Discovering Spikes in Random Matrices Using Approximate Message Passing

Yihan Zhang*

*IST Austria

Youth in High Dimensions 2023, Trieste

Institute of Science and Technology Austria

Spectral Analysis of Generalized Linear Models with General Gaussian Design




Yihan Zhang
IST Austria

Hong Chang Ji
IST Austria

Marco Mondelli
IST Austria

Ramji Venkataramanan
U of Cambridge

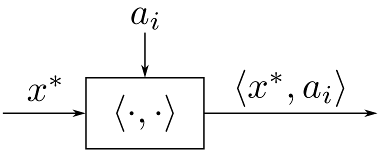
Generalized Linear Models

Generalized Linear Models

$$\xrightarrow{x^*}$$

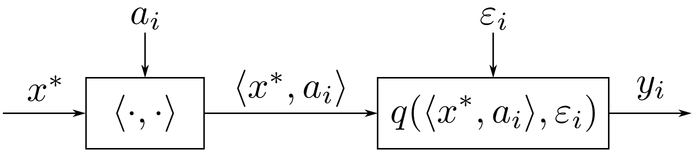
- **Unknown** parameter vector / signal: $x^* \in \sqrt{d}\,\mathbb{S}^{d-1}$

Generalized Linear Models



- **Unknown** parameter vector / signal: $x^* \in \sqrt{d}\,\mathbb{S}^{d-1}$
- **Known** covariate / sensing vectors: $a_i \in \mathbb{R}^d$, $1 \leqslant i \leqslant n$

Generalized Linear Models



- **Unknown** parameter vector / signal: $x^* \in \sqrt{d}\,\mathbb{S}^{d-1}$
- **Known** covariate / sensing vectors: $a_i \in \mathbb{R}^d$, $1 \leqslant i \leqslant n$
- **Known** link function / channel: $q : \mathbb{R}^2 \to \mathbb{R}$, or equivalently: $p(\cdot \mid \cdot)$
- **Unknown** noise: $\varepsilon_i \in \mathbb{R}$, $1 \leqslant i \leqslant n$
- **Observations** / measurements / responses: $y_i = q(\langle x^*, a_i \rangle, \varepsilon_i) \in \mathbb{R}$, $1 \leqslant i \leqslant n$

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$
- 1-bit compressed sensing: $y_i = \mathrm{sgn}(\langle x^*, a_i \rangle) + \varepsilon_i$

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$
- 1-bit compressed sensing: $y_i = \mathrm{sgn}(\langle x^*, a_i \rangle) + \varepsilon_i$
- Logistic regression: $y_i \,|\, \langle x^*, a_i \rangle \sim \frac{1}{1+\exp(\langle x^*, a_i \rangle)} \delta_0 + \frac{\exp(\langle x^*, a_i \rangle)}{1+\exp(\langle x^*, a_i \rangle)} \delta_1$

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$
- 1-bit compressed sensing: $y_i = \mathrm{sgn}(\langle x^*, a_i \rangle) + \varepsilon_i$
- Logistic regression: $y_i \,|\, \langle x^*, a_i \rangle \sim \frac{1}{1+\exp(\langle x^*, a_i \rangle)}\delta_0 + \frac{\exp(\langle x^*, a_i \rangle)}{1+\exp(\langle x^*, a_i \rangle)}\delta_1$
- Polynomial regression

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$
- 1-bit compressed sensing: $y_i = \mathrm{sgn}(\langle x^*, a_i \rangle) + \varepsilon_i$
- Logistic regression: $y_i \,|\, \langle x^*, a_i \rangle \sim \frac{1}{1+\exp(\langle x^*, a_i \rangle)} \delta_0 + \frac{\exp(\langle x^*, a_i \rangle)}{1+\exp(\langle x^*, a_i \rangle)} \delta_1$
- Polynomial regression
- Single-index model

Generalized Linear Models: examples

- Linear regression: $y_i = \langle x^*, a_i \rangle + \varepsilon_i$
- Phase retrieval: $y_i = |\langle x^*, a_i \rangle| + \varepsilon_i$
- 1-bit compressed sensing: $y_i = \mathrm{sgn}(\langle x^*, a_i \rangle) + \varepsilon_i$
- Logistic regression: $y_i \,|\, \langle x^*, a_i \rangle \sim \frac{1}{1+\exp(\langle x^*, a_i \rangle)}\delta_0 + \frac{\exp(\langle x^*, a_i \rangle)}{1+\exp(\langle x^*, a_i \rangle)}\delta_1$
- Polynomial regression
- Single-index model
- etc.

Random design — what's known

Prior art:

Random design — what's known

Prior art:

- i.i.d. Gaussian design: $a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, I_d/n)$, $1 \leqslant i \leqslant n$

Random design — what's known

Prior art:

- i.i.d. Gaussian design: $a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, I_d/n)$, $1 \leqslant i \leqslant n$
- (bi-)orthogonally invariant: $A = O\Lambda Q^\top \in \mathbb{R}^{n\times d}$, where

$$(O, Q) \sim \mathrm{Haar}(\mathbb{O}(n)) \otimes \mathrm{Haar}(\mathbb{O}(d))$$

$$\mathrm{ESD}(\Lambda) := \frac{1}{\min\{n, d\}} \sum_{i=1}^{\min\{n,d\}} \delta_{\Lambda_{i,i}} \to \mathrm{law}(\overline{\Lambda})$$

Random design — what's known

Prior art:

- i.i.d. Gaussian design: $a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, I_d/n)$, $1 \leqslant i \leqslant n$
- (bi-)orthogonally invariant: $A = O \Lambda Q^\top \in \mathbb{R}^{n \times d}$, where

$$(O, Q) \sim \mathrm{Haar}(\mathbb{O}(n)) \otimes \mathrm{Haar}(\mathbb{O}(d))$$

$$\mathrm{ESD}(\Lambda) := \frac{1}{\min\{n, d\}} \sum_{i=1}^{\min\{n,d\}} \delta_{\Lambda_{i,i}} \to \mathrm{law}(\overline{\Lambda})$$

- and everything in their universality class [BLM15, CL21, WZF22, DLS22]

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

or equivalently

$$A = \widetilde{A}\Sigma^{1/2}, \quad \text{where } \widetilde{A}_{i,j} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n),\ (i,j) \in [n] \times [d]$$

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

or equivalently

$$A = \widetilde{A}\Sigma^{1/2}, \quad \text{where } \widetilde{A}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n),\ (i,j) \in [n] \times [d]$$

Why **interesting**?

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

or equivalently

$$A = \widetilde{A}\Sigma^{1/2}, \quad \text{where } \widetilde{A}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n),\ (i,j) \in [n] \times [d]$$

Why **interesting**? — Anisotropic covariates commonly seen in practice, e.g., covariance with Toeplitz / circulant structure

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

or equivalently

$$A = \widetilde{A}\Sigma^{1/2}, \quad \text{where } \widetilde{A}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n),\ (i,j) \in [n] \times [d]$$

Why **interesting**? — Anisotropic covariates commonly seen in practice, e.g., covariance with Toeplitz / circulant structure
Why **nontrivial**?

General Gaussian design

Beyond isotropy and bi-rotational invariance:

$$a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, \Sigma/n), \quad 1 \leqslant i \leqslant n,$$

or equivalently

$$A = \widetilde{A}\Sigma^{1/2}, \quad \text{where } \widetilde{A}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n),\ (i,j) \in [n] \times [d]$$

Why **interesting**? — Anisotropic covariates commonly seen in practice, e.g., covariance with Toeplitz / circulant structure
Why **nontrivial**? — Non-Wigner (rows correlated), only left-rotationally invariant

Estimation from GLMs

Given $\boxed{(y_i, a_i)_{i=1}^n \in (\mathbb{R} \times \mathbb{R}^d)^n}$, statisticians seek an estimate $\widehat{x} = \widehat{x}(y, A) \in \mathbb{R}^d$ of x^*.

Estimation from GLMs

Given $\boxed{(y_i, a_i)_{i=1}^n \in (\mathbb{R} \times \mathbb{R}^d)^n}$, statisticians seek an estimate $\widehat{x} = \widehat{x}(y, A) \in \mathbb{R}^d$ of x^*.
Quality of estimation is measured by **overlap**:

$$\lim_{d \to \infty} \frac{|\langle \widehat{x}, x^* \rangle|}{\|\widehat{x}\|_2 \|x^*\|_2}$$

which is aimed to be maximized.

