Asymmetric Estimation of Low Rank Matrix: Statistical and Computational Limits

Yuchen Wu

Department of Statistics, Stanford University

8th August 2021

Professor Andrea Montanari
Stanford Stat & EE

A fundamental question

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^\intercal + \boldsymbol{Z} \in \mathbb{R}^{n\times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n\times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d\times r},$$

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n\times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n\times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d\times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda} \boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n \times r}, \; \boldsymbol{\Theta} \in \mathbb{R}^{d \times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^\intercal + \boldsymbol{Z} \in \mathbb{R}^{n\times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n\times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d\times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.
- Estimate $\boldsymbol{\Lambda}, \boldsymbol{\Theta}$.

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda} \boldsymbol{\Theta}^\mathsf{T} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n \times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d \times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.
- Estimate $\boldsymbol{\Lambda}, \boldsymbol{\Theta}$.

Study three problems

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda} \boldsymbol{\Theta}^\mathsf{T} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n \times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d \times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.
- Estimate $\boldsymbol{\Lambda}, \boldsymbol{\Theta}$.

Study three problems

- Information-theoretically possible?

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n\times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n\times r}, \ \boldsymbol{\Theta} \in \mathbb{R}^{d\times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.
- Estimate $\boldsymbol{\Lambda}, \boldsymbol{\Theta}$.

Study three problems

- Information-theoretically possible?
- Closed form MMSE?

A fundamental question

$$\boldsymbol{A} = s_n \boldsymbol{\Lambda} \boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}, \quad Z_{ij} \overset{iid}{\sim} \mathsf{N}(0,1), \quad \boldsymbol{\Lambda} \in \mathbb{R}^{n \times r}, \; \boldsymbol{\Theta} \in \mathbb{R}^{d \times r},$$

- r fixed, $n, d \to \infty$, $d/n \to \infty$.
- Λ_i: i-th row of $\boldsymbol{\Lambda}$, Θ_j: j-th row of $\boldsymbol{\Theta}$. $\Lambda_i \overset{iid}{\sim} \mu_\Lambda$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$.
- Estimate $\boldsymbol{\Lambda}, \boldsymbol{\Theta}$.

Study three problems

- Information-theoretically possible?
- Closed form MMSE?
- Computationally efficient?

Background

Background

- Statistical physics: $n/d \to \delta \in (0, \infty)$ heuristically computed optimal estimation error. (e.g. Lesieur et al., 2017)

Background

- Statistical physics: $n/d \to \delta \in (0, \infty)$ heuristically computed optimal estimation error. (e.g. Lesieur et al., 2017)

- Rigorous justification, $n/d \to \delta \in (0, \infty)$. (Lelarge and Miolane, 2019; Miolane, 2017)

Background

- Statistical physics: $n/d \to \delta \in (0, \infty)$ heuristically computed optimal estimation error. (e.g. Lesieur et al., 2017)
- Rigorous justification, $n/d \to \delta \in (0, \infty)$. (Lelarge and Miolane, 2019; Miolane, 2017)
- Approximate message passing (AMP) conjectured optimal among polynomial algorithms, $n/d \to \delta \in (0, \infty)$. (Montanari and Venkataramanan, 2021)

Overview of our results

Overview of our results

- $d/n \to \infty$.

Overview of our results

- $d/n \to \infty$.
- Introduce a new class of algorithms, characterize the asymptotic behavior of proposed algorithms.

Overview of our results

- $d/n \to \infty$.
- Introduce a new class of algorithms, characterize the asymptotic behavior of proposed algorithms.
- Consider various scalings of s_n, $\quad \boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^\intercal + \boldsymbol{Z}$.

Overview of our results

- $d/n \to \infty$.
- Introduce a new class of algorithms, characterize the asymptotic behavior of proposed algorithms.
- Consider various scalings of s_n, $\boldsymbol{A} = s_n \boldsymbol{\Lambda} \boldsymbol{\Theta}^\intercal + \boldsymbol{Z}$.
    - Strong signal regime: $s_n = 1/\sqrt{n}$.

