Optimality and Sub-optimality of Principal Component Analysis for Spiked Random Matrices

Amelia Perry
MIT

Joint work with:
Alex Wein (MIT), Afonso Bandeira (Courant/CDS), Ankur Moitra (MIT)

Random Matrices

Wigner Matrix

$$\frac{1}{\sqrt{n}}W \in \mathbb{R}^{n\times n} \text{ symmetric,}$$
$$W_{ij} \sim \mathcal{N}(0,1)$$

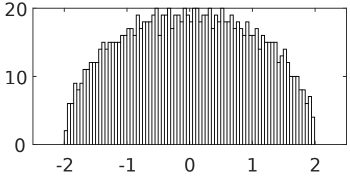
E. P. Wigner, AoM 1958.

V. A. Marchenko, L. A. Pastur, M.S 1967.

Random Matrices

Wigner Matrix

$$\frac{1}{\sqrt{n}} W \in \mathbb{R}^{n \times n} \text{ symmetric,}$$
$$W_{ij} \sim \mathcal{N}(0, 1)$$



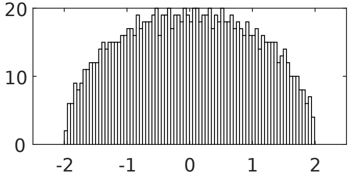
E. P. Wigner, AoM 1958.

V. A. Marchenko, L. A. Pastur, M.S 1967.

Random Matrices

Wigner Matrix

$\frac{1}{\sqrt{n}} W \in \mathbb{R}^{n \times n}$ symmetric,
$W_{ij} \sim \mathcal{N}(0, 1)$



Wishart Matrix

$\frac{1}{N} \sum_{k=1}^{N} y_k y_k^T$
$y_k \sim \mathcal{N}(0, I_n)$

E. P. Wigner, AoM 1958.

V. A. Marchenko, L. A. Pastur, M.S 1967.

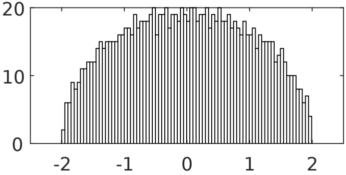
Random Matrices

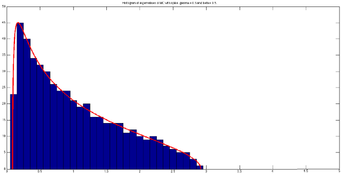
Wigner Matrix

$\frac{1}{\sqrt{n}} W \in \mathbb{R}^{n \times n}$ symmetric,
$W_{ij} \sim \mathcal{N}(0, 1)$

Wishart Matrix

$\frac{1}{N} \sum_{k=1}^{N} y_k y_k^T$
$y_k \sim \mathcal{N}(0, I_n)$





E. P. Wigner, AoM 1958.

V. A. Marchenko, L. A. Pastur, M.S 1967.

Spiked Models

Spiked Wigner Matrix

$$Y = \frac{1}{\sqrt{n}}W + \lambda x x^T$$

$\frac{1}{\sqrt{n}}W$ Wigner, $\|x\| = 1$.

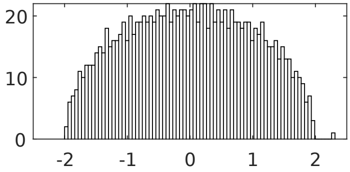
J. Baik, G. Ben Arous, S. Péché, AoP 2005.
D. Feral, S. Péché, CMP 2006.

Spiked Models

Spiked Wigner Matrix

$$Y = \frac{1}{\sqrt{n}} W + \lambda x x^T$$

$\frac{1}{\sqrt{n}} W$ Wigner, $\|x\| = 1$.



Visible on the largest eigenvalue when

$$\lambda > 1$$

J. Baik, G. Ben Arous, S. Péché, AoP 2005.

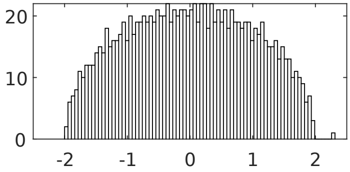
D. Feral, S. Péché, CMP 2006.

Spiked Models

Spiked Wigner Matrix

$$Y = \frac{1}{\sqrt{n}} W + \lambda x x^T$$

$\frac{1}{\sqrt{n}} W$ Wigner, $\|x\| = 1$.



Visible on the largest eigenvalue when

$$\lambda > 1$$

J. Baik, G. Ben Arous, S. Péché, AoP 2005.

D. Feral, S. Péché, CMP 2006.

Wishart Matrix

$$Y = \frac{1}{N} \sum_{k=1}^{N} y_k y_k^T$$

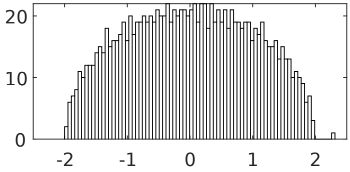
$$y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right), \ \|x\| = 1$$

Spiked Models

Spiked Wigner Matrix

$$Y = \frac{1}{\sqrt{n}} W + \lambda x x^T$$

$$\frac{1}{\sqrt{n}} W \text{ Wigner, } \|x\| = 1.$$



Visible on the largest eigenvalue when

$$\lambda > 1$$

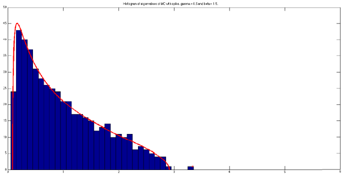
J. Baik, G. Ben Arous, S. Péché, AoP 2005.