Estimation from GLMs

Given $\boxed{(y_i, a_i)_{i=1}^n \in (\mathbb{R} \times \mathbb{R}^d)^n}$, statisticians seek an estimate $\widehat{x} = \widehat{x}(y, A) \in \mathbb{R}^d$ of x^*.
Quality of estimation is measured by **overlap**:

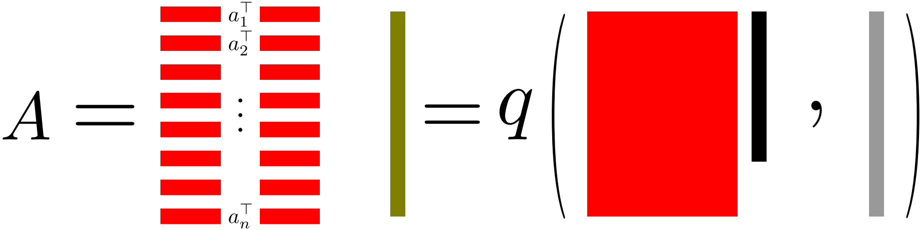
$$\lim_{d \to \infty} \frac{|\langle \widehat{x}, x^* \rangle|}{\|\widehat{x}\|_2 \|x^*\|_2}$$

which is aimed to be maximized.
Proportional regime: $n, d \to \infty$

$$\textbf{aspect ratio} = \frac{\#\text{observations}}{\#\text{parameter dimensions}} = \frac{n}{d} \to \delta \in (0, \infty).$$

Assumptions

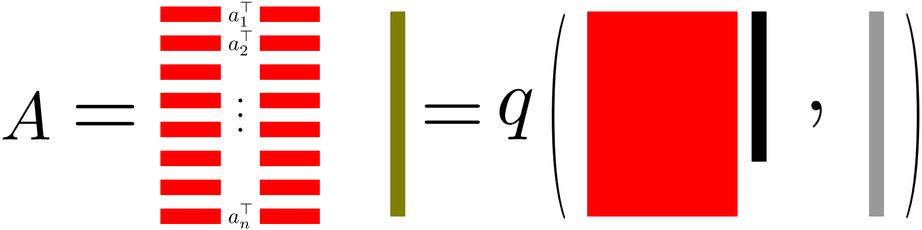
GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



[1]May not be necessary

Assumptions

GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



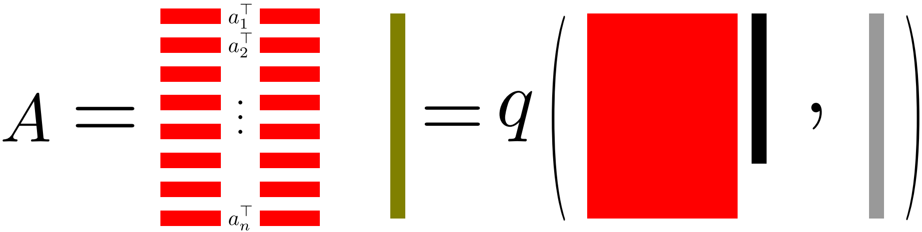
Assume

- $\Sigma \in \mathbb{R}^{d \times d}$ is **unknown**

[1]May not be necessary

Assumptions

GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



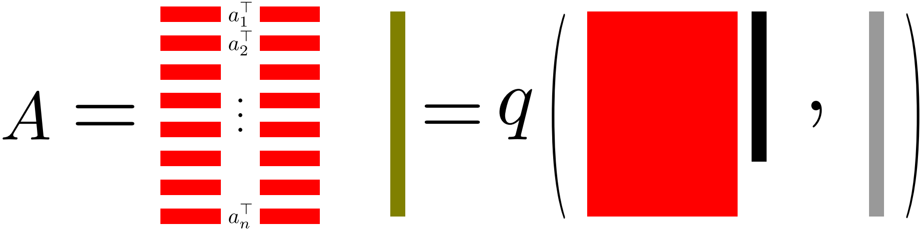
Assume

- $\Sigma \in \mathbb{R}^{d \times d}$ is **unknown**
- $\Sigma \succ 0_{d \times d}$ is strictly positive definite[1] with uniformly (in d) bounded $\|\Sigma\|_2$

[1] May not be necessary

Assumptions

GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



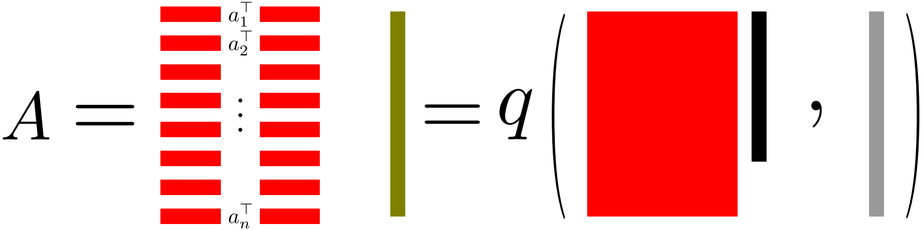
Assume

- $\Sigma \in \mathbb{R}^{d \times d}$ is **unknown**
- $\Sigma \succ 0_{d \times d}$ is strictly positive definite[1] with uniformly (in d) bounded $\|\Sigma\|_2$
- $\mathrm{ESD}(\Sigma) := \frac{1}{d} \sum_{i=1}^{d} \delta_{\lambda_i(\Sigma)} \to \mathrm{law}(\overline{\Sigma})$

[1] May not be necessary

Assumptions

GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



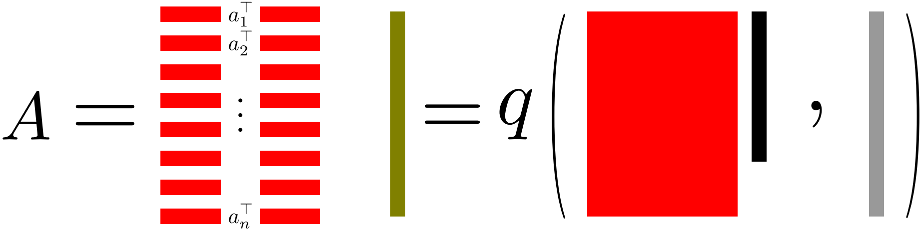
Assume

- $\Sigma \in \mathbb{R}^{d \times d}$ is **unknown**
- $\Sigma \succ 0_{d\times d}$ is strictly positive definite[1] with uniformly (in d) bounded $\|\Sigma\|_2$
- $\mathrm{ESD}(\Sigma) := \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(\Sigma)} \to \mathrm{law}(\overline{\Sigma})$
- **Unstructured** prior: $x^* \sim \mathrm{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$

[1] May not be necessary

Assumptions

GLM: $y = q(Ax^*, \varepsilon) \in \mathbb{R}^n$



Assume

- $\Sigma \in \mathbb{R}^{d \times d}$ is **unknown**
- $\Sigma \succ 0_{d \times d}$ is strictly positive definite[1] with uniformly (in d) bounded $\|\Sigma\|_2$
- $\mathrm{ESD}(\Sigma) := \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(\Sigma)} \to \mathrm{law}(\overline{\Sigma})$
- **Unstructured** prior: $x^* \sim \mathrm{Unif}(\sqrt{d}\,\mathbb{S}^{d-1})$
- $\frac{1}{n}\sum_{i=1}^{n} \delta_{\varepsilon_i} \to P_\varepsilon$

[1] May not be necessary

Main character: spectral estimator

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$,

$$\frac{1}{n} \sum_{i=1}^{n} a_i a_i^\top \in \mathbb{R}^{d \times d}$$

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$,

$$\frac{1}{n} \sum_{i=1}^{n} a_i a_i^\top \cdot y_i \in \mathbb{R}^{d \times d}$$

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$, pick our favorite **preprocessing function** $\mathcal{T} : \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n} \sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d \times d}.$$

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$, pick our favorite **preprocessing function** $\mathcal{T} \colon \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n} \sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d \times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n \times d}$, pick our favorite **preprocessing function** $\mathcal{T} \colon \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n} \sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d \times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

Why **spectral**?

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$, pick our favorite **preprocessing function** $\mathcal{T} \colon \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n}\sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d\times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

Why **spectral**?

- Computationally simple

[2]Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$, pick our favorite **preprocessing function** $\mathcal{T}: \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n}\sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d\times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

Why **spectral**?

- Computationally simple
- Warm start of AM (cf. Chandrasekher's talk), EM, AMP, GD, etc.

[2]Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$, pick our favorite **preprocessing function** $\mathcal{T}\colon \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n}\sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d\times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

Why **spectral**?

- Computationally simple
- Warm start of AM (cf. Chandrasekher's talk), EM, AMP, GD, etc.
- Interesting RMT problem per se

[2] Somewhat necessary

Main character: spectral estimator

Given $(y, A) \in \mathbb{R}^n \times \mathbb{R}^{n\times d}$, pick our favorite **preprocessing function** $\mathcal{T} \colon \mathbb{R} \to \mathbb{R}$

$$D = \frac{1}{n}\sum_{i=1}^{n} a_i a_i^\top \cdot \mathcal{T}(y_i) \in \mathbb{R}^{d\times d}.$$

The **spectral estimator** generates the top eigenvector of D:

$$\widehat{x}^{\mathrm{spec}}(\mathcal{T}) = v_1(D) \in \mathbb{S}^{d-1},$$

Why **spectral**?

