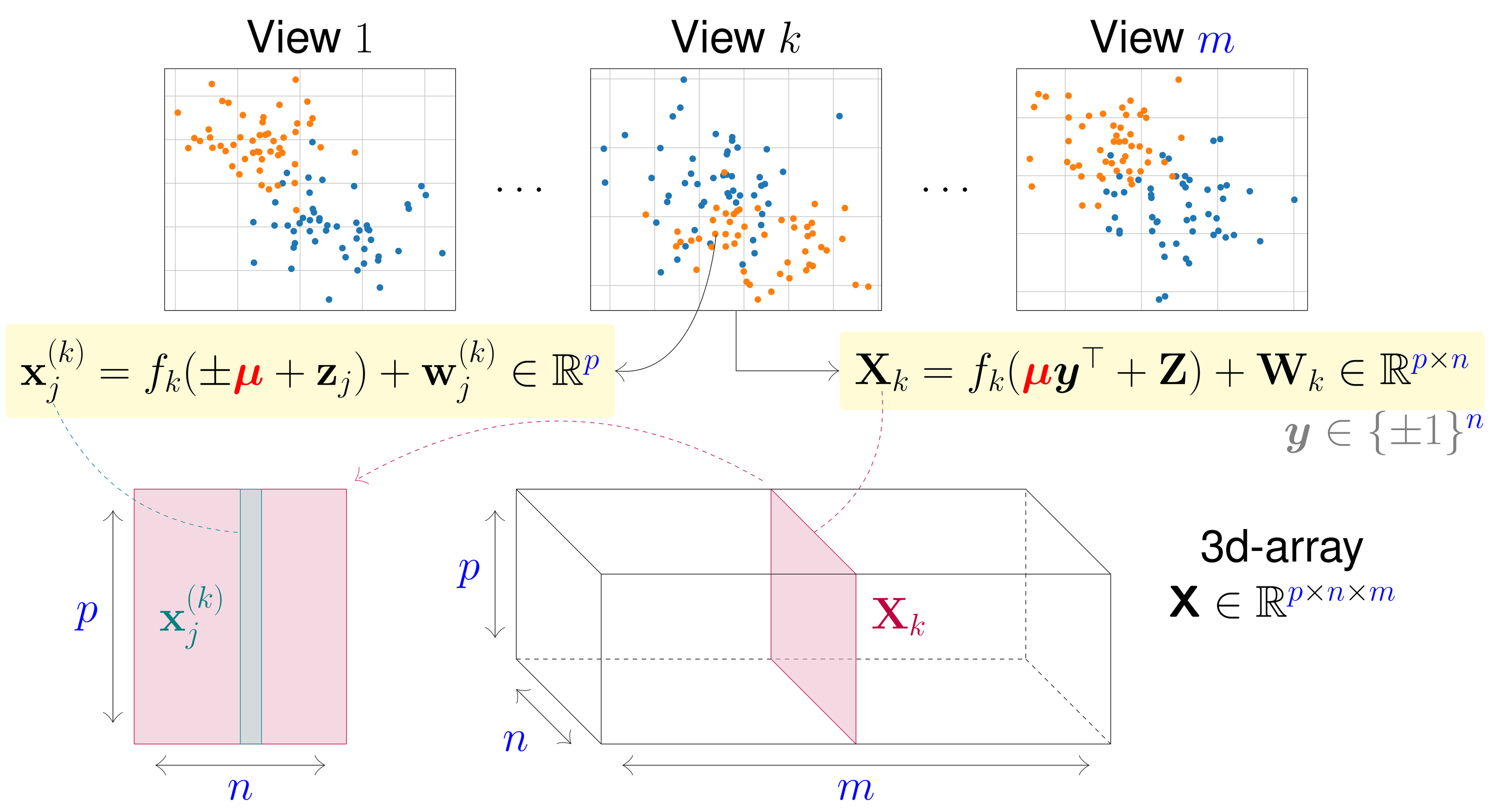
Performance Gaps in Multi-view Clustering under the Nested Matrix-Tensor Model