D. Feral, S. Péché, CMP 2006.

Wishart Matrix

$$Y = \frac{1}{N} \sum_{k=1}^{N} y_k y_k^T$$

$$y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right), \ \|x\| = 1$$



Visible on the largest eigenvalue when

$$|\beta| > \sqrt{\gamma}, \ \gamma = \frac{n}{N}, \ \beta \in [-1, \infty)$$

Statistical Questions

- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

Statistical Questions

- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

- Recovery: nontrivial correlation with truth: $\langle x, \hat{x} \rangle > \epsilon$

Statistical Questions

- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

- Recovery: nontrivial correlation with truth: $\langle x, \hat{x} \rangle > \epsilon$
- PCA solves both detection and recovery above threshold $\lambda > 1$ (Wigner)

Statistical Questions

- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

- Recovery: nontrivial correlation with truth: $\langle x, \hat{x} \rangle > \epsilon$
- PCA solves both detection and recovery above threshold $\lambda > 1$ (Wigner)
- Are they statistically possible below the threshold?

Statistical Questions

- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

- Recovery: nontrivial correlation with truth: $\langle x, \hat{x} \rangle > \epsilon$
- PCA solves both detection and recovery above threshold $\lambda > 1$ (Wigner)
- Are they statistically possible below the threshold?
- Need a prior on the spike $x \in \mathbb{R}^n$

Statistical Questions

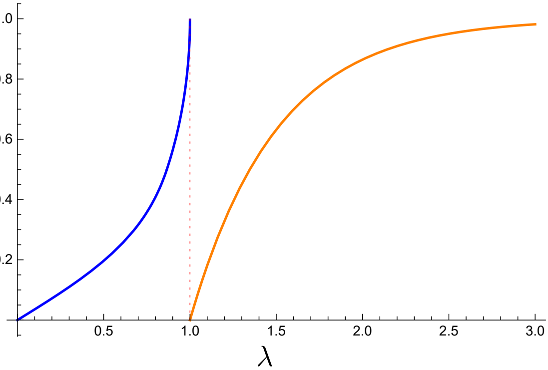
- Detection: distinguish reliably (error prob $\to 0$)

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T$$

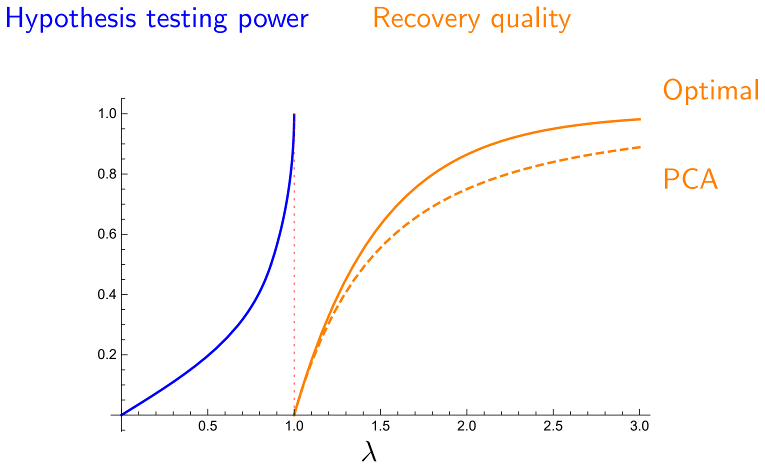
- Recovery: nontrivial correlation with truth: $\langle x, \hat{x} \rangle > \epsilon$
- PCA solves both detection and recovery above threshold $\lambda > 1$ (Wigner)
- Are they statistically possible below the threshold?
- Need a prior on the spike $x \in \mathbb{R}^n$
  - unit sphere
  - i.i.d. ± 1
  - sparse ± 1

Detection vs Recovery

Hypothesis testing power   Recovery quality

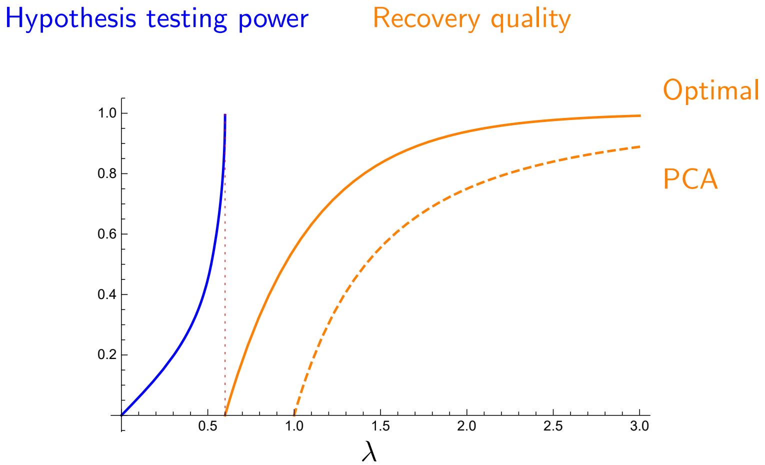


Detection vs Recovery



“PCA is optimal”

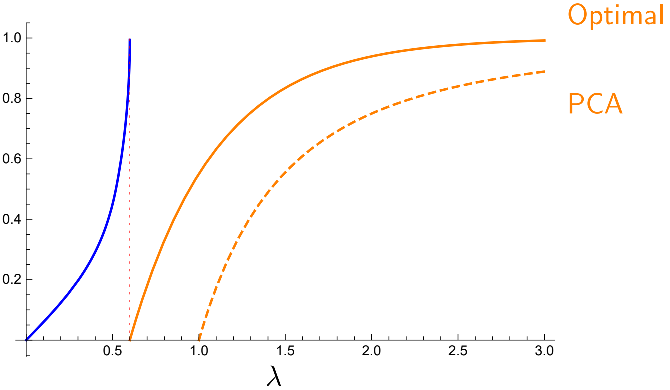
Detection vs Recovery



"PCA is sub-optimal"

Detection vs Recovery

Hypothesis testing power   Recovery quality



"PCA is sub-optimal"