Overview of our results

- $d/n \to \infty$.
- Introduce a new class of algorithms, characterize the asymptotic behavior of proposed algorithms.
- Consider various scalings of s_n, $\quad \boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^\mathsf{T} + \boldsymbol{Z}$.
  - Strong signal regime: $\quad s_n = 1/\sqrt{n}$.
  - Weak signal regime: $\quad s_n = 1/\sqrt[4]{nd}$.

Overview of our results

- $d/n \to \infty$.
- Introduce a new class of algorithms, characterize the asymptotic behavior of proposed algorithms.
- Consider various scalings of s_n, $\quad \boldsymbol{A} = s_n \boldsymbol{\Lambda}\boldsymbol{\Theta}^\mathsf{T} + \boldsymbol{Z}$.
  - Strong signal regime: $s_n = 1/\sqrt{n}$.
  - Weak signal regime: $s_n = 1/\sqrt[4]{nd}$.
- Application: Gaussian mixture clustering.

For simplicity, in this talk will focus on $r = 1$, will put cases with $r > 1$ in our manuscript.

Strong signal regime: estimating Λ

Strong signal regime: estimating $\boldsymbol{\Lambda}$

$\rho_\Theta = \mathbb{E}[\Theta^2]$, $\rho_\Lambda = \mathbb{E}[\Lambda^2]$

Strong signal regime: estimating $\boldsymbol{\Lambda}$

$\rho_\Theta = \mathbb{E}[\Theta^2]$, $\rho_\Lambda = \mathbb{E}[\Lambda^2]$

$$\boldsymbol{A} = \frac{1}{\sqrt{n}} \boldsymbol{\Lambda}^T \boldsymbol{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

Strong signal regime: estimating $\mathbf{\Lambda}$

$\rho_\Theta = \mathbb{E}[\Theta^2]$, $\rho_\Lambda = \mathbb{E}[\Lambda^2]$

$$\boldsymbol{A} = \frac{1}{\sqrt{n}} \mathbf{\Lambda}^T \mathbf{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

- $\hat{\mathbf{\Lambda}}_s \in \mathbb{R}^n$: top left singular vectors of $\boldsymbol{A}$, $\|\hat{\mathbf{\Lambda}}_s\|_2 = \sqrt{n}$, $\frac{1}{n}\langle \mathbf{\Lambda}, \hat{\mathbf{\Lambda}}_s \rangle > 0$.

Strong signal regime: estimating $\mathbf{\Lambda}$

$\rho_\Theta = \mathbb{E}[\Theta^2]$, $\rho_\Lambda = \mathbb{E}[\Lambda^2]$

$$\boldsymbol{A} = \frac{1}{\sqrt{n}} \mathbf{\Lambda}^T \mathbf{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

- $\hat{\mathbf{\Lambda}}_s \in \mathbb{R}^n$: top left singular vectors of $\boldsymbol{A}$, $\|\hat{\mathbf{\Lambda}}_s\|_2 = \sqrt{n}$, $\frac{1}{n}\langle \mathbf{\Lambda}, \hat{\mathbf{\Lambda}}_s \rangle > 0$.
- $\|\rho_\Lambda^{1/2} \hat{\mathbf{\Lambda}}_s - \mathbf{\Lambda}\|_2^2 / n \xrightarrow{P} 0$.

Strong signal regime: estimating Θ

Strong signal regime: estimating $\mathbf{\Theta}$

Suppose $\mathbf{\Lambda}$ is known,

Strong signal regime: estimating $\boldsymbol{\Theta}$

Suppose $\boldsymbol{\Lambda}$ is known,

$$p(\theta_j|\boldsymbol{\Lambda}, \boldsymbol{A}) \propto \exp\left(-\frac{1}{2n}\sum_{i=1}^{n}\Lambda_i^2\theta_j^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_{ij}\Lambda_i\theta_j\right)\mu_\Theta \mathrm{d}(\theta_j).$$

