Phase transitions in low-rank matrix estimation

May 11, 2017

Marc Lelarge & Léo Miolane
INRIA, ENS

Introduction

The statistical model

"Spiked Wigner" model

$$\underbrace{\mathbf{Y}}_{\text{observations}} = \sqrt{\frac{\lambda}{n}}\,\underbrace{\mathbf{X}\mathbf{X}^\intercal}_{\text{signal}} + \underbrace{\mathbf{Z}}_{\text{noise}}$$

- $\mathbf{X}$: vector of dimension n with entries $X_i \overset{\text{i.i.d.}}{\sim} P_0$. $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$.
- $Z_{i,j} = Z_{j,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.
- λ: signal-to-noise ratio.

Goal: recover the low-rank matrix $\mathbf{X}\mathbf{X}^\intercal$ from $\mathbf{Y}$.

Principal component analysis (PCA)

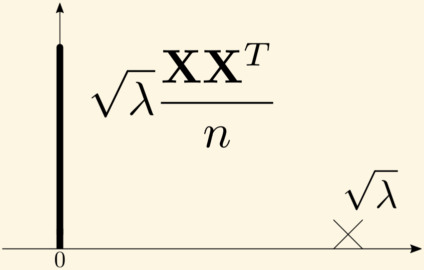
B.B.P. phase transition

- The matrix $\mathsf{Y}/\sqrt{n} = \sqrt{\lambda}\mathsf{X}\mathsf{X}^\intercal/n + \mathsf{Z}/\sqrt{n}$ is a perturbed low-rank matrix.
- Estimate X using the eigenvector $\hat{\mathsf{x}}_n$ associated with the largest eigenvalue μ_n of $\mathsf{Y}/\sqrt{n}$.

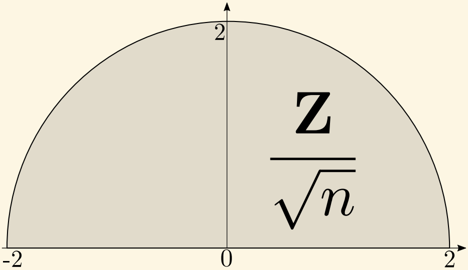
Principal component analysis (PCA)

B.B.P. phase transition

- The matrix $\mathsf{Y}/\sqrt{n} = \sqrt{\lambda}\mathsf{X}\mathsf{X}^\intercal/n + \mathsf{Z}/\sqrt{n}$ is a perturbed low-rank matrix.
- Estimate X using the eigenvector $\hat{\mathsf{x}}_n$ associated with the largest eigenvalue μ_n of $\mathsf{Y}/\sqrt{n}$.



Spectral density of the signal



Limiting spectral density of the noise

Principal component analysis (PCA)

B.B.P. phase transition

- The matrix $Y/\sqrt{n} = \sqrt{\lambda} X X^\intercal / n + Z/\sqrt{n}$ is a perturbed low-rank matrix.
- Estimate X using the eigenvector $\hat{x}_n$ associated with the largest eigenvalue μ_n of $Y/\sqrt{n}$.

B.B.P. phase transition

- if $\lambda \leq 1$ $\begin{cases} \mu_n & \longrightarrow 2 \\ X \cdot \hat{x}_n & \longrightarrow 0 \end{cases}$
- if $\lambda > 1$ $\begin{cases} \mu_n & \longrightarrow \sqrt{\lambda} + \frac{1}{\sqrt{\lambda}} > 2 \\ |X \cdot \hat{x}_n| & \longrightarrow \sqrt{1 - 1/\lambda} > 0 \end{cases}$

Baik et al., 2005; Benaych-Georges and Nadakuditi, 2011

Questions

- PCA fails when $\lambda \leq 1$, but is it still possible to recover the signal?

Questions

- PCA fails when $\lambda \leq 1$, but is it still possible to recover the signal?
- When $\lambda > 1$, is PCA optimal?

Questions

- PCA fails when $\lambda \leq 1$, but is it still possible to recover the signal?
- When $\lambda > 1$, is PCA optimal?
- More generally, what is the best achievable estimation performance in both regimes?