This talk: focus on detection threshold
(also hypothesis testing bounds, recovery threshold)

Our Results: 3 Scenarios

1. PCA achieves optimal threshold
   (e.g. Wigner with spherical or Rademacher prior)

Our Results: 3 Scenarios

1. PCA achieves optimal threshold
   (e.g. Wigner with spherical or Rademacher prior)

2. Can beat PCA with an efficient algorithm
   (e.g. non-Gaussian Wigner)

Our Results: 3 Scenarios

1. PCA achieves optimal threshold
   (e.g. Wigner with spherical or Rademacher prior)

2. Can beat PCA with an efficient algorithm
   (e.g. non-Gaussian Wigner)

3. Can beat PCA, but only with an inefficient algorithm
   (e.g. sparse priors, Wishart)

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n \to \infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n \to \infty} P_n(A_n) = 0.$$

- If distributions are contiguous, there is no reliable test to distinguish them.

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

- If distributions are contiguous, there is no reliable test to distinguish them.
  Proof: Let A_n be the event that distinguisher says 'P_n'. □

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

- If distributions are contiguous, there is no reliable test to distinguish them.
  Proof: Let A_n be the event that distinguisher says 'P_n'. $\square$

Theorem (Second Moment)

If $\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 = O(1)$, *then P_n is contiguous to Q_n.*

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

- If distributions are contiguous, there is no reliable test to distinguish them.
  Proof: Let A_n be the event that distinguisher says 'P_n'. □

Theorem (Second Moment)

If $\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 = O(1)$, *then* P_n *is contiguous to* Q_n.

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

Theorem (Second Moment)

If $\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 = O(1)$, *then* P_n *is contiguous to* Q_n.

Proof:

L. Le Cam, 1960.

Lower Bound Concept: Contiguity

- Sequence of distributions P_n is contiguous to Q_n if for any sequence of events A_n,

$$\lim_{n\to\infty} Q_n(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n(A_n) = 0.$$

Theorem (Second Moment)
If $\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 = O(1)$, *then* P_n *is contiguous to* Q_n.

Proof:

$$\underbrace{\int_{A_n} dP_n}_{P_n(A_n)} = \int_{A_n} \frac{dP_n}{dQ_n} dQ_n \le \Bigg(\underbrace{\int_{A_n}\left(\frac{dP_n}{dQ_n}\right)^2 dQ_n}_{\le \mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2}\Bigg)^{\frac{1}{2}} \Bigg(\underbrace{\int_{A_n} dQ_n}_{Q_n(A_n)}\Bigg)^{\frac{1}{2}}$$

L. Le Cam, 1960.

Case 1: PCA is optimal

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

- E.g. if the prior $\mathcal{X}$ is uniform on the unit sphere (and $\lambda < 1$):

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp \left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

- E.g. if the prior $\mathcal{X}$ is uniform on the unit sphere (and $\lambda < 1$):

$$\lim_{n \to \infty} \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp \left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right) = \left(1 - \lambda^2\right)^{-\frac{1}{2}} < \infty$$

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp \left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

- E.g. if the prior $\mathcal{X}$ is uniform on the unit sphere (and $\lambda < 1$):

$$\lim_{n \to \infty} \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp \left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right) = \left(1 - \lambda^2\right)^{-\frac{1}{2}} < \infty$$

The distributions are contiguous below the spectral threshold [MRZ15]

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

- E.g. if the prior $\mathcal{X}$ is uniform on the unit sphere (and $\lambda < 1$):

$$\lim_{n \to \infty} \mathbb{E}_{x,x' \sim \mathcal{X}} \ \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right) = \left(1 - \lambda^2\right)^{-\frac{1}{2}} < \infty$$

The distributions are contiguous below the spectral threshold [MRZ15]

- Same can be shown for the Wishart case [OMH13]

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner Second Moment

- Taking $P_n : \frac{1}{\sqrt{n}} W + \lambda x x^T$ with $x \sim \mathcal{X}$, and $Q_n : \frac{1}{\sqrt{n}} W$

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 = \mathbb{E}_{x,x' \sim \mathcal{X}} \, \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right)$$

- E.g. if the prior $\mathcal{X}$ is uniform on the unit sphere (and $\lambda < 1$):

$$\lim_{n \to \infty} \mathbb{E}_{x,x' \sim \mathcal{X}} \, \exp\left(\frac{\lambda^2 n}{2} \langle x, x' \rangle^2 \right) = \left(1 - \lambda^2\right)^{-\frac{1}{2}} < \infty$$

  The distributions are contiguous below the spectral threshold [MRZ15]

- Same can be shown for the Wishart case [OMH13]

But what about when we know more about the spike?

[MRZ15] A. Montanari, D. Reichman, O. Zeitouni, NIPS 2015.

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Wigner, Rademacher Prior

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T,$$

$$x \sim \mathrm{Unif}\left\{\pm \frac{1}{\sqrt{n}}\right\}^n.$$

Related: J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner, Rademacher Prior

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T,$$

$$x \sim \mathrm{Unif}\left\{ \pm \frac{1}{\sqrt{n}} \right\}^n.$$

Still Contiguous for $\lambda < 1$

Related: J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner, Rademacher Prior

$$Y \sim \frac{1}{\sqrt{n}} W \text{ (Wigner)} \quad \text{vs} \quad Y \sim \frac{1}{\sqrt{n}} W + \lambda x x^T,$$

$$x \sim \mathrm{Unif}\left\{ \pm \frac{1}{\sqrt{n}} \right\}^n.$$

Still Contiguous for $\lambda < 1$

- Contiguity argument goes through for a general class of priors!

