Nonlinear Laplacians:
Tunable principal component analysis under directional prior information

Yuxin Ma, Dmitriy Kunisky

NEURAL INFORMATION PROCESSING SYSTEMS

JOHNS HOPKINS UNIVERSITY

NeurIPS

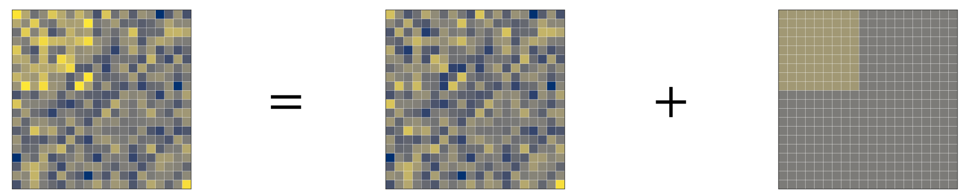
arXiv

Motivation

Gaussian planted submatrix problem

Randomly sample indices $S \subseteq [n] = \{1, 2, \dots, n\}$ with $|S| = \beta\sqrt{n}$, where $\beta \geq 0$ is a signal-to-noise parameter. Observe symmetric matrix $\boldsymbol{Y} = \boldsymbol{Y}^{(n)} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ with

$$\boldsymbol{Y} = \underbrace{(\text{i.i.d. } \mathcal{N}(0,1) \text{ noise})}_{=:\boldsymbol{W}} + \underbrace{\boldsymbol{1}_S \boldsymbol{1}_S^\top}_{\text{rank-one signal}}$$



Consider two tasks:

- **Detection**: Decide if $\boldsymbol{Y}$ is uniformly random ($\beta = 0$) or contains a signal ($\beta > 0$).
- **Recovery**: When $\beta > 0$, estimate the hidden subset S.

More generally, we consider $\boldsymbol{Y} = \boldsymbol{W} + \beta\sqrt{n}\boldsymbol{x}\boldsymbol{x}^\top$, where the signal vector $\boldsymbol{x} \in \mathbb{S}^{n-1}$ is sampled with directional prior information: $\langle \boldsymbol{x}, \boldsymbol{1} \rangle > 0$.

Naive spectral algorithm: directly perform PCA

Algorithm: Decide $\beta > 0$ if the top eigenvalue $\lambda_1(\boldsymbol{Y})$ is unusually large, and estimate $\boldsymbol{x}$ by the associated eigenvector $v_1(\boldsymbol{Y})$. It is imple and effective, yet agnostic to any prior information about the signal $\boldsymbol{x}$.

Question: Can we "boost" PCA by using the top eigenpair of a deformed matrix $\lambda_1(H(\boldsymbol{Y})), v_1(H(\boldsymbol{Y}))$? How to design the deformation H to tailor to prior information about the signal?

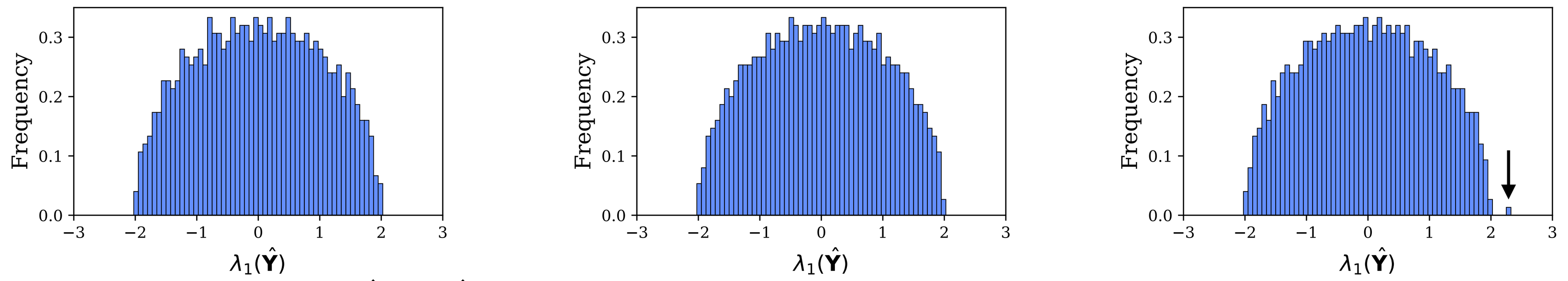
Background: Naive Spectral Algorithm

Let $\hat{\boldsymbol{Y}} = \frac{\boldsymbol{Y}}{\sqrt{n}}, \hat{\boldsymbol{W}} = \frac{\boldsymbol{W}}{\sqrt{n}}$. Then $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{W}} + \beta\boldsymbol{x}\boldsymbol{x}^\top$.