- Computationally simple
- Warm start of AM (cf. Chandrasekher's talk), EM, AMP, GD, etc.
- Interesting RMT problem per se

Assume

- $\mathcal{T}$ is bounded from above and below, not constantly 0, and[2] $\boxed{\sup \operatorname{supp}(\mathcal{T}) > 0}$

[2]Somewhat necessary

Main result

Theorem (Eigenvalues & overlap)

Let $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta), \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta), \eta(\overline{\Sigma}, \mathcal{T}, \delta)$ *be defined as [insert complicated expressions].*

If $\boxed{\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)}$*, then*

$$\lim_{d \to \infty} \lambda_1(D) = \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta), \tag{1}$$

$$\lim_{d \to \infty} \lambda_2(D) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta), \tag{2}$$

$$\lim_{d \to \infty} \frac{|\langle v_1(D), x^* \rangle|}{\|v_1(D)\|_2 \|x^*\|_2} = \eta(\overline{\Sigma}, \mathcal{T}, \delta) > 0 \tag{3}$$

Main result

Theorem (Eigenvalues & overlap)

Let $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta), \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta), \eta(\overline{\Sigma}, \mathcal{T}, \delta)$ be defined as [insert complicated expressions].
If $\boxed{\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)}$, then

$$\lim_{d\to\infty} \lambda_1(D) = \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta), \tag{1}$$

$$\lim_{d\to\infty} \lambda_2(D) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta), \tag{2}$$

$$\lim_{d\to\infty} \frac{|\langle v_1(D), x^*\rangle|}{\|v_1(D)\|_2\|x^*\|_2} = \eta(\overline{\Sigma}, \mathcal{T}, \delta) > 0 \tag{3}$$

- The condition $\boxed{\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)}$ is very likely the **phase transition threshold**, though we do not have converse result **:-(**

In case you care about arcane formulas...

In case you care about arcane formulas...

Define random variables $(\overline{G}, \overline{\varepsilon}) \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right) \otimes P_\varepsilon$, $\overline{Y} = q(\overline{G}, \overline{\varepsilon})$.

In case you care about arcane formulas...

Define random variables $(\overline{G}, \overline{\varepsilon}) \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right) \otimes P_\varepsilon$, $\overline{Y} = q(\overline{G}, \overline{\varepsilon})$.

Denote by $(a(\overline{\Sigma}, \mathcal{T}, \delta), \gamma(\overline{\Sigma}, \mathcal{T}, \delta))$ the unique solution to

$$1 = \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a-\mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

$$1 = \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a-\mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

In case you care about arcane formulas...

Define random variables $(\overline{G}, \overline{\varepsilon}) \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right) \otimes P_\varepsilon$, $\overline{Y} = q(\overline{G}, \overline{\varepsilon})$.

Denote by $(a(\overline{\Sigma}, \mathcal{T}, \delta), \gamma(\overline{\Sigma}, \mathcal{T}, \delta))$ the unique solution to

$$1 = \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

$$1 = \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

Then $\boxed{\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) := a(\overline{\Sigma}, \mathcal{T}, \delta)\gamma(\overline{\Sigma}, \mathcal{T}, \delta)}$.

In case you care about arcane formulas...

Define random variables $(\overline{G}, \overline{\varepsilon}) \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right) \otimes P_\varepsilon$, $\overline{Y} = q(\overline{G}, \overline{\varepsilon})$.

Denote by $(a(\overline{\Sigma}, \mathcal{T}, \delta), \gamma(\overline{\Sigma}, \mathcal{T}, \delta))$ the unique solution to

$$1 = \frac{1}{\mathbb{E}[\overline{\Sigma}]}\mathbb{E}\left[\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\frac{\mathcal{T}(\overline{Y})}{a - \mathcal{T}(\overline{Y})}\right]\mathbb{E}\left[\frac{\overline{\Sigma}^2}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a-\mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

$$1 = \frac{1}{\delta}\mathbb{E}\left[\frac{\overline{\Sigma}}{\gamma - \mathbb{E}\left[\frac{\mathcal{T}(\overline{Y})}{a-\mathcal{T}(\overline{Y})}\right]\overline{\Sigma}}\right]$$

Then $\boxed{\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) := a(\overline{\Sigma}, \mathcal{T}, \delta)\gamma(\overline{\Sigma}, \mathcal{T}, \delta)}$.

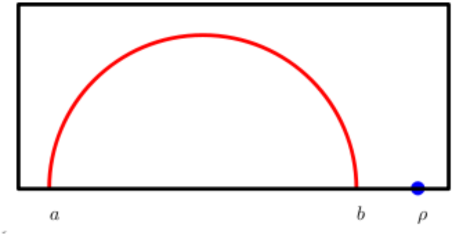
The other parameters $\boxed{\lambda_2(\overline{\Sigma}, \mathcal{T}, \delta), \eta(\overline{\Sigma}, \mathcal{T}, \delta)}$ admit similarly messy expressions.

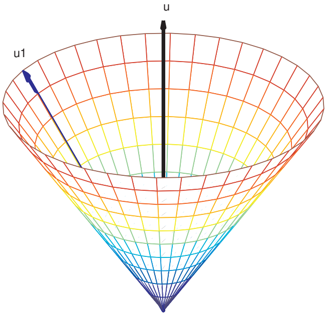
Basic intuition

$$D = A^\top T A \in \mathbb{R}^{d\times d}, \text{ where } A = \begin{bmatrix} \text{---}\; a_1^\top \;\text{---} \\ \cdots \\ \text{---}\; a_n^\top \;\text{---} \end{bmatrix} \in \mathbb{R}^{n\times d}, \quad T = \begin{bmatrix} \mathcal{T}(y_1) & & \\ & \ddots & \\ & & \mathcal{T}(y_n) \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Basic intuition

$$D = A^\top T A \in \mathbb{R}^{d\times d},\ \text{where } A = \begin{bmatrix} - & a_1^\top & - \\ & \cdots & \\ - & a_n^\top & - \end{bmatrix} \in \mathbb{R}^{n\times d},\ T = \begin{bmatrix} \mathcal{T}(y_1) & & \\ & \ddots & \\ & & \mathcal{T}(y_n) \end{bmatrix} \in \mathbb{R}^{n\times n}$$





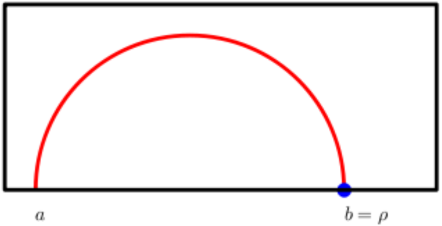
Outlier eigenvalue $\lim_{d\to\infty} \lambda_1(D) > \lim_{d\to\infty} \lambda_2(D)$

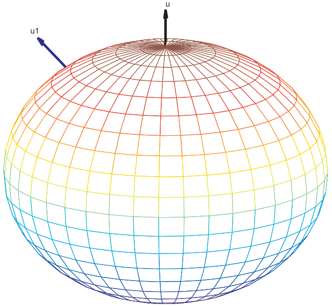
Positive overlap $\lim_{d\to\infty} \frac{|\langle v_1(D), x^* \rangle|}{\|v_1(D)\|_2 \|x^*\|_2} > 0$

Supercritical regime $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$

Basic intuition

$$D = A^\top T A \in \mathbb{R}^{d\times d}, \text{ where } A = \begin{bmatrix} — a_1^\top — \\ \cdots \\ — a_n^\top — \end{bmatrix} \in \mathbb{R}^{n\times d}, \quad T = \begin{bmatrix} \mathcal{T}(y_1) & & \\ & \ddots & \\ & & \mathcal{T}(y_n) \end{bmatrix} \in \mathbb{R}^{n\times n}$$





No outlier $\lim_{d\to\infty} \lambda_1(D) = \lim_{d\to\infty} \lambda_2(D)$

Zero overlap $\lim_{d\to\infty} \frac{|\langle v_1(D), x^*\rangle|}{\|v_1(D)\|_2 \|x^*\|_2} = 0$

Subcritical regime $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) \leqslant \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$

Spiked, or non-spiked, that is the question

Spectral matrix: $D = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2}$

where

$$T = \mathrm{diag}(\mathcal{T}(y)) = \mathrm{diag}(\mathcal{T}(q(\widetilde{A}\Sigma^{1/2}x^*, \varepsilon)))$$

Spiked, or non-spiked, that is the question

$$\text{Spectral matrix:} \quad D = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2}$$

v.s.

$$\textbf{Separable covariance matrix?} \quad \widehat{D} = \Sigma^{1/2} \widehat{A}^\top T \widehat{A} \Sigma^{1/2}$$

where

$$T = \operatorname{diag}(\mathcal{T}(y)) = \operatorname{diag}(\mathcal{T}(q(\widetilde{A}\Sigma^{1/2} x^*, \varepsilon)))$$

and

$$\widehat{A} \overset{\mathrm{d}}{=} \widetilde{A}, \text{ but } \textbf{independent} \text{ of } T$$

Eigenvalue interlacing

GLM: $\quad y = q(\widetilde{A}\Sigma x^*, \varepsilon) \in \mathbb{R}^n$

Spectral matrix: $\quad D = \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} \in \mathbb{R}^{d\times d}$

Eigenvalue interlacing

$$\text{GLM:} \quad y = q(\widetilde{A}\Sigma x^*, \varepsilon) \in \mathbb{R}^n$$

$$\text{Spectral matrix:} \quad D = \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} \in \mathbb{R}^{d\times d}$$

Let us ‘take out’ the spike.