Related: J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Proof Details: Bounding the (Wigner) Second Moment

$$\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 \quad = \quad \mathbb{E}\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X}$$

Proof Details: Bounding the (Wigner) Second Moment

$$\begin{aligned}\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 &= \mathbb{E}\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X} \\ &= \int_0^\infty \Pr\left[\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \geq t\right] dt\end{aligned}$$

Proof Details: Bounding the (Wigner) Second Moment

$$
\begin{aligned}
\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 &= \mathbb{E}\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X} \\
&= \int_0^\infty \Pr\left[\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \geq t\right] dt \\
&= 2\int_0^\infty \Pr\left[\langle x, x'\rangle \geq \sqrt{\frac{2\log t}{\lambda^2 n}}\right] dt
\end{aligned}
$$

Proof Details: Bounding the (Wigner) Second Moment

$$
\begin{aligned}
\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 &= \mathbb{E}\,\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X} \\
&= \int_0^\infty \Pr\left[\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \geq t\right] dt \\
&= 2\int_0^\infty \Pr\left[\langle x, x'\rangle \geq \sqrt{\frac{2\log t}{\lambda^2 n}}\right] dt \\
&= 2\int_0^1 \Pr\left[\langle x, x'\rangle \geq u\right] \exp\left(\frac{\lambda^2 n}{2}u^2\right) \lambda^2 n\, u\, du
\end{aligned}
$$

Proof Details: Bounding the (Wigner) Second Moment

$$
\begin{aligned}
\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 &= \mathbb{E}\,\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X} \\
&= \int_0^\infty \Pr\left[\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \geq t\right] dt \\
&= 2\int_0^\infty \Pr\left[\langle x, x'\rangle \geq \sqrt{\frac{2\log t}{\lambda^2 n}}\right] dt \\
&= 2\int_0^1 \Pr\left[\langle x, x'\rangle \geq u\right] \exp\left(\frac{\lambda^2 n}{2} u^2\right) \lambda^2 n\, u\, du \\
&\to 2\int_0^1 \lambda^2 n\, u \exp\left[n\left(\frac{\lambda^2}{2} u^2 - R(u)\right)\right] du
\end{aligned}
$$

Rate function: $R(u) = \lim_{n\to\infty} \frac{-1}{n} \log \Pr\left[\langle x, x'\rangle \geq u\right]$

Proof Details: Bounding the (Wigner) Second Moment

$$\begin{aligned}
\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 &= \mathbb{E}\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \qquad x, x' \sim \mathcal{X} \\
&= \int_0^\infty \Pr\left[\exp\left(\frac{\lambda^2 n}{2}\langle x, x'\rangle^2\right) \geq t\right] dt \\
&= 2\int_0^\infty \Pr\left[\langle x, x'\rangle \geq \sqrt{\frac{2\log t}{\lambda^2 n}}\right] dt \\
&= 2\int_0^1 \Pr\left[\langle x, x'\rangle \geq u\right] \exp\left(\frac{\lambda^2 n}{2}u^2\right) \lambda^2 n\, u\, du \\
&\to 2\int_0^1 \lambda^2 n\, u \exp\left[n\left(\frac{\lambda^2}{2}u^2 - R(u)\right)\right] du
\end{aligned}$$

Rate function: $R(u) = \lim_{n\to\infty} \frac{-1}{n} \log \Pr\left[\langle x, x'\rangle \geq u\right]$

In other words: $\Pr\left[\langle x, x'\rangle \geq u\right] \approx \exp(-n\, R(u))$

Proof Details: Bounding the (Wigner) Second Moment

$$\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 \quad \to \quad 2\int_0^1 \lambda^2 n\, u \exp\left[n\left(\frac{\lambda^2}{2}u^2 - R(u)\right)\right] du$$

Proof Details: Bounding the (Wigner) Second Moment

$$\mathbb{E}_{Q_n} \left(\frac{dP_n}{dQ_n} \right)^2 \quad \to \quad 2 \int_0^1 \lambda^2 n\, u \exp \left[n \left(\frac{\lambda^2}{2} u^2 - R(u) \right) \right] du$$

- Bounded iff $\frac{\lambda^2}{2} u^2 - R(u) < 0$ for all $u \in (0, 1)$

Proof Details: Bounding the (Wigner) Second Moment

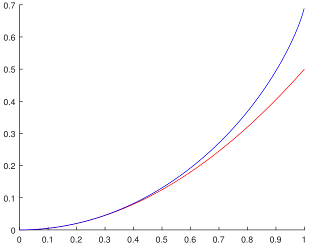
$$\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 \quad \to \quad 2\int_0^1 \lambda^2 n\, u \exp\left[n\left(\frac{\lambda^2}{2}u^2 - R(u)\right)\right] du$$

- Bounded iff $\frac{\lambda^2}{2}u^2 - R(u) < 0$ for all $u \in (0,1)$
- How big a parabola can you fit underneath the rate function?

Proof Details: Bounding the (Wigner) Second Moment

$$\mathbb{E}_{Q_n}\left(\frac{dP_n}{dQ_n}\right)^2 \quad \to \quad 2\int_0^1 \lambda^2 n\, u \exp\left[n\left(\frac{\lambda^2}{2}u^2 - R(u)\right)\right] du$$

- Bounded iff $\frac{\lambda^2}{2}u^2 - R(u) < 0$ for all $u \in (0,1)$
- How big a parabola can you fit underneath the rate function?
- E.g. Rademacher prior (± 1) has $R(u) = \log 2 - h\left(\frac{1+u}{2}\right)$ where $h(p) = -p\log p - (1-p)\log(1-p)$



$R(u)$
$\frac{1}{2}u^2$ $\quad (\lambda = 1)$

Case 2: PCA can be efficiently beaten

What if noise is not Gaussian?

$$Y = \frac{1}{\sqrt{n}} W + \lambda x x^T$$

$x \sim \mathrm{Unif}\{\mathbb{S}^{n-1}\}$, $W \in \mathbb{R}^{n \times n}$

but $W_{ij} \sim p(w)$ such that $\mathbb{E}w = 0$, $\mathbb{E}w^2 = 1$.