Strong signal regime: estimating $\boldsymbol{\Theta}$

Suppose $\boldsymbol{\Lambda}$ is known,

$$p(\theta_j|\boldsymbol{\Lambda}, \boldsymbol{A}) \propto \exp\left(-\frac{1}{2n}\sum_{i=1}^{n}\Lambda_i^2\theta_j^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_{ij}\Lambda_i\theta_j\right)\mu_\Theta \mathrm{d}(\theta_j).$$

Theorem 1

Let $G \overset{iid}{\sim} \mathsf{N}(0,1)$, then for any $\hat{\boldsymbol{\Theta}} \in \mathbb{R}^n$ being an estimator of $\boldsymbol{\Theta}$ based on observation $\boldsymbol{A}$, we have

$$\liminf_{n,d\to\infty} \mathbb{E}\left[\frac{1}{d}\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_2^2\right] \geq \mathbb{E}\left[\left(\Theta - \mathbb{E}[\Theta|\rho_\Lambda^{1/2}\Theta + G]\right)^2\right],$$

Strong signal regime: estimating Θ

Strong signal regime: estimating $\mathbf{\Theta}$

- $(\hat{\mathbf{\Theta}}_s, \hat{\mathbf{\Lambda}}_s)$: top singular vectors pair of $\boldsymbol{A}$, norm being $\sqrt{d}$ and $\sqrt{n}$.

Strong signal regime: estimating $\mathbf{\Theta}$

- $(\hat{\mathbf{\Theta}}_s, \hat{\mathbf{\Lambda}}_s)$: top singular vectors pair of $\boldsymbol{A}$, norm being $\sqrt{d}$ and $\sqrt{n}$.
- $F(y) \triangleq \mathbb{E}\left[\Theta \mid \frac{\rho_\Lambda^{1/2}\Theta+G}{\sqrt{1+\rho_\Theta\rho_\Lambda}} = y\right]$, $\hat{\mathbf{\Theta}} = F(\hat{\mathbf{\Theta}}_s)$.

Theorem 2

Assume F is Lipschitz continuous, $\langle \mathbf{\Theta}, \hat{\mathbf{\Theta}}_s \rangle > 0$. Then $\hat{\mathbf{\Theta}}$ is an estimator which achieves the lower bound proposed in Theorem 1:

$$\frac{1}{d}\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2^2 \xrightarrow{P} \mathbb{E}[\Theta^2] - \mathbb{E}\left[\mathbb{E}[\Theta|\rho_\Lambda^{1/2}\Theta + G]^2\right].$$

Weak signal regime: estimating Θ

Weak signal regime: estimating $\mathbf{\Theta}$

$$\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \mathbf{\Lambda}^T \mathbf{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

Weak signal regime: estimating $\mathbf{\Theta}$

$$\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda}^T \boldsymbol{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

Theorem 3

For any $\hat{\boldsymbol{\Theta}} \in \mathbb{R}^d$ being an estimator of $\boldsymbol{\Theta}$ based on observation $\boldsymbol{A}$,

$$\liminf_{n,d \to \infty} \mathbb{E}[\frac{1}{d} \| \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \|_F^2] \geq \mathrm{Var}[\Theta].$$

The above lower bound is achieved by the naive estimator $\hat{\boldsymbol{\Theta}} = \vec{1}_d \mathbb{E}[\Theta] \in \mathbb{R}^d$.

Weak signal regime: estimating $\mathbf{\Theta}$

$$\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda}^T \mathbf{\Theta} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}.$$

Theorem 3

For any $\hat{\mathbf{\Theta}} \in \mathbb{R}^d$ being an estimator of $\mathbf{\Theta}$ based on observation $\boldsymbol{A}$,

$$\liminf_{n,d \to \infty} \mathbb{E}[\frac{1}{d} \|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\|_F^2] \geq \mathrm{Var}[\Theta].$$

The above lower bound is achieved by the naive estimator $\hat{\mathbf{\Theta}} = \vec{1}_d \mathbb{E}[\Theta] \in \mathbb{R}^d$.