Eigenvalue interlacing

GLM: $\quad y = q(\widetilde{A}\Sigma x^*, \varepsilon) \in \mathbb{R}^n$

Spectral matrix: $\quad D = \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} \in \mathbb{R}^{d\times d}$

Let us 'take out' the spike. **Fact**: for any $g \in \mathbb{R}^n$

$$\widetilde{A} \overset{\mathrm{d}}{=} \Pi_g \widetilde{A} + \Pi_g^\perp \widehat{A}, \quad \Pi_g = \frac{gg^\top}{\|g\|_2^2} \in \mathbb{R}^{n\times n}, \quad \Pi_g^\perp = I_n - \Pi_g \in \mathbb{R}^{n\times n},$$

$\widehat{A} \overset{\mathrm{d}}{=} \widetilde{A}$ **independent** of everything else

Eigenvalue interlacing

GLM: $\quad y = q(\widetilde{A}\Sigma x^*, \varepsilon) \in \mathbb{R}^n$

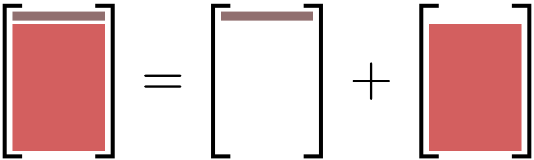
Spectral matrix: $\quad D = \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} \in \mathbb{R}^{d\times d}$

Let us 'take out' the spike. **Fact**: for any $g \in \mathbb{R}^n$

$$\widetilde{A} \overset{\mathrm{d}}{=} \Pi_g \widetilde{A} + \Pi_g^\perp \widehat{A}, \quad \Pi_g = \frac{gg^\top}{\|g\|_2^2} \in \mathbb{R}^{n\times n}, \quad \Pi_g^\perp = I_n - \Pi_g \in \mathbb{R}^{n\times n},$$

$\widehat{A} \overset{\mathrm{d}}{=} \widetilde{A}$ **independent** of everything else

If $g = e_1 \in \mathbb{R}^n$, then $\widetilde{A} \overset{\mathrm{d}}{=} e_1 e_1^\top \widetilde{A} + (I_n - e_1 e_1^\top)\widehat{A}$



Eigenvalue interlacing

$$\text{GLM:} \quad y = q(\widetilde{A}\Sigma x^*, \varepsilon) \in \mathbb{R}^n$$

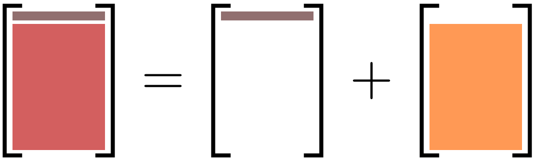
$$\text{Spectral matrix:} \quad D = \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} \in \mathbb{R}^{d\times d}$$

Let us 'take out' the spike. **Fact**: for any $g \in \mathbb{R}^n$

$$\widetilde{A} \overset{\mathrm{d}}{=} \Pi_g \widetilde{A} + \Pi_g^\perp \widehat{A}, \quad \Pi_g = \frac{gg^\top}{\|g\|_2^2} \in \mathbb{R}^{n\times n}, \quad \Pi_g^\perp = I_n - \Pi_g \in \mathbb{R}^{n\times n},$$

$\widehat{A} \overset{\mathrm{d}}{=} \widetilde{A}$ **independent** of everything else

If $g = e_1 \in \mathbb{R}^n$, then $\widetilde{A} \overset{\mathrm{d}}{=} e_1 e_1^\top \widetilde{A} + (I_n - e_1 e_1^\top)\widehat{A}$



Eigenvalue interlacing (cont'd)

With $g = \widetilde{A}\Sigma^{1/2}x^*$, D becomes

$$D \overset{\mathrm{d}}{=} \left(\underbrace{\frac{\Sigma^{1/2}\widetilde{A}^\top g}{\|g\|_2}\frac{g^\top}{\|g\|_2}}_{\text{rank 1}} + \Sigma^{1/2}\widehat{A}^\top \Pi_g^\perp \right) T \left(\underbrace{\frac{g}{\|g\|_2}\frac{g^\top \widetilde{A}\Sigma^{1/2}}{\|g\|_2}}_{\text{rank 1}} + \Pi_g^\perp \widehat{A}\Sigma^{1/2} \right)$$

Eigenvalue interlacing (cont'd)

With $g = \widetilde{A}\Sigma^{1/2}x^*$, D becomes

$$D \overset{\mathrm{d}}{=} \Bigg(\underbrace{\frac{\Sigma^{1/2}\widetilde{A}^\top g}{\|g\|_2}\frac{g^\top}{\|g\|_2}}_{\text{rank 1}} + \Sigma^{1/2}\widehat{A}^\top\Pi_g^\perp \Bigg) T \Bigg(\underbrace{\frac{g}{\|g\|_2}\frac{g^\top\widetilde{A}\Sigma^{1/2}}{\|g\|_2}}_{\text{rank 1}} + \Pi_g^\perp\widehat{A}\Sigma^{1/2} \Bigg)$$

The 'null' matrix $\widehat{D}$ has no spike:

$$\widehat{D} = \Sigma^{1/2}\widehat{A}^\top\Pi_g^\perp T \Pi_g^\perp\widehat{A}\Sigma^{1/2}$$

Eigenvalue interlacing (cont'd)

With $g = \widetilde{A}\Sigma^{1/2}x^*$, D becomes

$$D \overset{\mathrm{d}}{=} \Bigg(\underbrace{\frac{\Sigma^{1/2}\widetilde{A}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2}}_{\text{rank 1}} + \Sigma^{1/2}\widehat{A}^\top \Pi_g^\perp \Bigg) T \Bigg(\underbrace{\frac{g}{\|g\|_2} \frac{g^\top \widetilde{A}\Sigma^{1/2}}{\|g\|_2}}_{\text{rank 1}} + \Pi_g^\perp \widehat{A}\Sigma^{1/2} \Bigg)$$

The 'null' matrix $\widehat{D}$ has no spike:

$$\widehat{D} = \Sigma^{1/2}\widehat{A}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{A}\Sigma^{1/2} \overset{\mathrm{d}}{=} \Sigma^{1/2} \underbrace{\check{A}^\top}_{d\times(n-1)} \underbrace{\check{T}}_{(n-1)\times(n-1)} \underbrace{\check{A}}_{(n-1)\times d} \Sigma^{1/2},$$

where $\lambda_n(T) \geqslant \lambda_{n-1}(\check{T}) \geqslant \lambda_{n-1}(T) \cdots \geqslant \lambda_2(\check{T}) \geqslant \lambda_2(T) \geqslant \lambda_1(\check{T}) \geqslant \lambda_1(T)$

Eigenvalue interlacing (cont'd)

With $g = \widetilde{A}\Sigma^{1/2}x^*$, D becomes

$$D \overset{\mathrm{d}}{=} \Bigg(\underbrace{\frac{\Sigma^{1/2}\widetilde{A}^\top g}{\|g\|_2} \frac{g^\top}{\|g\|_2}}_{\text{rank 1}} + \Sigma^{1/2}\widehat{A}^\top \Pi_g^\perp \Bigg) T \Bigg(\underbrace{\frac{g}{\|g\|_2} \frac{g^\top \widetilde{A}\Sigma^{1/2}}{\|g\|_2}}_{\text{rank 1}} + \Pi_g^\perp \widehat{A}\Sigma^{1/2} \Bigg)$$

The 'null' matrix $\widehat{D}$ has no spike:

$$\widehat{D} = \Sigma^{1/2}\widehat{A}^\top \Pi_g^\perp T \Pi_g^\perp \widehat{A}\Sigma^{1/2} \overset{\mathrm{d}}{=} \Sigma^{1/2} \underbrace{\check{A}^\top}_{d\times(n-1)} \underbrace{\check{T}}_{(n-1)\times(n-1)} \underbrace{\check{A}}_{(n-1)\times d} \Sigma^{1/2},$$

where $\lambda_n(T) \geqslant \lambda_{n-1}(\check{T}) \geqslant \lambda_{n-1}(T) \cdots \geqslant \lambda_2(\check{T}) \geqslant \lambda_2(T) \geqslant \lambda_1(\check{T}) \geqslant \lambda_1(T)$