Theorem 1: Baik-Ben Arous-Péché (BBP) transition [FP07]

- If $\beta \leq \beta_* = 1$, then almost surely, $\lambda_1(\hat{\boldsymbol{Y}}^{(n)}) \to 2, \quad |\langle \boldsymbol{v}_1(\hat{\boldsymbol{Y}}^{(n)}), \boldsymbol{x}\rangle| \to 0.$
- If $\beta > \beta_* = 1$, then almost surely,

$$\lambda_1(\hat{\boldsymbol{Y}}^{(n)}) \to \beta + 1/\beta > 2, \quad |\langle \boldsymbol{v}_1(\hat{\boldsymbol{Y}}^{(n)}), \boldsymbol{x}\rangle|^2 \to 1 - 1/\beta^2 > 0.$$



$\beta = 0$: eigenvalues of $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{W}}$ follow **Wigner's semicircle law**. $0 < \beta \leq 1$: the signal is too weak to create an outlier eigenvalue. $\beta > 1$: the signal creates an outlier outside the bulk.

$$\boxed{\text{Nontrivial detection and recovery} \iff \beta > \beta_* = 1}$$

Nonlinear Laplacians Spectral Algorithm

Definition: σ-Laplacian spectral algorithm

σ-Laplacian matrix: $\boldsymbol{L} = \boldsymbol{L}_\sigma(\hat{\boldsymbol{Y}}) := \hat{\boldsymbol{Y}} + \mathrm{diag}(\sigma(\hat{\boldsymbol{Y}}\boldsymbol{1}))$ where the non-decreasing, bounded, Lipschitz function $\sigma : \mathbb{R} \to \mathbb{R}$ applies entrywise to the vector $\hat{\boldsymbol{Y}}\boldsymbol{1} \in \mathbb{R}^n$.

Algorithm: Decide $\beta > 0$ if the top eigenvalue $\lambda_1(\boldsymbol{L})$ is unusually large, and estimate $\boldsymbol{x}$ by the associated eigenvector $v_1(\boldsymbol{L})$.

Why does it work?

$$\boldsymbol{L} = \underbrace{\hat{\boldsymbol{W}}}_{\text{noise}} + \underbrace{\beta\boldsymbol{x}\boldsymbol{x}^\top + \mathrm{diag}(\sigma(\hat{\boldsymbol{W}}\boldsymbol{1} + \beta\langle \boldsymbol{x}, \boldsymbol{1}\rangle\boldsymbol{x}))}_{\text{new signal}},$$

- Under prior information $\langle \boldsymbol{x}, \boldsymbol{1}\rangle > 0$, we know that $\hat{\boldsymbol{Y}}\boldsymbol{1} = \beta\langle \boldsymbol{x}, \boldsymbol{1}\rangle\boldsymbol{x} + \hat{\boldsymbol{W}}\boldsymbol{1}$ is somewhat correlated with $\boldsymbol{x}$.
- With non-decreasing σ, the diagonal matrix becomes larger entrywise as β increases, effectively boosting the original signal $\beta\boldsymbol{x}\boldsymbol{x}^\top$.
- **Notice**: bounded σ is necessary, because $\|\hat{\boldsymbol{Y}}\| = O(1)$ while $\|\mathrm{diag}(\hat{\boldsymbol{Y}}\boldsymbol{1})\| = \|\hat{\boldsymbol{Y}}\boldsymbol{1}\|_\infty = \Theta(\sqrt{\log n})$.

Theoretical Results

Rigorous characterization of β_* as a function of σ

Theorem 2: BBP transition of the σ-Laplacian matrix, for Gaussian planted submatrix

A small variant $\tilde{\boldsymbol{L}}^{(n)}$ of the σ-Laplacian has a BBP transition around a different $\beta_* = \beta_*(\sigma)$. Define

$$G(z) = \beta \mathop{\mathbb{E}}_{g\sim\mathcal{N}(0,1)} (z - \sigma(\beta + g))^{-1}, \quad H(z) = z + \mathop{\mathbb{E}}_{g\sim\mathcal{N}(0,1)} (z - \sigma(g))^{-1}.$$

Let $\theta = \theta(\sigma)$ solves $G(\theta) = 1$ if such $\theta > \mathrm{edge}^+(\sigma)$ exists, and $\theta = \mathrm{edge}^+(\sigma)$ otherwise. Then, let $\beta_* = \beta_*(\sigma)$ solves $H'(\theta_\sigma(\beta_*)) = 0$.