Proof: Suppose we know $\boldsymbol{\Lambda}$, compute the minimum MSE.

Information-theoretic lower bound on $\mathbf{\Lambda}$, $s_n = 1/\sqrt[4]{nd}$

Information-theoretic lower bound on $\boldsymbol{\Lambda}$, $s_n = 1/\sqrt[4]{nd}$

$$\mathsf{MMSE}_n(\boldsymbol{A}) = \min_{\hat{M}(\boldsymbol{A})} \frac{1}{n^2} \mathbb{E}\left[\left\| \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\intercal - \hat{M}(\boldsymbol{A}) \right\|_F^2 \right].$$

Symmetric model

$$\boldsymbol{Y} = \frac{\rho_\Theta^2}{n}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\intercal + \boldsymbol{W}, \qquad \boldsymbol{W} \sim \frac{1}{\sqrt{n}}\mathsf{GOE}(n),$$

$$\mathsf{MMSE}_n(\boldsymbol{Y}) = \min_{\hat{M}(\boldsymbol{Y})} \frac{1}{n^2}\mathbb{E}\left[\left\|\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\intercal - \hat{M}(\boldsymbol{Y})\right\|_F^2\right], \text{ (Lelarge and Miolane, 2019).}$$

Information-theoretic lower bound on $\boldsymbol{\Lambda}$, $s_n = 1/\sqrt[4]{nd}$

Theorem 4

Assume $\mathbb{E}[\Theta] = \mathbb{E}[\Theta^3] = 0$, Θ, Λ are sub-gaussian random variables, and $d/n \to \infty$, then for all but countably many values of ρ_Θ, we have

$$\liminf_{n,d\to\infty} \mathsf{MMSE}_n(\boldsymbol{A}) \geq \lim_{n,d\to\infty} \mathsf{MMSE}_n(\boldsymbol{Y}).$$

Furthermore, $\lim_{n,d\to\infty} \mathsf{MMSE}_n(\boldsymbol{A}) = \lim_{n,d\to\infty} \mathsf{MMSE}_n(\boldsymbol{Y})$ holds when

(a) $d/n^3 \to \infty$,

(b) $n^{6/5}(\log d)^{-8/5}(\log n)^{-2/5} \gg d \gg n$.

Application: Gaussian mixture clustering

Application: Gaussian mixture clustering

- d features, n observations, $d \gg n$.

Application: Gaussian mixture clustering

- d features, n observations, $d \gg n$.
- $\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda} \boldsymbol{\Theta}^\intercal + \boldsymbol{Z} \in \mathbb{R}^{n \times d}$,

Application: Gaussian mixture clustering

- d features, n observations, $d \gg n$.
- $\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda} \boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}$,
- $\Lambda_i \overset{iid}{\sim} \mathrm{Unif}\{-1, +1\}$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$ with $\mathbb{E}[\Theta] = \mathbb{E}[\Theta^3] = 0$, $Z_{ij} \overset{iid}{\sim} \mathsf{N}(0, 1)$.

Application: Gaussian mixture clustering

- d features, n observations, $d \gg n$.
- $\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda} \boldsymbol{\Theta}^{\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}$,
- $\Lambda_i \overset{iid}{\sim} \mathrm{Unif}\{-1, +1\}$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$ with $\mathbb{E}[\Theta] = \mathbb{E}[\Theta^3] = 0$, $Z_{ij} \overset{iid}{\sim} \mathsf{N}(0, 1)$.
- Metric we consider:

$$\mathsf{Overlap}_n = \max_{s = \pm 1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \hat{\Lambda}_i = s \Lambda_i \right\}$$

