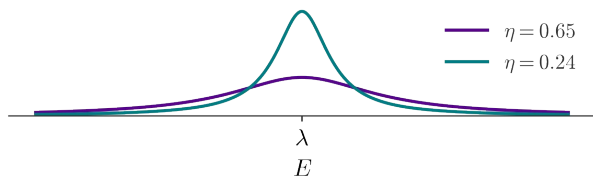
$$x^T A x - \text{Tr } A = \sum_{i=1}^p a_{ii} (X_i^2 - 1) + 2 \sum_{i=1}^p X_i \sum_{k < i} a_{ik} X_k$$

A transform that detects eigenvalues

$$m_n(z) = \frac{1}{n} \text{Tr}(B_n - zI)^{-1} = \frac{1}{n} \sum_{k=1}^{p_n} \frac{1}{\lambda_k - z}$$

$$\text{Im } m_n(E + i\eta) = \frac{1}{n} \sum_{k=1}^{p_n} \frac{\eta}{(\lambda_k - E)^2 + \eta^2}, \quad \eta > 0$$

Each summand is largest when E equals its eigenvalue.



η sets the width.

Smaller η gives finer detail.

$$m_{c,n}(z) = \frac{p_n}{n} \int \frac{1}{\lambda - z} dF_{c,n}(\lambda)$$

$F_{c,n}$: comparison
probability distribution

The role of stability

$$u_n(z) = m_n(z) - \frac{1 - p_n/n}{z}, \quad u_{c,n}(z) = m_{c,n}(z) - \frac{1 - p_n/n}{z}$$

$$f_n(v) = -\frac{1}{v} + \frac{1}{n} \sum_{i=1}^{p_n} \frac{t_i}{1 + t_i v}, \quad f_n(u_{c,n}(z)) = z$$

$$t_i = \lambda_i(T_n), \quad \text{Im } u_{c,n}(z) > 0$$

1 Room around the target interval

A neighborhood of I outside
the comparison support.

2 Denominators away from zero

$$|u_{c,n}(z)| \geq c, \quad \min_i |1 + t_i u_{c,n}(z)| \geq c$$

3 Derivative away from zero

$$|f'_n(u_{c,n}(z))| \geq c > 0$$

The main steps of the proof

1 Delete a column

The omitted sample is independent of the remaining matrix.

2 Apply the quadratic-form bound

Condition on the remaining columns to estimate the resolvent terms.

3 Compare the transforms

Use stability to bound the difference between the two transforms.

4 Show that I contains no eigenvalues

Sum nonnegative terms to detect an eigenvalue in I .

Excluding eigenvalues with dependent coordinates

Independent columns and martingale-difference coordinates

Moment bounds, support separation, and stability

Thank you