- If $\beta \leq \beta_*$, then almost surely, $\lambda_1(\tilde{\boldsymbol{L}}^{(n)}) \to$ edge of bulk, $\quad |\langle \boldsymbol{v}_1(\tilde{\boldsymbol{L}}^{(n)}), \boldsymbol{x}\rangle| \to 0.$
- If $\beta > \beta_*$, then almost surely,

$$\lambda_1(\tilde{\boldsymbol{L}}^{(n)}) \to H(\theta_\sigma(\beta)) > \text{edge of bulk}, \quad |\langle \boldsymbol{v}_1(\tilde{\boldsymbol{L}}^{(n)}), \boldsymbol{x}\rangle|^2 \to -\beta^{-1}G'(\theta_\sigma(\beta))^{-1}H'(\theta_\sigma(\beta)) > 0.$$

Proof idea: $\boldsymbol{L}$ is a full-rank perturbation of the Gaussian random matrix $\hat{\boldsymbol{W}}$. This resembles the spiked matrix model studied in [CDMFF11] using the free probability theory.

Picking the nonlinearity σ that optimizes β_*

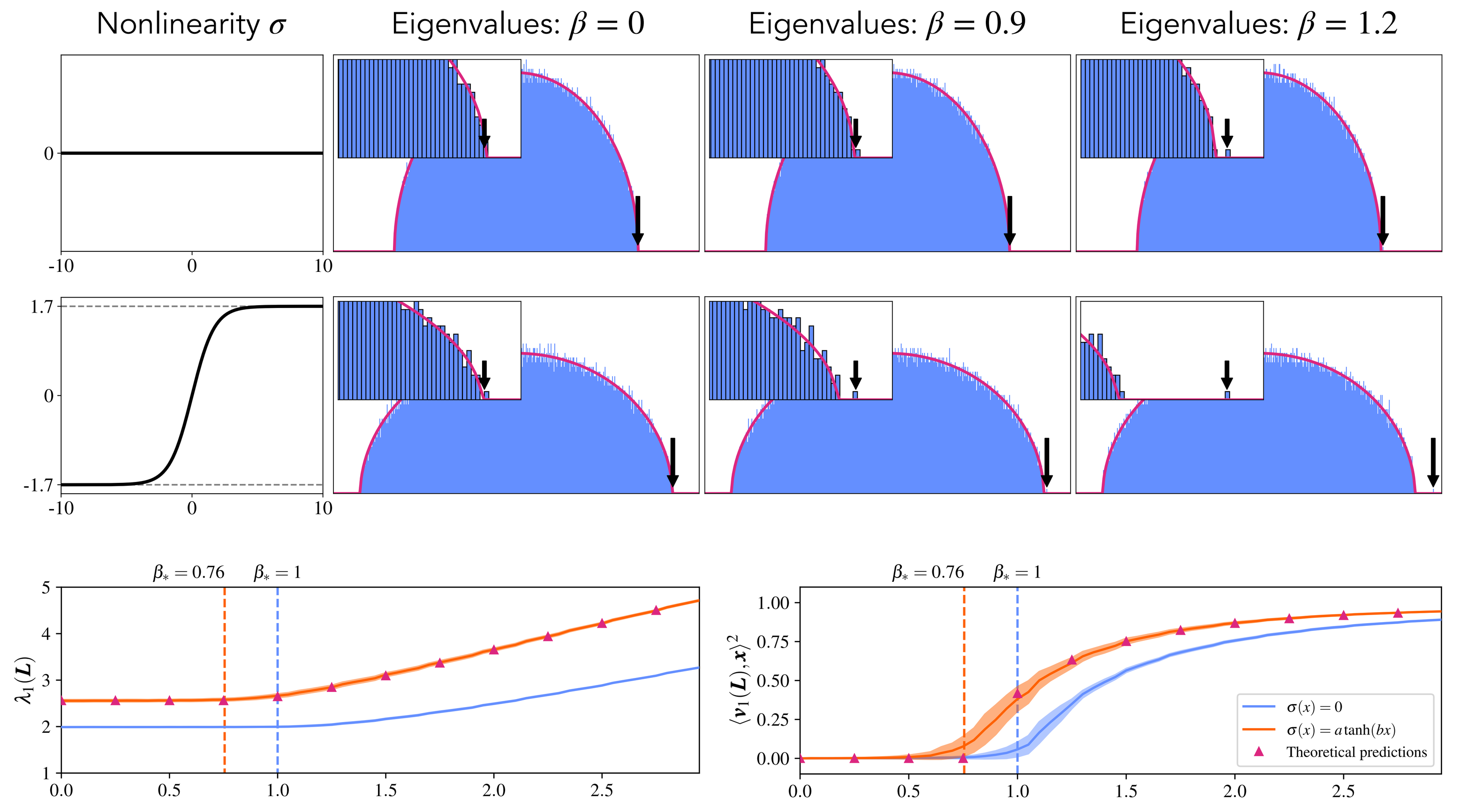
Corollary 3: σ-Laplacian spectral algorithm improves naive spectral algorithm