Eigenvalues of D and $\widehat{D}$ are interlaced: $\lambda_{i+1}(\widehat{D}) \leqslant \lambda_i(D) \leqslant \lambda_{i-1}(\widehat{D})$ for $2 \leqslant i \leqslant d-1$

$$\begin{array}{cccccccc} \lambda_d(D) & \lambda_{d-1}(D) & \lambda_{d-2}(D) & \lambda_{d-3}(D) & \cdots & \lambda_3(D) & \lambda_2(D) & \lambda_1(D) \\ \\ \lambda_d(\widehat{D}) & \lambda_{d-1}(\widehat{D}) & \lambda_{d-2}(\widehat{D}) & \lambda_{d-3}(\widehat{D}) & \cdots & \lambda_3(\widehat{D}) & \lambda_2(\widehat{D}) & \lambda_1(\widehat{D}) \end{array}$$

Right edge of the bulk

$$\lim_{d\to\infty} \lambda_3(\widehat{D}) \quad \leq \quad \lim_{d\to\infty} \lambda_2(D) \quad \leq \quad \lim_{d\to\infty} \lambda_1(\widehat{D})$$

$$\sup \operatorname{supp}(\underbrace{\,\overline{\mu}_{\widehat{D}}\,}_{\Big\uparrow \atop \mathrm{ESD}(\widehat{D})})$$

Right edge of the bulk

$$\lim_{d\to\infty}\lambda_3(\widehat{D}) \quad \le \quad \lim_{d\to\infty}\lambda_2(D) \quad \le \quad \lim_{d\to\infty}\lambda_1(\widehat{D})$$

$$\sup\operatorname{supp}(\underbrace{\overline{\mu}_{\widehat{D}}}_{\substack{\uparrow \\ \mathrm{ESD}(\widehat{D})}})$$

If $\mathcal{T} \geqslant 0$, $\widehat{D} = \Sigma^{1/2}\widehat{A}^\top T \widehat{A}\Sigma^{1/2}$ is PSD and is well-understood in RMT under the name of ‘separable covariance matrix’.

Right edge of the bulk

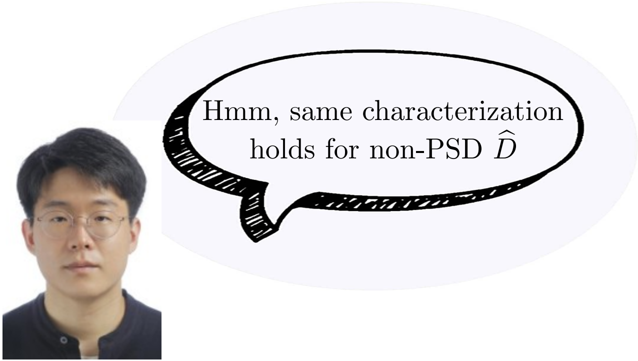
$$\lim_{d\to\infty}\lambda_3(\widehat{D}) \quad \le \quad \lim_{d\to\infty}\lambda_2(D) \quad \le \quad \lim_{d\to\infty}\lambda_1(\widehat{D})$$

$$\sup\operatorname{supp}(\underbrace{\overline{\mu}_{\widehat{D}}}_{\substack{\uparrow \\ \mathrm{ESD}(\widehat{D})}})$$

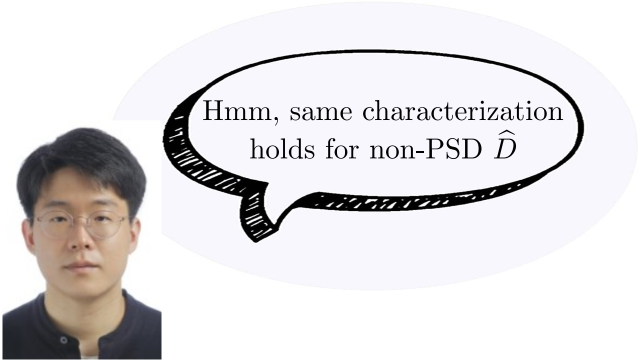
If $\mathcal{T} \geqslant 0$, $\widehat{D} = \Sigma^{1/2}\widehat{A}^\top T\widehat{A}\Sigma^{1/2}$ is PSD and is well-understood in RMT under the name of ‘separable covariance matrix’.

CAUTION **But**, we only assume $\sup\operatorname{supp}(\mathcal{T}) > 0$.

Right edge of the bulk (cont'd)

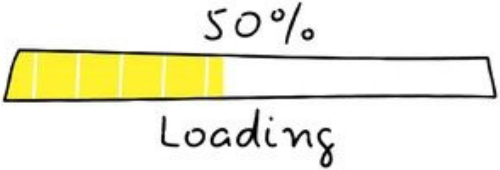


Right edge of the bulk (cont'd)

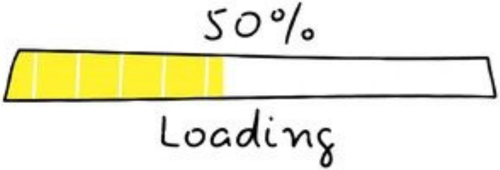


$$\lim_{d\to\infty} \lambda_2(D) = \sup \operatorname{supp}(\overline{\mu}_{\widehat{D}}) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$$

Where are we now?



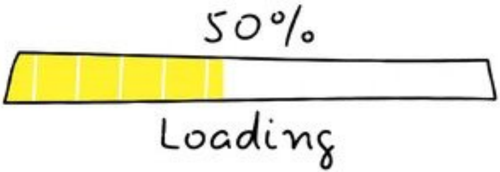
Where are we now?



We have identified (2) the right edge of the bulk:

$$\lim_{d\to\infty} \lambda_2(D) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$$

Where are we now?



We have identified (2) the right edge of the bulk:

$$\lim_{d\to\infty} \lambda_2(D) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$$

But where is (1) the spike

$$\lim_{d\to\infty} \lambda_1(D)?$$

And how large is (3) the overlap

$$\lim_{d\to\infty} \frac{|\langle v_1(D), x^* \rangle|}{\|v_1(D)\|_2 \|x^*\|_2}?$$

A hammer: Generalized Approximate Message Passing

Power iteration:

$$v^{t+1} = \frac{\Sigma^{1/2}\widetilde{A}^\top T\widetilde{A}\Sigma^{1/2}v^t}{\left\|\Sigma^{1/2}\widetilde{A}^\top T\widetilde{A}\Sigma^{1/2}v^t\right\|_2}$$

A hammer: Generalized Approximate Message Passing

Power iteration:

$$v^{t+1} = \frac{\Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} v^t}{\left\| \Sigma^{1/2}\widetilde{A}^\top T \widetilde{A}\Sigma^{1/2} v^t \right\|_2}$$

(**G**eneralized) **A**pproximate **M**essage **P**assing (**GAMP**)
Given $g_t \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, f_{t+1} \colon \mathbb{R}^d \to \mathbb{R}^d$ for every $t \geqslant 1$

$$u^t = \widetilde{A}\widetilde{v}^t - b_t \widetilde{u}^{t-1} \in \mathbb{R}^n, \quad \widetilde{u}^t = g_t(u^t; y) \in \mathbb{R}^n, \quad c_t = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \frac{\partial g_t(u^t; y)_i}{\partial u_i^t} \in \mathbb{R},$$

$$v^{t+1} = \widetilde{A}^\top \widetilde{u}^t - c_t \widetilde{v}^t \in \mathbb{R}^d, \quad \widetilde{v}^{t+1} = f_{t+1}(v^{t+1}) \in \mathbb{R}^d, \quad b_{t+1} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^d \frac{\partial f_{t+1}(v^{t+1})_i}{\partial v_i^{t+1}} \in \mathbb{R}.$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$u^t = \widetilde{A} f_t(v^t) - b_t g_{t-1}(u^{t-1}), \quad v^{t+1} = \widetilde{A}^\top g_t(u^t) - c_t f_t(v^t).$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$u = \widetilde{A} f(v) - b g(u), \quad v = \widetilde{A}^\top g(u) - c f(v).$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$u = \widetilde{A} f(v) - b g(u), \quad v = \widetilde{A}^\top g(u) - c f(v).$$

Choose

$$g(u) = (aI_n - T)^{-1} T u \in \mathbb{R}^n, \quad f(v) = (\gamma I_d - c\Sigma)^{-1} \Sigma v,$$
$$c = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} g(u), \quad b = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} f(v),$$

where (a, γ) are such that

$$b = 1, \quad \lim_{d\to\infty} \frac{1}{d} \|f(v)\|_2^2 = 1$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$\gamma \Sigma^{-1} f(v) = \frac{1}{a} \widetilde{A}^\top T \widetilde{A} f(v)$$

Choose

$$g(u) = (aI_n - T)^{-1} T u \in \mathbb{R}^n, \quad f(v) = (\gamma I_d - c\Sigma)^{-1} \Sigma v,$$
$$c = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} g(u), \quad b = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} f(v),$$

where (a, γ) are such that

$$b = 1, \quad \lim_{d\to\infty} \frac{1}{d} \|f(v)\|_2^2 = 1$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$a\gamma \Sigma^{-1/2} f(v) = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} \Sigma^{-1/2} f(v)$$