MMSE and information-theoretic threshold

Goal

$$\mathrm{MMSE}_n = \min_{\hat{\theta}} \frac{1}{n^2}\mathbb{E}\left\|\mathbf{X}\mathbf{X}^\intercal - \hat{\theta}(\mathbf{Y})\right\|^2$$
$$= \frac{1}{n^2}\sum_{1\le i,j\le n}(X_iX_j - \mathbb{E}[X_iX_j|\mathbf{Y}])^2 \le \underbrace{\mathbb{E}[X^2]^2}_{\text{Dummy MSE}}$$

MMSE and information-theoretic threshold

Goal

$$\mathrm{MMSE}_n = \min_{\hat\theta} \frac{1}{n^2}\mathbb{E}\left\|\mathbf{X}\mathbf{X}^\intercal - \hat\theta(\mathbf{Y})\right\|^2$$

$$= \frac{1}{n^2}\sum_{1\le i,j\le n}(X_iX_j - \mathbb{E}[X_iX_j|\mathbf{Y}])^2 \le \underbrace{\mathbb{E}[X^2]^2}_{\text{Dummy MSE}}$$

Information-theoretic threshold

1. Compute $\lim_{n\to\infty} \mathrm{MMSE}_n$
2. Deduce the information-theoretic threshold, i.e. the critical value λ_c such that
   - if $\lambda > \lambda_c$, $\lim_{n\to\infty} \mathrm{MMSE}_n <$ Dummy MSE
   - if $\lambda < \lambda_c$, $\lim_{n\to\infty} \mathrm{MMSE}_n =$ Dummy MSE

Connection with statistical physics

A planted spin glass model

- Compute the MMSE for $Y = \sqrt{\frac{\lambda}{n}} X X^{\intercal} + Z$

Connection with statistical physics

A planted spin glass model

- Compute the MMSE for $Y = \sqrt{\frac{\lambda}{n}} X X^\intercal + Z$
- Study the posterior $\mathbb{P}(x \mid Y) = \frac{1}{Z_n} P_0(x) \exp(H_n(x))$ where

$$H_n(\mathbf{x}) = \sum_{i<j} \sqrt{\frac{\lambda}{n}} Y_{i,j} x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2$$

$$= \sum_{i<j} \underbrace{\sqrt{\frac{\lambda}{n}} Z_{i,j} x_i x_j}_{\text{SK}} + \underbrace{\frac{\lambda}{n} X_i X_j x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2}_{\text{planted solution}}$$

Connection with statistical physics

A planted spin glass model

- Compute the MMSE for $\mathrm{Y} = \sqrt{\frac{\lambda}{n}}\mathrm{X}\mathrm{X}^\intercal + \mathrm{Z}$
- Study the posterior $\mathbb{P}(\mathrm{x} \mid \mathrm{Y}) = \frac{1}{Z_n} P_0(\mathrm{x}) \exp(H_n(\mathrm{x}))$ where

$$H_n(\mathbf{x}) = \sum_{i<j} \sqrt{\frac{\lambda}{n}} Y_{i,j} x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2$$
$$= \sum_{i<j} \underbrace{\sqrt{\frac{\lambda}{n}} Z_{i,j} x_i x_j}_{\text{SK}} + \underbrace{\frac{\lambda}{n} X_i X_j x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2}_{\text{planted solution}}$$

- Compute the limit of the free energy $F_n = \frac{1}{n}\mathbb{E} \log Z_n$ because

$$\text{Constant} - F_n = \frac{1}{n} I(\mathbf{X}; \mathbf{Y}) \xrightarrow{\partial \lambda} \text{MMSE}$$

Replica symmetric formula

The scalar channel

Lesieur et al., 2015 conjectured that the problem is characterized par the scalar channel:

$$Y_0 = \sqrt{\gamma} X_0 + Z_0$$

and the scalar free energy: $\mathcal{F}(\gamma) = \mathbb{E}\left[\log \sum_{x_0} P_0(x_0) e^{\sqrt{\gamma} Y_0 x_0 - \frac{\gamma}{2} x_0^2}\right]$