D. Feral, S. Péché, CMP 2006.

T. Tao, V. Vu, MAoRMT 2012.

What if noise is not Gaussian?

$$Y = \frac{1}{\sqrt{n}} W + \lambda x x^T$$

$x \sim \mathrm{Unif}\{\mathbb{S}^{n-1}\}$, $W \in \mathbb{R}^{n \times n}$

but $W_{ij} \sim p(w)$ such that $\mathbb{E}w = 0$, $\mathbb{E}w^2 = 1$.

Universality: spectral properties are unchanged...

D. Feral, S. Péché, CMP 2006.

T. Tao, V. Vu, MAoRMT 2012.

Can you tell which one is which?

For $W_{ij} \sim \mathrm{Unif}(\pm 1)$, $\lambda < 1$

$$W + \lambda\sqrt{n}xx^T \quad \text{vs} \quad W$$

Can you tell which one is which?

For $W_{ij} \sim \text{Unif}(\pm 1)$, $\lambda < 1$

$$W + \lambda\sqrt{n}xx^T \quad \text{vs} \quad W$$

-1.0000	1.0000	-1.0000	-1.0000	1.0000	-1.0000
1.0000	1.0000	1.0000	-1.0000	-1.0000	1.0000
-1.0000	1.0000	1.0000	-1.0000	-1.0000	-1.0000
-1.0000	-1.0000	-1.0000	1.0000	-1.0000	1.0000
1.0000	-1.0000	-1.0000	-1.0000	-1.0000	1.0000
-1.0000	1.0000	-1.0000	1.0000	1.0000	1.0000

vs

-0.9988	1.0011	-1.0007	-0.9997	0.9990	-1.0014
1.0011	1.0010	0.9993	-0.9997	-1.0010	0.9987
-1.0007	0.9993	1.0004	-1.0002	-0.9994	-0.9991
-0.9997	-0.9997	-1.0002	1.0001	-1.0002	0.9997
0.9990	-1.0010	-0.9994	-1.0002	-0.9991	1.0012
-1.0014	0.9987	-0.9991	0.9997	1.0012	1.0017

Can you tell which one is which?

For $W_{ij} \sim \mathrm{Unif}(\pm 1)$, $\lambda < 1$

$$W + \lambda\sqrt{n}xx^T \quad \text{vs} \quad W$$

-1.0000	1.0000	-1.0000	-1.0000	1.0000	-1.0000
1.0000	1.0000	1.0000	-1.0000	-1.0000	1.0000
-1.0000	1.0000	1.0000	-1.0000	-1.0000	-1.0000
-1.0000	-1.0000	-1.0000	1.0000	-1.0000	1.0000
1.0000	-1.0000	-1.0000	-1.0000	-1.0000	1.0000
-1.0000	1.0000	-1.0000	1.0000	1.0000	1.0000

vs

-0.9988	1.0011	-1.0007	-0.9997	0.9990	-1.0014
1.0011	1.0010	0.9993	-0.9997	-1.0010	0.9987
-1.0007	0.9993	1.0004	-1.0002	-0.9994	-0.9991
-0.9997	-0.9997	-1.0002	1.0001	-1.0002	0.9997
0.9990	-1.0010	-0.9994	-1.0002	-0.9991	1.0012
-1.0014	0.9987	-0.9991	0.9997	1.0012	1.0017

Let's restrict ourselves to when the density $p(w)$ is smooth.

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$f(Y_{ij}) \quad = \quad f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right)$$

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$\begin{aligned} f(Y_{ij}) &= f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right) \\ &\approx f\left(W_{ij}\right) + f'\left(W_{ij}\right)\lambda\sqrt{n}x_i x_j \end{aligned}$$

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$\begin{aligned}
f(Y_{ij}) &= f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right) \\
&\approx f\left(W_{ij}\right) + f'\left(W_{ij}\right)\lambda\sqrt{n}x_i x_j \\
&\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j - \left(f'\left(W_{ij}\right) - \mathbb{E}f'\left(W_{ij}\right)\right)\lambda\sqrt{n}x_i x_j
\end{aligned}$$

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$
\begin{aligned}
f(Y_{ij}) &= f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right) \\
&\approx f\left(W_{ij}\right) + f'\left(W_{ij}\right)\lambda\sqrt{n}x_i x_j \\
&\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j - \left(f'\left(W_{ij}\right) - \mathbb{E}f'\left(W_{ij}\right)\right)\lambda\sqrt{n}x_i x_j \\
&\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j .
\end{aligned}
$$

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$
\begin{aligned}
f(Y_{ij}) &= f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right) \\
&\approx f\left(W_{ij}\right) + f'\left(W_{ij}\right)\lambda\sqrt{n}x_i x_j \\
&\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j - \left(f'\left(W_{ij}\right) - \mathbb{E}f'\left(W_{ij}\right)\right)\lambda\sqrt{n}x_i x_j \\
&\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j.
\end{aligned}
$$

- It is (close to) a new spiked Wigner matrix with

$$
\lambda' = \frac{\lambda \mathbb{E} f'\left(W_{ij}\right)}{\sqrt{\mathbb{E} f^2\left(W_{ij}\right)}}.
$$

Entrywise function

If noise drawn from non-Gaussian $p(w)$: we will beat PCA by applying some function $f : \mathbb{R} \to \mathbb{R}$ entrywise to our matrix $Y = W + \lambda\sqrt{n}xx^\top$, followed by PCA.