Choose

$$g(u) = (aI_n - T)^{-1} Tu \in \mathbb{R}^n, \quad f(v) = (\gamma I_d - c\Sigma)^{-1} \Sigma v,$$
$$c = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} g(u), \quad b = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} f(v),$$

where (a, γ) are such that

$$b = 1, \quad \lim_{d\to\infty} \frac{1}{d} \|f(v)\|_2^2 = 1$$

Simulating power iteration using GAMP

Eigenequation:

$$v = \Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2} v$$

GAMP:

$$\underbrace{a\gamma}_{\text{eigval}} \underbrace{\Sigma^{-1/2} f(v)}_{\text{eigvec}} = \underbrace{\Sigma^{1/2} \widetilde{A}^\top T \widetilde{A} \Sigma^{1/2}}_{D} \underbrace{\Sigma^{-1/2} f(v)}_{\text{eigvec}}$$

Choose

$$g(u) = (aI_n - T)^{-1} T u \in \mathbb{R}^n, \quad f(v) = (\gamma I_d - c\Sigma)^{-1} \Sigma v,$$

$$c = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} g(u), \quad b = \lim_{n\to\infty} \frac{1}{n} \operatorname{div} f(v),$$

where (a, γ) are such that

$$b = 1, \quad \lim_{d\to\infty} \frac{1}{d} \|f(v)\|_2^2 = 1$$

State evolution

For each $t \geqslant 1$, define the random vector

$$V_t := \chi_t \Sigma^{1/2} X^* + \sigma_t W_t \in \mathbb{R}^d$$

where

$$(X^*, \sigma_1 W_1, \cdots, \sigma_t W_t) \to \mathcal{N}(0_d, I_d) \otimes \mathcal{N}(0_{d \times t}, \Psi_t \otimes I_d)$$

State evolution

For each $t \geqslant 1$, define the random vector

$$V_t := \chi_t \Sigma^{1/2} X^* + \sigma_t W_t \in \mathbb{R}^d$$

where

$$(X^*, \sigma_1 W_1, \cdots, \sigma_t W_t) \to \mathcal{N}(0_d, I_d) \otimes \mathcal{N}(0_{d\times t}, \Psi_t \otimes I_d)$$

Theorem (State evolution)

For any uniformly pseudo-Lipschitz $h \colon \mathbb{R}^{d\times t} \to \mathbb{R}$

$$\lim_{d\to\infty} h(v^1, v^2, \cdots, v^t) - \mathbb{E}[h(V_1, V_2, \cdots, V_t)] = 0$$

Crux of the argument: locating the spike

$$v^{t+1} = \frac{D}{a\gamma} v^t + e^t$$

Crux of the argument: locating the spike

$$v^{t+1} = \frac{D}{a\gamma} v^t$$

Crux of the argument: locating the spike

$$v^{t+t'} = \left(\frac{D}{a\gamma}\right)^{t'} v^t$$

Crux of the argument: locating the spike

$$\frac{1}{d}\left\|v^{t+t'}\right\|_2^2 = \frac{1}{d}\left\|\left(\frac{D}{a\gamma}\right)^{t'} v^t\right\|_2^2$$

Crux of the argument: locating the spike

$$\frac{1}{d}\left\|v^{t+t'}\right\|_2^2 = \left(\frac{\lambda_1(D)}{a\gamma}\right)^{2t'} \frac{\left\langle v_1(D),\, v^t\right\rangle^2}{d} + \sum_{i=2}^{d}\left(\frac{\lambda_i(D)}{a\gamma}\right)^{2t'} \frac{\left\langle v_i(D),\, v^t\right\rangle^2}{d}$$

Crux of the argument: locating the spike

$$
\begin{aligned}
\lim_{d\to\infty} \frac{1}{d}\left\| v^{t+t'} \right\|_2^2 &= \lim_{d\to\infty} \left(\frac{\lambda_1(D)}{a\gamma}\right)^{2t'} \frac{\left\langle v_1(D),\, v^t \right\rangle^2}{d} \\
&\quad + \lim_{d\to\infty} \sum_{i=2}^{d} \left(\frac{\lambda_i(D)}{a\gamma}\right)^{2t'} \frac{\left\langle v_i(D),\, v^t \right\rangle^2}{d}
\end{aligned}
$$

Crux of the argument: locating the spike

$$\lim_{d\to\infty} \frac{1}{d}\left\| v^{t+t'} \right\|_2^2 = \lim_{d\to\infty} \left(\frac{\lambda_1(D)}{a\gamma} \right)^{2t'} \frac{\left\langle v_1(D), v^t \right\rangle^2}{d}$$

provided

$$\lim_{d\to\infty} \lambda_2(D) = \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta) < a\gamma =: \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta)$$

Crux of the argument: locating the spike

$$\lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\left\|v^{t+t'}\right\|_2^2=\lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left(\frac{\lambda_1(D)}{a\gamma}\right)^{2t'}\frac{\left\langle v_1(D),v^t\right\rangle^2}{d}$$

Crux of the argument: locating the spike

$$\nu^2 = \lim_{t' \to \infty} \lim_{t \to \infty} \lim_{d \to \infty} \left(\frac{\lambda_1(D)}{a\gamma} \right)^{2t'} \frac{\langle v_1(D), v^t \rangle^2}{d}$$

where $\nu > 0$ is a constant derived from the state evolution

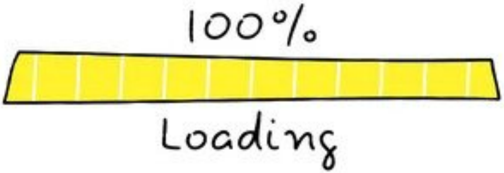
Crux of the argument: locating the spike

$$\nu^2 = \lim_{t'\to\infty} \lim_{t\to\infty} \lim_{d\to\infty} \left(\frac{\lambda_1(D)}{a\gamma}\right)^{2t'} \frac{\left\langle v_1(D), v^t \right\rangle^2}{d}$$

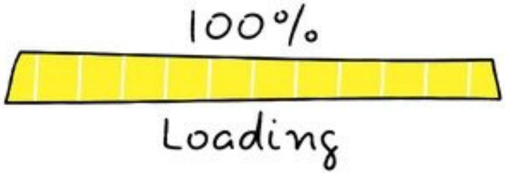
which implies

$$\lim_{d\to\infty} \lambda_1(D) = a\gamma = \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta), \quad \lim_{t\to\infty} \lim_{d\to\infty} \frac{\left\langle v_1(D), v^t \right\rangle^2}{\|v_1(D)\|_2^2 \|v^t\|_2^2} = 1$$

Where are we now?



Where are we now?



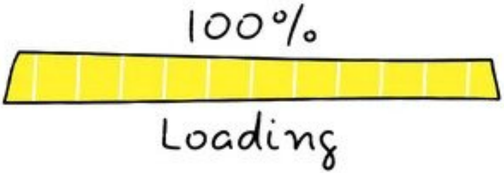
If $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$,

$$(1) \quad \lim_{d\to\infty} \lambda_1(D) = \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta),$$

$$(3) \quad \lim_{d\to\infty} \frac{|\langle v_1(D), x^* \rangle|}{\|v_1(D)\|_2 \|x^*\|_2} = \lim_{t\to\infty} \lim_{d\to\infty} \frac{|\langle v^t, x^* \rangle|}{\|v^t\|_2 \|x^*\|_2} = \eta(\overline{\Sigma}, \mathcal{T}, \delta) > 0,$$

where $\eta(\overline{\Sigma}, \mathcal{T}, \delta)$ can be derived from the state evolution.

Where are we now?



If $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$,

$$(1) \qquad \lim_{d\to\infty} \lambda_1(D) = \lambda_1(\overline{\Sigma}, \mathcal{T}, \delta),$$

$$(3) \qquad \lim_{d\to\infty} \frac{|\langle v_1(D), x^* \rangle|}{\|v_1(D)\|_2 \|x^*\|_2} = \lim_{t\to\infty} \lim_{d\to\infty} \frac{\left|\left\langle v^t, x^* \right\rangle\right|}{\|v^t\|_2 \|x^*\|_2} = \eta(\overline{\Sigma}, \mathcal{T}, \delta) > 0,$$

where $\eta(\overline{\Sigma}, \mathcal{T}, \delta)$ can be derived from the state evolution.

We are done!

Bonus: Opt spec threshold & preprocessing function

Morally, $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$ is equivalent to $\delta > \delta^*(\overline{\Sigma}, \mathcal{T})$.

Bonus: Opt spec threshold & preprocessing function

Morally, $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$ is equivalent to $\delta > \delta^*(\overline{\Sigma}, \mathcal{T})$.

Q: which $\mathcal{T}$ yields the minimal $\delta^*(\overline{\Sigma}, \mathcal{T})$?