Application: Gaussian mixture clustering

- d features, n observations, $d \gg n$.
- $\boldsymbol{A} = \frac{1}{\sqrt[4]{nd}} \boldsymbol{\Lambda} \boldsymbol{\Theta}^\mathsf{T} + \boldsymbol{Z} \in \mathbb{R}^{n \times d}$,
- $\Lambda_i \overset{iid}{\sim} \mathrm{Unif}\{-1, +1\}$, $\Theta_j \overset{iid}{\sim} \mu_\Theta$ with $\mathbb{E}[\Theta] = \mathbb{E}[\Theta^3] = 0$, $Z_{ij} \overset{iid}{\sim} \mathsf{N}(0, 1)$.
- Metric we consider:

$$\mathsf{Overlap}_n = \max_{s = \pm 1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \left\{ \hat{\Lambda}_i = s \Lambda_i \right\}$$

Application: Gaussian mixture clustering

Partial recovery: $\exists \epsilon > 0$, such that $\mathbb{P}\left(\text{Overlap}_n \geq \frac{1}{2} + \epsilon\right) \to 1$.

Application: Gaussian mixture clustering

Partial recovery: $\exists \epsilon > 0$, such that $\mathbb{P}\left(\text{Overlap}_n \geq \frac{1}{2} + \epsilon\right) \to 1$.

$\rho_\Theta = \mathbb{E}[\Theta^2]$

Application: Gaussian mixture clustering

Partial recovery: $\exists \epsilon > 0$, such that $\mathbb{P}\left(\text{Overlap}_n \geq \frac{1}{2} + \epsilon\right) \to 1$.

$\rho_\Theta = \mathbb{E}[\Theta^2]$

Two regimes:

Application: Gaussian mixture clustering

Partial recovery: $\exists\epsilon > 0$, such that $\mathbb{P}\left(\text{Overlap}_n \geq \frac{1}{2} + \epsilon\right) \to 1$.

$\rho_\Theta = \mathbb{E}[\Theta^2]$

Two regimes:

1. $\rho_\Theta \leq 1$: partial recovery impossible.

Application: Gaussian mixture clustering

Partial recovery: $\exists \epsilon > 0$, such that $\mathbb{P}\left(\mathsf{Overlap}_n \geq \frac{1}{2} + \epsilon\right) \to 1$.

$\rho_\Theta = \mathbb{E}[\Theta^2]$

Two regimes:

1. $\rho_\Theta \leq 1$: partial recovery impossible.
2. $\rho_\Theta > 1$: partial recovery possible via AMP,

$$\mathsf{Overlap}_n \xrightarrow{a.s.} \Phi\left(\rho_\Theta \sqrt{q^*(\rho_\Theta^2)}\right),$$

$\Phi(x) = \mathbb{P}(G \leq x)$.

$$\mathcal{F} : (s, q) \in \mathbb{R}_+^2 \mapsto -\frac{s}{4}q^2 + \mathbb{E}\log\left(\int \exp\left(\sqrt{sq}Z\lambda + sq\lambda\Lambda - \frac{s}{2}q\lambda^2\right) \mathrm{d}P_\Lambda(\lambda)\right),$$

$$q^*(s) : \text{ maximizer of } q \mapsto \mathcal{F}(s, q).$$

Application: Gaussian mixture clustering

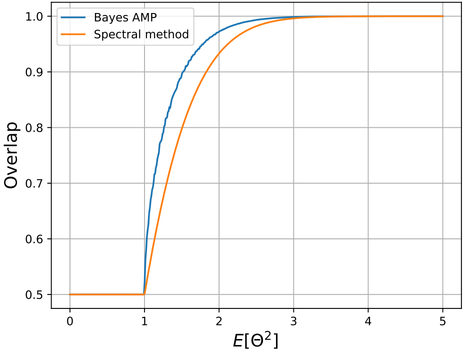

Figure: Phase transition diagram for high-dimensional GMM

References I

Lelarge, M. & Miolane, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields, 173*(3), 859–929.

Lesieur, T., Krzakala, F. & Zdeborová, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment, 2017*(7), 073403.

Miolane, L. (2017). Fundamental limits of low-rank matrix estimation: The non-symmetric case. *arXiv preprint arXiv:1702.00473.*

Montanari, A. & Venkataramanan, R. (2021). Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics, 49*(1), 321–345.

Thank You!