$$\begin{aligned} f(Y_{ij}) &= f\left(W_{ij} + \lambda\sqrt{n}x_i x_j\right) \\ &\approx f\left(W_{ij}\right) + f'\left(W_{ij}\right)\lambda\sqrt{n}x_i x_j \\ &\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j - \left(f'\left(W_{ij}\right) - \mathbb{E}f'\left(W_{ij}\right)\right)\lambda\sqrt{n}x_i x_j \\ &\approx f\left(W_{ij}\right) + \mathbb{E}[f'\left(W_{ij}\right)]\lambda\sqrt{n}x_i x_j. \end{aligned}$$

- It is (close to) a new spiked Wigner matrix with

$$\lambda' = \frac{\lambda\mathbb{E}f'\left(W_{ij}\right)}{\sqrt{\mathbb{E}f^2\left(W_{ij}\right)}}.$$

- Calculus of variations gives optimal choice of f:

$$f(w) = \frac{-p'(w)}{p(w)}$$

Pre-transformed PCA

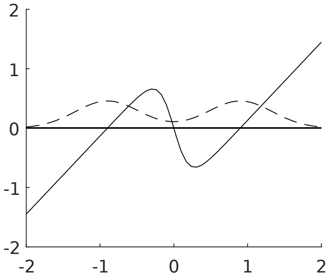
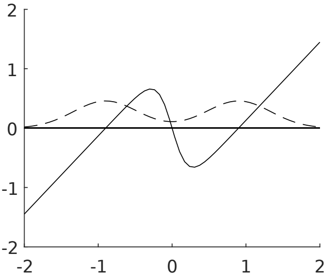


Figure: Dashed: $p(w)$,
Solid: $f(w) = -p'(w)/p(w)$

Related: T. Lesieur, F. Krzakala, L. Zdeborová, Allerton 2015;
F. Krzakala, J. Xu, L. Zdeborová, 2016.

Pre-transformed PCA



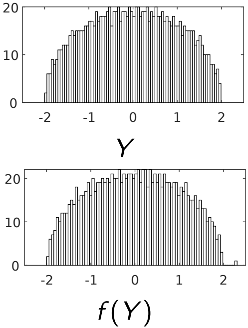
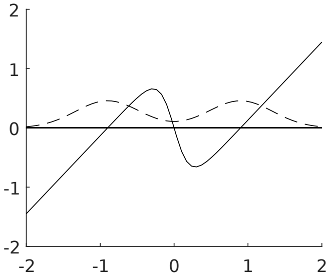


Figure: Dashed: $p(w)$,
Solid: $f(w) = -p'(w)/p(w)$

Related: T. Lesieur, F. Krzakala, L. Zdeborová, Allerton 2015;
F. Krzakala, J. Xu, L. Zdeborová, 2016.

Pre-transformed PCA



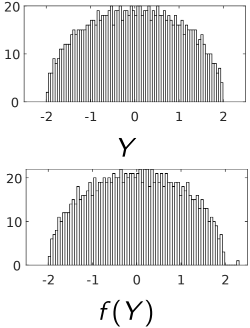
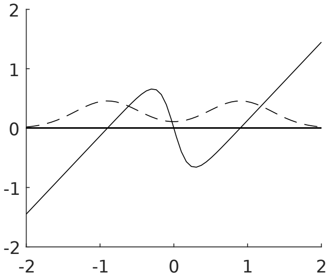


Figure: Dashed: $p(w)$,
Solid: $f(w) = -p'(w)/p(w)$

- New threshold at $\lambda = \frac{1}{\sqrt{F_p}}, \quad F_p = \mathbb{E}_p \left(\frac{p'(w)}{p(w)}\right)^2 \geq 1.$

Related: T. Lesieur, F. Krzakala, L. Zdeborová, Allerton 2015;
F. Krzakala, J. Xu, L. Zdeborová, 2016.

Pre-transformed PCA



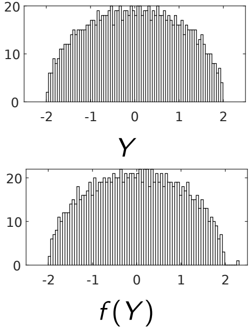


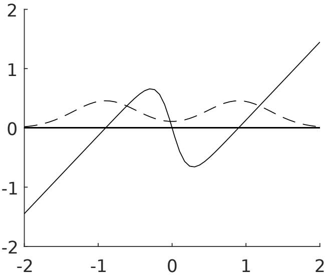
Figure: Dashed: $p(w)$,
Solid: $f(w) = -p'(w)/p(w)$

- New threshold at $\lambda = \frac{1}{\sqrt{F_p}}$, $\quad F_p = \mathbb{E}_p \left(\frac{p'(w)}{p(w)}\right)^2 \geq 1.$

Gaussian is hardest: $F_p = 1.$

Related: T. Lesieur, F. Krzakala, L. Zdeborová, Allerton 2015;
F. Krzakala, J. Xu, L. Zdeborová, 2016.

Pre-transformed PCA



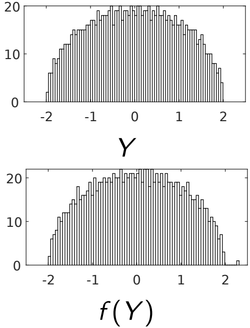


Figure: Dashed: $p(w)$,
Solid: $f(w) = -p'(w)/p(w)$

- New threshold at $\lambda = \frac{1}{\sqrt{F_p}}$, $\quad F_p = \mathbb{E}_p \left(\frac{p'(w)}{p(w)} \right)^2 \geq 1.$

  Gaussian is hardest: $F_p = 1$.

- Theorem: detection is impossible below this threshold!

Related: T. Lesieur, F. Krzakala, L. Zdeborová, Allerton 2015;
F. Krzakala, J. Xu, L. Zdeborová, 2016.