With a nonlinearity σ obtained by numerically optimizing $\beta_*(\sigma)$, the variant $\tilde{\boldsymbol{L}}$ of the σ-Laplacian has a BBP transition around $\beta_*(\sigma) \approx 0.76 < 1 = \beta_*(0)$.

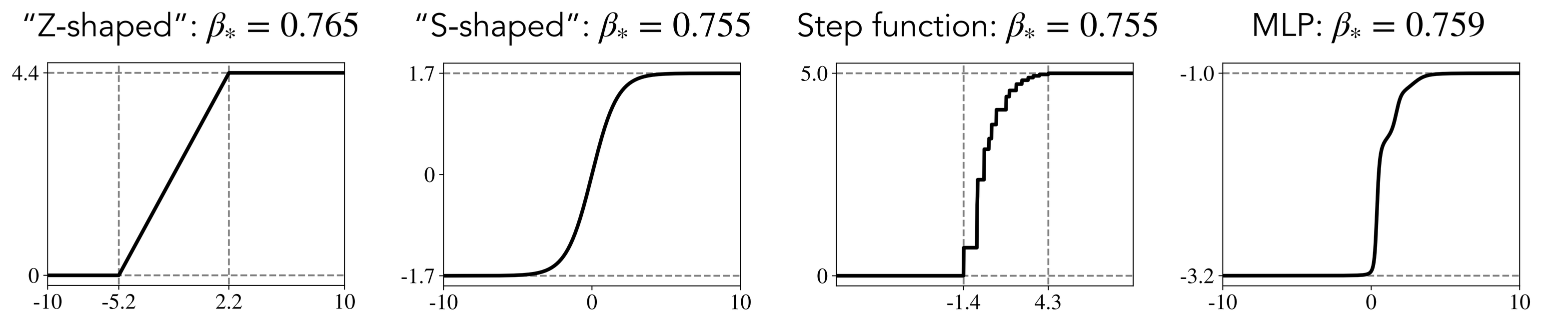
$$\boxed{\text{Nontrivial detection and recovery} \iff \beta > \beta_*(\sigma) \approx 0.76}$$

Numerical Results

BBP transition of σ-Laplacian v.s. naive spectral algorithm



Effective choices of nonlinearity



Conclusion and Future Work

Lessons learned

Spectral algorithms are a simple design pattern for tractable algorithms in high dimension. They can be boosted by:

- Combining with weak information.
- Optimizing over a simple restricted class of spectral algorithms.

Future directions

- **Generalized directional priors:** How to handle more general directional prior information, such as signals constrained to non-convex cones?
- **More flexible spectral algorithms:** Is it possible to optimize over a more flexible class of deformations $H : \mathbb{R}^{n\times n}_{\mathrm{sym}} \to \mathbb{R}^{n\times n}_{\mathrm{sym}}$ to achieve even better performance?

References

[FP07] Féral, Delphine, and Sandrine Péché. "The largest eigenvalue of rank one deformation of large Wigner matrices." Communications in Mathematical Physics 272.1 (2007): 185-228.

[CDMFF11] Capitaine, Mireille, et al. "Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices." Electron. J. Probab 16.64 (2011): 1750-1792.