Hugo Lebeau[1] Mohamed E. A. Seddik[2] J. Henrique de M. Goulart[3]

[1]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG [2]Technology Innovation Institute [3]Univ. Toulouse, INP-ENSEEIHT, IRIT, CNRS

Motivation: a Multi-view Clustering Problem



- Tensor representation: $\mathbf{X}_{:,:,k} = \mathbf{X}_k$ for all $k \in \{1, \ldots, m\}$.
- **Assumption:** $f_k(\boldsymbol{M}) = h_k \boldsymbol{M}$

  $\rightsquigarrow \quad \boxed{\mathbf{X} = (\boldsymbol{\mu} \boldsymbol{y}^\top + \mathbf{Z}) \otimes \boldsymbol{h} + \mathbf{W}} \in \mathbb{R}^{p \times n \times m}.$
- **Clustering problem:** reconstruct $\boldsymbol{y}$ from the observation $\mathbf{X}$.

Nested Matrix-Tensor Model

$$\mathbf{T} = \beta_T \mathbf{M} \otimes \boldsymbol{z} + \frac{1}{\sqrt{n_T}} \mathbf{W} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \qquad \mathbf{M} = \beta_M \boldsymbol{x} \boldsymbol{y}^\top + \frac{1}{\sqrt{n_M}} \mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$$

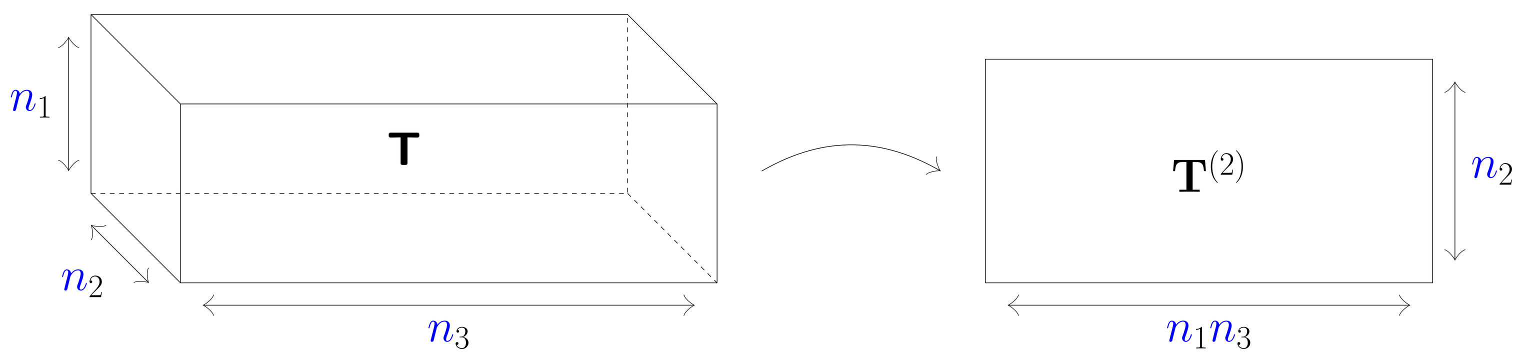
- $n_M = n_1 + n_2, \quad n_T = n_1 + n_2 + n_3, \quad Z_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad W_{i,j,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$
- $\mathbf{M}$ is a standard *spiked random matrix model.*
- "Double noise" structure: $\mathbf{Z}$ and $\mathbf{W}$.
- (β_M, β_T) control the signal-to-noise ratio.