Bonus: Opt spec threshold & preprocessing function

Morally, $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$ is equivalent to $\delta > \delta^*(\overline{\Sigma}, \mathcal{T})$.
Q: which $\mathcal{T}$ yields the minimal $\delta^*(\overline{\Sigma}, \mathcal{T})$?

Theorem (Minimum threshold, optimal $\mathcal{T}^*$)

The optimal spectral threshold $\delta^*(\overline{\Sigma}) = \inf_{\mathcal{T}} \delta^*(\overline{\Sigma}, \mathcal{T})$ *is given by the solution to*

$$\delta^* = \frac{\mathbb{E}[\overline{\Sigma}]^2}{\mathbb{E}[\overline{\Sigma}^2]} \left(\int \frac{\mathbb{E}\left[p(y \,|\, \overline{G})\left(\frac{\delta^*}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2 - 1\right)\right]^2}{\mathbb{E}\left[p(y \,|\, \overline{G})\right]} \, \mathrm{d}y \right)^{-1},$$

and is obtained by

$$\mathcal{T}^*(y) = 1 - \frac{\mathbb{E}\left[p(y \,|\, \overline{G})\right]}{\mathbb{E}\left[p(y \,|\, \overline{G})\left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]}\overline{G}^2\right)\right]}, \quad \text{where } \overline{G} \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right).$$

Bonus: Opt spec threshold & preprocessing function

Morally, $\lambda_1(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2(\overline{\Sigma}, \mathcal{T}, \delta)$ is equivalent to $\delta > \delta^*(\overline{\Sigma}, \mathcal{T})$.
Q: which $\mathcal{T}$ yields the minimal $\delta^*(\overline{\Sigma}, \mathcal{T})$?

Theorem (Minimum threshold, optimal $\mathcal{T}^*$)

The optimal spectral threshold $\delta^*(\overline{\Sigma}) = \inf_{\mathcal{T}} \delta^*(\overline{\Sigma}, \mathcal{T})$ is given by the solution to

$$\delta^* = \frac{\mathbb{E}[\overline{\Sigma}]^2}{\mathbb{E}[\overline{\Sigma}^2]} \left(\int \frac{\mathbb{E}\left[p(y \,|\, \overline{G}) \left(\frac{\delta^*}{\mathbb{E}[\overline{\Sigma}]} \overline{G}^2 - 1 \right)\right]^2}{\mathbb{E}\left[p(y \,|\, \overline{G})\right]} \, \mathrm{d}y \right)^{-1},$$

and is obtained by

$$\mathcal{T}^*(y) = 1 - \frac{\mathbb{E}\left[p(y \,|\, \overline{G})\right]}{\mathbb{E}\left[p(y \,|\, \overline{G}) \left(\frac{\delta}{\mathbb{E}[\overline{\Sigma}]} \overline{G}^2 \right)\right]}, \quad \text{where } \overline{G} \sim \mathcal{N}\left(0, \frac{\mathbb{E}[\overline{\Sigma}]}{\delta}\right).$$

- $\mathcal{T}^*$ is impossible to guess

Reflection on opt threshold & preprocessor

- $\delta^*(\overline{\Sigma})$ depends on $\overline{\Sigma}$ only through its first two moments $\mathbb{E}[\overline{\Sigma}], \mathbb{E}[\overline{\Sigma}^2]$

Reflection on opt threshold & preprocessor

- $\delta^*(\overline{\Sigma})$ depends on $\overline{\Sigma}$ only through its first two moments $\mathbb{E}[\overline{\Sigma}], \mathbb{E}[\overline{\Sigma}^2]$
- Our results are expressed using $\mathrm{ESD}(\Sigma)$; equivalently, consider $\mathrm{ESD}(A^\top A)$:

$$\mathrm{ESD}(\Sigma) \to \mathrm{law}(\overline{\Sigma}),$$
$$\mathrm{ESD}(A^\top A) = \mathrm{ESD}(\Sigma^{1/2}\widetilde{A}^\top \widetilde{A}\Sigma^{1/2}) \to \mathsf{MP}_{1/\delta} \boxtimes \mathrm{law}(\overline{\Sigma}) =: \mathrm{law}(\overline{\Lambda})$$

Reflection on opt threshold & preprocessor

- $\delta^*(\overline{\Sigma})$ depends on $\overline{\Sigma}$ only through its first two moments $\mathbb{E}[\overline{\Sigma}], \mathbb{E}[\overline{\Sigma}^2]$
- Our results are expressed using $\mathrm{ESD}(\Sigma)$; equivalently, consider $\mathrm{ESD}(A^\top A)$:

$$\mathrm{ESD}(\Sigma) \to \mathrm{law}(\overline{\Sigma}),$$

$$\mathrm{ESD}(A^\top A) = \mathrm{ESD}(\Sigma^{1/2}\widetilde{A}^\top \widetilde{A}\Sigma^{1/2}) \to \mathsf{MP}_{1/\delta} \boxtimes \mathrm{law}(\overline{\Sigma}) =: \mathrm{law}(\overline{\Lambda})$$

Moments of $\overline{\Sigma}$ and $\overline{\Lambda}$ are related:

$$\mathbb{E}[\overline{\Lambda}] = \mathbb{E}[\overline{\Sigma}], \quad \mathbb{E}[\overline{\Lambda}^2] = \mathbb{E}[\overline{\Sigma}^2] + \frac{1}{\delta}\mathbb{E}[\overline{\Sigma}]^2$$

Reflection on opt threshold & preprocessor

- $\delta^*(\overline{\Sigma})$ depends on $\overline{\Sigma}$ only through its first two moments $\mathbb{E}[\overline{\Sigma}], \mathbb{E}[\overline{\Sigma}^2]$
- Our results are expressed using $\mathrm{ESD}(\Sigma)$; equivalently, consider $\mathrm{ESD}(A^\top A)$:

$$\mathrm{ESD}(\Sigma) \to \mathrm{law}(\overline{\Sigma}),$$

$$\mathrm{ESD}(A^\top A) = \mathrm{ESD}(\Sigma^{1/2}\widetilde{A}^\top \widetilde{A}\Sigma^{1/2}) \to \mathsf{MP}_{1/\delta} \boxtimes \mathrm{law}(\overline{\Sigma}) =: \mathrm{law}(\overline{\Lambda})$$

Moments of $\overline{\Sigma}$ and $\overline{\Lambda}$ are related:

$$\mathbb{E}[\overline{\Lambda}] = \mathbb{E}[\overline{\Sigma}], \quad \mathbb{E}[\overline{\Lambda}^2] = \mathbb{E}[\overline{\Sigma}^2] + \frac{1}{\delta}\mathbb{E}[\overline{\Sigma}]^2$$

Optimal spectral threshold $\delta^*(\overline{\Lambda})$ can be written as the solution δ^* to

$$\delta^* = \frac{\mathbb{E}[\overline{\Lambda}]^2}{\mathbb{E}[\overline{\Lambda}^2]}\left[\mathbf{1} + \left(\int \frac{\mathbb{E}\left[p(y \mid \overline{G})\left(\frac{\delta^*}{\mathbb{E}[\overline{\Lambda}]}\overline{G}^2 - 1\right)\right]^2}{\mathbb{E}\left[p(y \mid \overline{G})\right]} \mathrm{d}y\right)^{-1}\right]$$

Reflection on opt threshold & preprocessor

- $\delta^*(\overline{\Sigma})$ depends on $\overline{\Sigma}$ only through its first two moments $\mathbb{E}[\overline{\Sigma}], \mathbb{E}[\overline{\Sigma}^2]$
- Our results are expressed using $\mathrm{ESD}(\Sigma)$; equivalently, consider $\mathrm{ESD}(A^\top A)$:

$$\mathrm{ESD}(\Sigma) \to \mathrm{law}(\overline{\Sigma}),$$
$$\mathrm{ESD}(A^\top A) = \mathrm{ESD}(\Sigma^{1/2}\widetilde{A}^\top \widetilde{A}\Sigma^{1/2}) \to \mathsf{MP}_{1/\delta} \boxtimes \mathrm{law}(\overline{\Sigma}) =: \mathrm{law}(\overline{\Lambda})$$

Moments of $\overline{\Sigma}$ and $\overline{\Lambda}$ are related:

$$\mathbb{E}[\overline{\Lambda}] = \mathbb{E}[\overline{\Sigma}], \quad \mathbb{E}[\overline{\Lambda}^2] = \mathbb{E}[\overline{\Sigma}^2] + \frac{1}{\delta}\mathbb{E}[\overline{\Sigma}]^2$$

Optimal spectral threshold $\delta^*(\overline{\Lambda})$ can be written as the solution δ^* to

$$\delta^* = \frac{\mathbb{E}[\overline{\Lambda}]^2}{\mathbb{E}[\overline{\Lambda}^2]}\left[\mathbf{1} + \left(\int \frac{\mathbb{E}\left[p(y \mid \overline{G})\left(\frac{\delta^*}{\mathbb{E}[\overline{\Lambda}]}\overline{G}^2 - 1\right)\right]^2}{\mathbb{E}[p(y \mid \overline{G})]}\,\mathrm{d}y\right)^{-1}\right]$$

The formulas of $\delta^*, \mathcal{T}^*$ are **precisely the same** as those in [MLKZ20, MKLZ20].