Case 3: PCA can be inefficiently beaten

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$y_k \sim \mathcal{N}(0, I_n)$ vs $y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right)$, $x \sim \mathrm{Unif}\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$$y_k \sim \mathcal{N}(0, I_n) \qquad \text{vs} \qquad y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right),\ x \sim \mathrm{Unif}\left\{\pm\frac{1}{\sqrt{n}}\right\}^n$$

PCA is optimal for spherical prior: contiguous for $|\beta| < \sqrt{\gamma}$ [OMH13].

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$y_k \sim \mathcal{N}(0, I_n)$ vs $y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right)$, $x \sim \mathrm{Unif}\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$

PCA is optimal for spherical prior: contiguous for $|\beta| < \sqrt{\gamma}$ [OMH13].

For Rademacher prior, something surprising happens:

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$y_k \sim \mathcal{N}(0, I_n)$ vs $y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right)$, $x \sim \mathrm{Unif}\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$

PCA is optimal for spherical prior: contiguous for $|\beta| < \sqrt{\gamma}$ [OMH13].

For Rademacher prior, something surprising happens:

- If $\frac{n}{N} = \gamma < \frac{1}{3}$ then the models are contiguous for $|\beta| < \sqrt{\gamma}$

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$$y_k \sim \mathcal{N}(0, I_n) \qquad \text{vs} \qquad y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right),\ x \sim \mathrm{Unif}\left\{\pm\frac{1}{\sqrt{n}}\right\}^n$$

PCA is optimal for spherical prior: contiguous for $|\beta| < \sqrt{\gamma}$ [OMH13].

For Rademacher prior, something surprising happens:

- If $\frac{n}{N} = \gamma < \frac{1}{3}$ then the models are contiguous for $|\beta| < \sqrt{\gamma}$
- But... for $\gamma > 0.698$ a computationally inefficient procedure distinguishes the two models for some $\beta \in (-\sqrt{\gamma}, 0)$ (below the spectral threshold).

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Spiked Wishart model, Rademacher prior

$$\frac{1}{N}\sum_{k=1}^{N} y_k y_k^T$$

$$y_k \sim \mathcal{N}(0, I_n) \qquad \text{vs} \qquad y_k \sim \mathcal{N}\left(0, I_n + \beta x x^T\right),\ x \sim \mathrm{Unif}\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$$

PCA is optimal for spherical prior: contiguous for $|\beta| < \sqrt{\gamma}$ [OMH13].

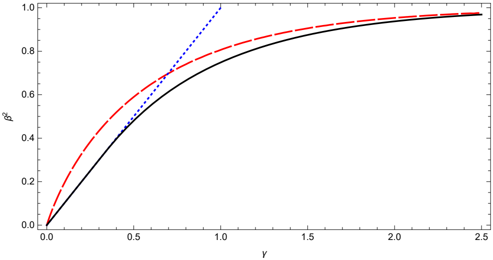
For Rademacher prior, something surprising happens:

- If $\frac{n}{N} = \gamma < \frac{1}{3}$ then the models are contiguous for $|\beta| < \sqrt{\gamma}$
- But. . . for $\gamma > 0.698$ a computationally inefficient procedure distinguishes the two models for some $\beta \in (-\sqrt{\gamma}, 0)$ (below the spectral threshold).

Is there a computational gap?

[OMH13] A. Onatski, M. J. Moreira, M. Hallin, AoS 2013.

Rademacher Spiked Wishart, Negative β



- PCA: succeeds above the line
- inefficient algorithm: succeeds above the line
- contiguity lower bound: impossible below the line

Improvement: Conditioning

- Goal: P_n contiguous to Q_n: $Q_n(A_n) \to 0 \;\Rightarrow\; P_n(A_n) \to 0$

J. Banks, C. Moore, J. Neeman, P. Netrapalli, COLT 2016.

Improvement: Conditioning

- Goal: P_n contiguous to Q_n: $Q_n(A_n) \to 0 \;\Rightarrow\; P_n(A_n) \to 0$

- Sufficient to show $\tilde{P}_n$ is contiguous to Q_n, where $\tilde{P}_n, P_n$ agree with prob $1 - o(1)$

J. Banks, C. Moore, J. Neeman, P. Netrapalli, COLT 2016.

Improvement: Conditioning

- Goal: P_n contiguous to Q_n: $Q_n(A_n) \to 0 \;\Rightarrow\; P_n(A_n) \to 0$

- Sufficient to show $\tilde{P}_n$ is contiguous to Q_n, where $\tilde{P}_n, P_n$ agree with prob $1 - o(1)$

- Choose $\tilde{P}_n$ to make $\mathbb{E}_{Q_n} \left(\frac{d\tilde{P}_n}{dQ_n} \right)^2$ small

J. Banks, C. Moore, J. Neeman, P. Netrapalli, COLT 2016.

Improvement: Conditioning

- Goal: P_n contiguous to Q_n: $Q_n(A_n) \to 0 \;\Rightarrow\; P_n(A_n) \to 0$

- Sufficient to show $\tilde{P}_n$ is contiguous to Q_n, where $\tilde{P}_n, P_n$ agree with prob $1 - o(1)$

- Choose $\tilde{P}_n$ to make $\mathbb{E}_{Q_n} \left(\frac{d\tilde{P}_n}{dQ_n} \right)^2$ small

- Condition away from rare bad events

J. Banks, C. Moore, J. Neeman, P. Netrapalli, COLT 2016.

Wigner with Sparse Rademacher Prior

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner with Sparse Rademacher Prior

- Sparsity $\rho \in [0, 1]$

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner with Sparse Rademacher Prior

- Sparsity $\rho \in [0, 1]$
- Prior $\mathcal{X}_\rho$: i.i.d. $x_i \sim \frac{1}{\sqrt{\rho n}} \begin{cases} +1 & \text{w.p. } \rho/2 \\ -1 & \text{w.p. } \rho/2 \\ 0 & \text{w.p. } 1-\rho \end{cases}$