Best rank-one approximation problem ⚠️ NP-hard

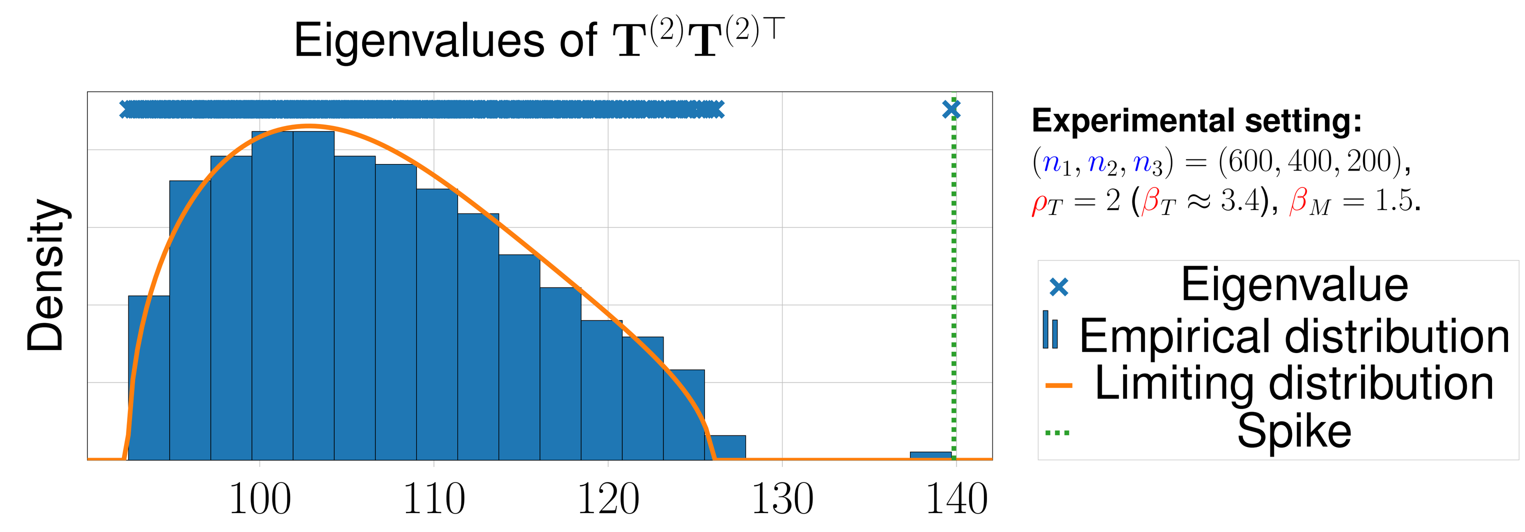
$$(\mathbf{x}_\star, \mathbf{y}_\star, \mathbf{z}_\star) \in \underset{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) \in \mathbb{S}^{n_1-1} \times \mathbb{S}^{n_2-1} \times \mathbb{S}^{n_3-1}}{\arg\max} \langle \mathbf{T}, \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} \rangle$$

- **Tractable algorithmic approach:** estimate $\boldsymbol{y}$ via the unfolding

$$\mathbf{T}^{(2)} = \beta_T \beta_M \boldsymbol{y} (\boldsymbol{x} \boxtimes \boldsymbol{z})^\top + \frac{\beta_T}{\sqrt{n_M}} \mathbf{Z}^\top (\boldsymbol{I}_{n_1} \boxtimes \boldsymbol{z})^\top + \frac{1}{\sqrt{n_T}} \mathbf{W}^{(2)}.$$



Random Matrix Analysis along the Second Mode



- The dominant eigenvector of $\mathbf{T}^{(2)}\mathbf{T}^{(2)\top}$ is an estimator of $\boldsymbol{y}$.
- Random matrix analysis in the **asymptotic regime**

$$n_1, n_2, n_3 \to +\infty \quad \text{with} \quad 0 < \lim \frac{n_\ell}{n_1 + n_2 + n_3} \overset{\text{def}}{=} c_\ell < +\infty \quad \text{for all } \ell \in \{1, 2, 3\}.$$