Reflection on opt threshold & preprocessor (cont'd)

- $\mathcal{T}^*$ and δ^* depend on $p(\cdot \mid \cdot)$ only through its 0th & 2nd Hermite coefficients:

$$\sigma_0(y) = \mathbb{E}\left[p\left(y \middle| \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\, \overline{W} \right) \right], \quad \sigma_2(y) = \mathbb{E}\left[p\left(y \middle| \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\, \overline{W} \right) \left(\overline{W}^2 - 1 \right) \right],$$

where $\overline{W} \sim \mathcal{N}(0,1)$

Reflection on opt threshold & preprocessor (cont'd)

- $\mathcal{T}^*$ and δ^* depend on $p(\cdot \mid \cdot)$ only through its 0th & 2nd Hermite coefficients:

$$\sigma_0(y) = \mathbb{E}\left[p\left(y \,\middle|\, \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\,\overline{W} \right) \right], \quad \sigma_2(y) = \mathbb{E}\left[p\left(y \,\middle|\, \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\,\overline{W} \right) \left(\overline{W}^2 - 1 \right) \right],$$

  where $\overline{W} \sim \mathcal{N}(0,1)$
  - Information exponent [BAGJ21]

Reflection on opt threshold & preprocessor (cont'd)

- $\mathcal{T}^*$ and δ^* depend on $p(\cdot \mid \cdot)$ only through its 0th & 2nd Hermite coefficients:

$$\sigma_0(y) = \mathbb{E}\left[p\left(y \,\middle|\, \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\,\overline{W} \right) \right], \quad \sigma_2(y) = \mathbb{E}\left[p\left(y \,\middle|\, \sqrt{\frac{\mathbb{E}[\overline{\Sigma}]}{\delta}}\,\overline{W} \right) \left(\overline{W}^2 - 1 \right) \right],$$

  where $\overline{W} \sim \mathcal{N}(0,1)$
  - Information exponent [BAGJ21]
  - Leap complexity [ABM23] (See Boix's talk)

Spectral estimator with known Σ

We assumed Σ is unknown.

Spectral estimator with known Σ

We assumed Σ is unknown. What if it is **known**?

Spectral estimator with known Σ

We assumed Σ is unknown. What if it is **known**? A naive spectral estimator is to first 'whiten' A:

$$D_\triangle = \Sigma^{-1/2} A^\top T A \Sigma^{-1/2} = \widetilde{A}^\top T \widetilde{A},$$

$$T = \mathrm{diag}(\mathcal{T}(y)), \quad y = q(\widetilde{A}\Sigma^{1/2} x^*, \varepsilon)$$

and then output

$$\widehat{x}_\triangle^{\mathrm{spec}}(\mathcal{T}) = v_1(D_\triangle) \in \mathbb{S}^{d-1}$$

Spectral estimator with known Σ

We assumed Σ is unknown. What if it is **known**? A naive spectral estimator is to first 'whiten' A:

$$D_{\triangle} = \Sigma^{-1/2} A^\top T A \Sigma^{-1/2} = \widetilde{A}^\top T \widetilde{A},$$

$$T = \operatorname{diag}(\mathcal{T}(y)), \quad y = q(\widetilde{A}\Sigma^{1/2} x^*, \varepsilon)$$

and then output

$$\widehat{x}_{\triangle}^{\mathrm{spec}}(\mathcal{T}) = v_1(D_{\triangle}) \in \mathbb{S}^{d-1}$$

Its overlap can be similarly derived: if $\lambda_1^{\triangle}(\overline{\Sigma}, \mathcal{T}, \delta) > \lambda_2^{\triangle}(\overline{\Sigma}, \mathcal{T}, \delta)$

$$\frac{|\langle v_1(D_{\triangle}), x^* \rangle|}{\|v_1(D_{\triangle})\|_2 \|x^*\|_2} \to \eta_{\triangle}(\overline{\Sigma}, \mathcal{T}, \delta) > 0$$

Spectral estimator with $\Sigma = I_d$

With $\Sigma = I_d$ (i.e., i.i.d. Gaussian design), the spectral estimator becomes

$$\widehat{x}_{\square}^{\mathrm{spec}}(\mathcal{T}) = v_1(D_{\square}) \in \mathbb{S}^{d-1},$$
$$D_{\square} = \widetilde{A}^{\top} T_{\square} \widetilde{A}, \quad T_{\square} = \mathrm{diag}(\mathcal{T}(y_{\square})), \quad y_{\square} = p(\widetilde{A} x^*, \varepsilon).$$

Its overlap has been understood [LL20, MM19]

$$\frac{|\langle v_1(D_{\square}), x^* \rangle|}{\|v_1(D_{\square})\|_2 \|x^*\|_2} \to \eta_{\square}(\mathcal{T}, \delta)$$

- Phase transition

Outlook

Outlook

- Only "half" (supercritical part) of a phase transition phenomenon **:-(**

Outlook

- Only "half" (supercritical part) of a phase transition phenomenon **:-(**
  - RMT proof?

Outlook

- Only "half" (supercritical part) of a phase transition phenomenon **:-(**
  - RMT proof?
- More applications of this methodology

Outlook

- Only “half” (supercritical part) of a phase transition phenomenon **:-(**
  - RMT proof?
- More applications of this methodology
  - Mixture of GLMs with **correlated** signals (in prep **:-I**)

Outlook

- Only "half" (supercritical part) of a phase transition phenomenon **:-(**
    - RMT proof?
- More applications of this methodology
    - Mixture of GLMs with **correlated** signals (in prep **:-I**)
    - Low rank matrix estimation with heteroscedastic noise (in prep **:-I**)

Outlook

- Only “half” (supercritical part) of a phase transition phenomenon **:-(**
  - RMT proof?
- More applications of this methodology
  - Mixture of GLMs with **correlated** signals (in prep **:-|**)
  - Low rank matrix estimation with heteroscedastic noise (in prep **:-|**)
- Miraculous coincidence with physics prediction for **right** invariant ensemble [MLKZ20, MKLZ20]. **Universality**? (See Dudeja’s talk)

Outlook

- Only "half" (supercritical part) of a phase transition phenomenon :-(
    - RMT proof?
- More applications of this methodology
    - Mixture of GLMs with **correlated** signals (in prep :-|)
    - Low rank matrix estimation with heteroscedastic noise (in prep :-|)
- Miraculous coincidence with physics prediction for **right** invariant ensemble [MLKZ20, MKLZ20]. **Universality**? (See Dudeja's talk)
- **IT** limit (i.e., **MMSE**) known for $A = \widetilde{A}B$ (where B arbitrary) [MLKZ20]. Is spectral estimator IT-opt in terms of threshold and/or overlap?

THANK YOU!

zephyr.z798@gmail.com

Emmanuel Abbe, Enric Boix-Adserà, and Theodor Misiakiewicz.
SGD learning on neural networks: leap complexity and saddle-to-saddle dynamics.
CoRR, abs/2302.11055, 2023.

Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath.
Online stochastic gradient descent on non-convex losses from high-dimensional inference.
J. Mach. Learn. Res., 22:Paper No. 106, 51, 2021.

Mohsen Bayati, Marc Lelarge, and Andrea Montanari.
Universality in polytope phase transitions and message passing algorithms.
Ann. Appl. Probab., 25(2):753–822, 2015.

Wei-Kuo Chen and Wai-Kit Lam.
Universality of approximate message passing algorithms.
Electron. J. Probab., 26:Paper No. 36, 44, 2021.

Rishabh Dudeja, Yue M Lu, and Subhabrata Sen.
Universality of approximate message passing with semi-random matrices.
arXiv preprint arXiv:2204.04281, 2022.

Yue M Lu and Gen Li.
Phase transitions of spectral initialization for high-dimensional non-convex estimation.
Information and Inference: A Journal of the IMA, 9(3):507–541, 2020.

Antoine Maillard, Florent Krzakala, Yue M Lu, and Lenka Zdeborová.
Construction of optimal spectral methods in phase retrieval.
arXiv preprint arXiv:2012.04524, 2020.

Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová.
Phase retrieval in high dimensions: Statistical and computational phase transitions.
arXiv preprint arXiv:2006.05228, 2020.

Marco Mondelli and Andrea Montanari.
Fundamental limits of weak recovery with applications to phase retrieval.
Found. Comput. Math., 19(3):703–773, 2019.

Tianhao Wang, Xinyi Zhong, and Zhou Fan.
Universality of approximate message passing algorithms and tensor networks.
CoRR, abs/2206.13037, 2022.