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner with Sparse Rademacher Prior

- Sparsity $\rho \in [0, 1]$
- Prior $\mathcal{X}_\rho$: i.i.d. $x_i \sim \frac{1}{\sqrt{\rho n}} \begin{cases} +1 & \text{w.p. } \rho/2 \\ -1 & \text{w.p. } \rho/2 \\ 0 & \text{w.p. } 1-\rho \end{cases}$
- P_n : spiked Wigner with prior $\mathcal{X}_\rho$, Q_n : unspiked

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner with Sparse Rademacher Prior

- Sparsity $\rho \in [0, 1]$
- Prior $\mathcal{X}_\rho$: i.i.d. $x_i \sim \frac{1}{\sqrt{\rho n}} \begin{cases} +1 & \text{w.p. } \rho/2 \\ -1 & \text{w.p. } \rho/2 \\ 0 & \text{w.p. } 1-\rho \end{cases}$
- P_n : spiked Wigner with prior $\mathcal{X}_\rho$, Q_n : unspiked
- $\tilde{P}_n$: change prior to $\tilde{\mathcal{X}}_\rho$: condition on close-to-typical proportion of nonzeros

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Wigner with Sparse Rademacher Prior

- Sparsity $\rho \in [0, 1]$
- Prior $\mathcal{X}_\rho$: i.i.d. $x_i \sim \frac{1}{\sqrt{\rho n}} \begin{cases} +1 & \text{w.p. } \rho/2 \\ -1 & \text{w.p. } \rho/2 \\ 0 & \text{w.p. } 1-\rho \end{cases}$
- P_n : spiked Wigner with prior $\mathcal{X}_\rho$, Q_n : unspiked
- $\tilde{P}_n$: change prior to $\tilde{\mathcal{X}}_\rho$: condition on close-to-typical proportion of nonzeros

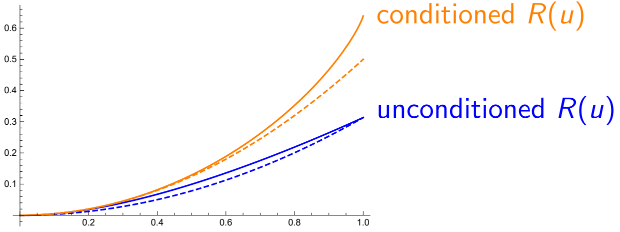
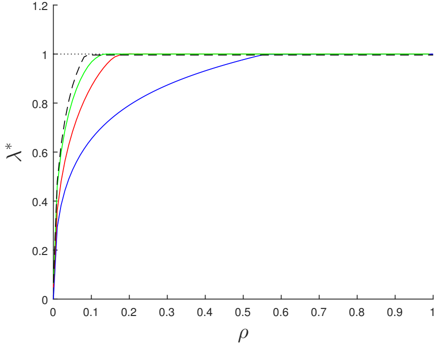


Figure: $\rho = 0.2$

J. Banks, C. Moore, R. Vershynin, J. Xu, 2016.

Sparse Rademacher: Results



- unconditioned
- conditioned
- noise-conditioned (upcoming)
- replica prediction (truth)

Summary

- 3 scenarios:

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)
  - PCA beaten efficiently (non-Gaussian Wigner)

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)
  - PCA beaten efficiently (non-Gaussian Wigner)
  - PCA beaten inefficiently (Rademacher Wishart; sparse Rademacher Wigner)

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)
  - PCA beaten efficiently (non-Gaussian Wigner)
  - PCA beaten inefficiently (Rademacher Wishart; sparse Rademacher Wigner)
- Second moment method: simple, widely-applicable technique for lower bounds.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)
  - PCA beaten efficiently (non-Gaussian Wigner)
  - PCA beaten inefficiently (Rademacher Wishart; sparse Rademacher Wigner)

- Second moment method: simple, widely-applicable technique for lower bounds.

- General formulation in terms of rate functions of priors.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Summary

- 3 scenarios:
  - PCA optimal (Wigner with spherical or Rademacher prior)
  - PCA beaten efficiently (non-Gaussian Wigner)
  - PCA beaten inefficiently (Rademacher Wishart; sparse Rademacher Wigner)
- Second moment method: simple, widely-applicable technique for lower bounds.
- General formulation in terms of rate functions of priors.
- Conditioning method: cut out rare events with an extreme effect on second moment

---

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

- Is it possible to match the replica prediction with a simple method? Or are there fundamental limits to these methods?

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

- Is it possible to match the replica prediction with a simple method? Or are there fundamental limits to these methods?
- In Case 3, are there true statistical-to-computational gaps, or can some method detect or recover efficiently?

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

- Is it possible to match the replica prediction with a simple method? Or are there fundamental limits to these methods?
- In Case 3, are there true statistical-to-computational gaps, or can some method detect or recover efficiently?
- Can we establish sum-of-squares lower bounds for these computational problems?

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

- Is it possible to match the replica prediction with a simple method? Or are there fundamental limits to these methods?
- In Case 3, are there true statistical-to-computational gaps, or can some method detect or recover efficiently?
- Can we establish sum-of-squares lower bounds for these computational problems?
- Is there a connection between SOS lower bounds and replica predictions of hardness?

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+

Further Directions

- Is it possible to match the replica prediction with a simple method? Or are there fundamental limits to these methods?
- In Case 3, are there true statistical-to-computational gaps, or can some method detect or recover efficiently?
- Can we establish sum-of-squares lower bounds for these computational problems?
- Is there a connection between SOS lower bounds and replica predictions of hardness?

Thanks! Questions?

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Optimality and Sub-optimality of PCA for Spiked Random Matrices and Synchronization," 2016.

A. Perry, A. S. Wein, A. S. Bandeira, A. Moitra, "Statistical limits of spiked tensor models," 2016+