Non-trivial regime

$$\beta_M = \Theta(1) \quad \text{and} \quad \beta_T = \Theta(n_T^{1/4}).$$

Limiting Spectral Distribution

The centered-and-scaled matrix

$$\frac{n_T}{\sqrt{n_1 n_2 n_3}} \left[\mathbf{T}^{(2)} \mathbf{T}^{(2)\top} - \frac{n_1 n_3}{n_T} \boldsymbol{I}_{n_2} \right]$$

has a limiting spectral distribution $\tilde{\nu}$ whose Stieltjes transform $\tilde{m}(\tilde{s})$ satisfies

$$\frac{\rho_T c_2}{1 - c_3} \tilde{m}^3(\tilde{s}) + \left(1 + \tilde{s} \frac{\rho_T c_2}{1 - c_3}\right) \tilde{m}^2(\tilde{s}) + \left(\tilde{s} + \frac{\rho_T (c_2 - c_1)}{1 - c_3}\right) \tilde{m}(\tilde{s}) + 1 = 0$$

for all $\tilde{s} \in \mathbb{C} \setminus \operatorname{supp} \tilde{\nu}$ with $\rho_T = \lim \frac{\beta_T^2 n_T}{\sqrt{n_1 n_2 n_3}}$.

Phase Transition and Spike Behavior

Let $$\tilde{\xi} = \frac{\rho_T}{\beta_M^2} \left(\frac{c_1}{1 - c_3} + \beta_M^2 \right) \left(\frac{c_2}{1 - c_3} + \beta_M^2 \right) + \frac{1}{\rho_T \left(\frac{c_2}{1-c_3} + \beta_M^2 \right)}$$

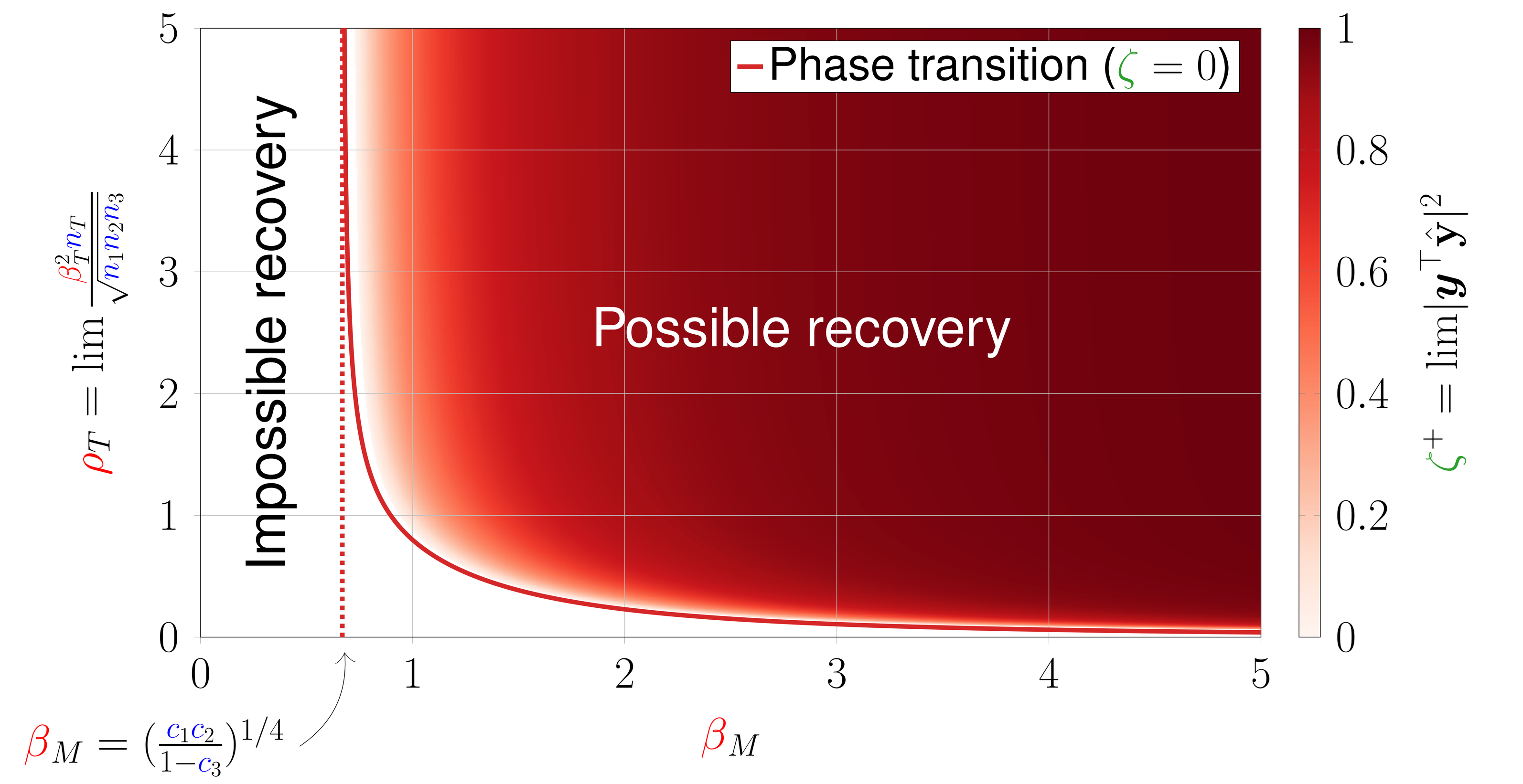
and $$\zeta = 1 - \frac{1}{\beta_M^2 \left(\frac{c_2}{1-c_3} + \beta_M^2 \right)} \left[\left(\frac{\beta_M^2}{\rho_T \left(\frac{c_2}{1-c_3} + \beta_M^2 \right)} \right)^2 + \frac{c_2}{1 - c_3} \left(\frac{c_1}{1 - c_3} + \beta_M^2 \right) \right].$$