Replica symmetric formula

The scalar channel

Lesieur et al., 2015 conjectured that the problem is characterized par the scalar channel:

$$Y_0 = \sqrt{\gamma} X_0 + Z_0$$

and the scalar free energy: $\mathcal{F}(\gamma) = \mathbb{E}\left[\log \sum_{x_0} P_0(x_0) e^{\sqrt{\gamma} Y_0 x_0 - \frac{\gamma}{2} x_0^2}\right]$

Replica symmetric formula

$$F_n \xrightarrow[n\to\infty]{} \sup_{q \geq 0} \mathcal{F}(\lambda q) - \frac{\lambda}{4} q^2$$

$$\mathrm{MMSE}_n \xrightarrow[n\to\infty]{} \mathbb{E}_{P_0}[X^2]^2 - q^*(\lambda)^2$$

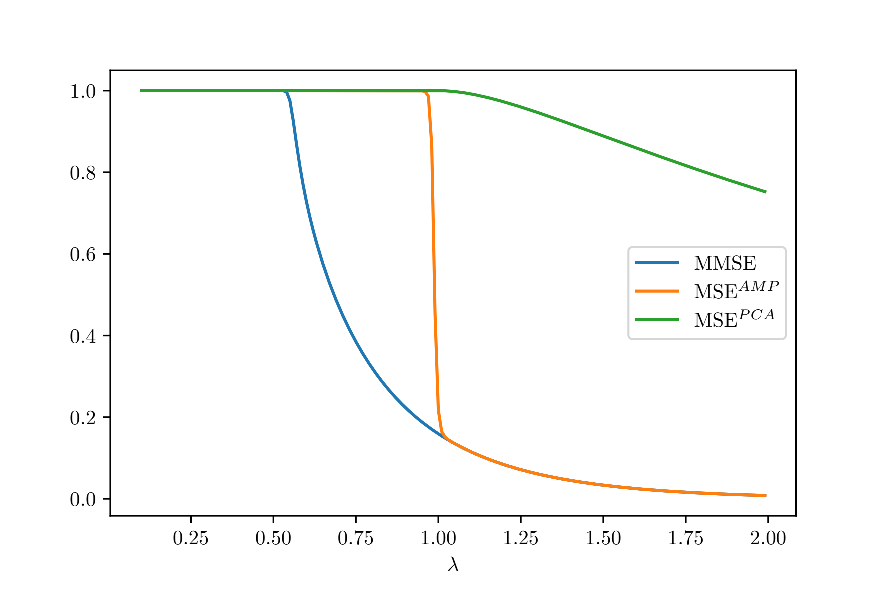
Proved by Barbier et al., 2016, extended by Lelarge and Miolane, 2016.

Some curves

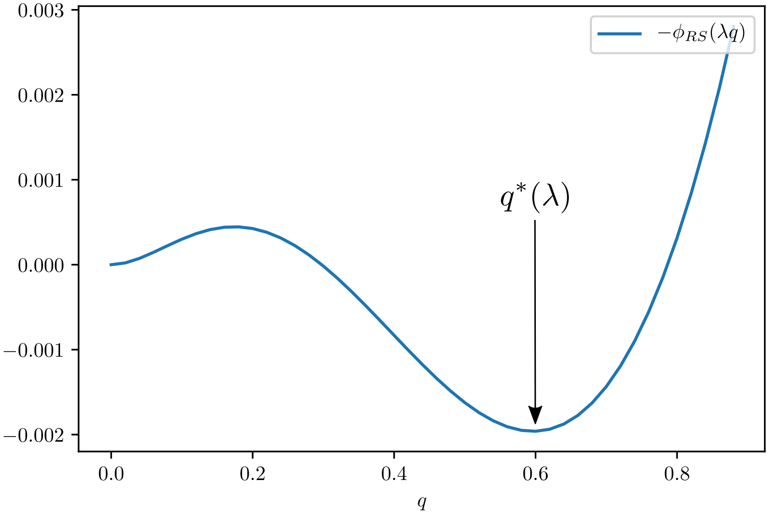
- We will plot the MMSE and $\mathrm{MSE}^{\mathrm{PCA}}$ curves when P_0 is of the form

$$\begin{cases} P_0(\sqrt{(1-p)/p}) & = \; p \\ P_0(-\sqrt{p/(1-p)}) & = \; 1-p \end{cases}$$