If $\zeta > 0$, then,

$$\lambda_{\max} \xrightarrow[n_1, n_2, n_3 \to +\infty]{\text{a.s.}} \tilde{\xi} \quad \text{and} \quad \left| \boldsymbol{y}^\top \hat{\mathbf{y}} \right|^2 \xrightarrow[n_1, n_2, n_3 \to +\infty]{\text{a.s.}} \zeta$$

where $\lambda_{\max}$ is the dominant eigenvalue of $\frac{n_T}{\sqrt{n_1 n_2 n_3}} \left[\mathbf{T}^{(2)} \mathbf{T}^{(2)\top} - \frac{n_1 n_3}{n_T} \boldsymbol{I}_{n_2} \right]$ and $\hat{\mathbf{y}}$ is the corresponding eigenvector.

Phase Diagram



Performance of Multi-view Spectral Clustering

- Back to the original problem

$$\mathbf{X} = (\boldsymbol{\mu} \bar{\boldsymbol{y}}^\top + \mathbf{Z}) \otimes \boldsymbol{h} + \mathbf{W} \quad \text{with} \quad \bar{\boldsymbol{y}} = \frac{\boldsymbol{y}}{\sqrt{n}} \text{ and } \begin{cases} Z_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{p+n}\right) \\ W_{i,j,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{p+n+m}\right) \end{cases}.$$

- $(\|\boldsymbol{\mu}\|, \|\boldsymbol{h}\|) \longleftrightarrow (\beta_M, \beta_T)$ and $\rho = \|\boldsymbol{h}\|^2 \frac{p+n+m}{\sqrt{pnm}}$.

$$\sqrt{\frac{n}{1 - \zeta^+}} \left(\hat{y}_j - \sqrt{\zeta^+} \bar{y}_j \right) \xrightarrow[p,n,m \to +\infty]{\mathcal{D}} \mathcal{N}(0, 1) \qquad \text{for all } j \in \{1, \ldots, n\}.$$

Performance Gaps