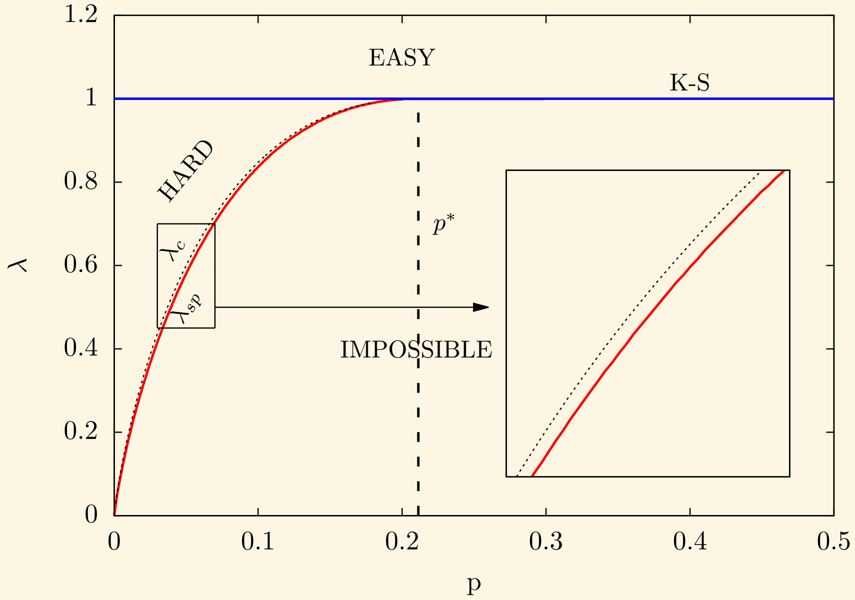
  for some $p \in (0,1)$.
- One can show that the corresponding matrix estimation problem is, in some sense, equivalent to the community detection problem with 2 asymmetric communities.



MMSE, MSE^{PCA} and MSE^{AMP}, asymmetric SBM: $p = 0.05$.



"Free energy lanscape", $p = 0.05$, $\lambda = 0.63$.



Phase diagram from Caltagirone et al., 2016

Thank you for your attention.

Any questions?

References I

- Baik, Jinho, Gérard Ben Arous, and Sandrine Péché (2005). "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *Annals of Probability*, pp. 1643–1697.
- Barbier, Jean et al. (2016). "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula". In: *Advances in Neural Information Processing Systems*, pp. 424–432.
- Benaych-Georges, Florent and Raj Rao Nadakuditi (2011). "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices". In: *Advances in Mathematics* 227.1, pp. 494–521.
- Caltagirone, Francesco, Marc Lelarge, and Léo Miolane (2016). "Recovering asymmetric communities in the stochastic block model". In: *arXiv preprint arXiv:1610.03680*.
- Lelarge, Marc and Léo Miolane (2016). "Fundamental limits of symmetric low-rank matrix estimation". In: *arXiv preprint arXiv:1611.03888*.

References II

- Lesieur, Thibault, Florent Krzakala, and Lenka Zdeborová (2015). "MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel". In: *53rd Annual Allerton Conference on Communication, Control, and Computing, Allerton 2015, Allerton Park & Retreat Center, Monticello, IL, USA, September 29 - October 2, 2015*. IEEE, pp. 680–687. ISBN: 978-1-5090-1824-6. DOI: 10.1109/ALLERTON.2015.7447070. URL: http://dx.doi.org/10.1109/ALLERTON.2015.7447070.