Optimal Permutation Estimation in Crowd-Sourcing Problems

Alexandra Carpentier

Universität Potsdam

Based on joint works with Emmanuel Pilliat (Uni Montpellier and INRAE) and Nicolas Verzelen (INRAE)

June, 1st 2023
Typical Dataset in Crowd-Sourcing

Cifar10H dataset: 10000 images, 10 labels.
Typical Dataset in Crowd-Sourcing

Cifar10H dataset: 10000 images, 10 labels.

- Identification a worker: annotator_id
Typical Dataset in Crowd-Sourcing

Cifar10H dataset: 10000 images, 10 labels.

- Identification a worker: annotator_id
- Evaluation on a given image: correct_guess
Typical Dataset in Crowd-Sourcing

- Identification a worker: annotator_id
- Evaluation on a given image: correct_guess
This Talk

We consider a **ranking** problem:

- Given the observation of the correctness of answers of \( n \) experts on \( d \) questions,
- We want to rank the experts according to their ability.

**Question**: how well can we recover their ranking in a minimax sense?
Example of Possible Data

10 questions

\[
\begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{pmatrix}
\]

4 experts

0: Wrong answer 1: Correct answer
Example of Possible Questions

4 experts

\[
\begin{pmatrix}
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{pmatrix}
\]

0: Wrong answer 1: Correct answer

Bad Experts  Good Experts
## Example of Possible Data

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>1 0 1 0 0 0 1 0 1 1</td>
<td>0 0 1 1 1 1 0 1 1 1</td>
<td>0 0 0 0 1 0 1 1 0 1</td>
<td>0 0 1 1 1 1 1 1 1 1</td>
</tr>
</tbody>
</table>

0: Wrong answer 1: Correct answer

- **Hard Questions**
- **Easy Questions**
Example of Possible Data

\[\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & \end{pmatrix}\]

0: Wrong answer  1: Correct answer

This talk: **Ranking of Experts**
Example of Possible Data

4 experts

\[
\begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{pmatrix}
\]

0: Wrong answer 1: Correct answer

This talk: Ranking of Experts

Under Known Difficulty of the questions
Experts/Questions Setting

Experts $i \in \{1, \ldots, n\}$ and questions $k \in \{1, \ldots, d\}$. We observe for all $i, k$:

$$Y_{ik} \sim \text{Bern}(M_{ik}).$$

<table>
<thead>
<tr>
<th>1: Correct</th>
<th>0: Wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 1 0 0 0 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>0 0 1 1 1 1 0 1 1 1</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 1 0 1 1 0 1</td>
<td></td>
</tr>
<tr>
<td>0 0 1 1 1 1 1 1 1 1</td>
<td></td>
</tr>
</tbody>
</table>
Experts/Questions Setting

**Experts** $i \in \{1, \ldots, n\}$ and **questions** $k \in \{1, \ldots, d\}$. We observe for all $i, k$:

$$ Y_{ik} \sim \text{Bern}(M_{ik}) \cdot $$

expert $i$ correct at question $k$  
$$ \Leftrightarrow Y_{ik} = 1 \cdot $$
Experts/Questions Setting

Experts $i \in \{1, \ldots, n\}$ and questions $k \in \{1, \ldots, d\}$. We observe for all $i, k$:

$$Y_{ik} \sim \text{Bern}(M_{ik}) .$$

expert $i$ correct at question $k$  
$\Leftrightarrow Y_{ik} = 1$ .

1: Correct 0: Wrong

$$
\begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
$$

$\blacklozenge$ $M_{ik} = 1/2$: random choice of expert $i$ at question $k$  
$\blacklozenge$ $M_{ik} = 1$: Expert $i$ knows perfectly the answer of question $k$
Statistical Models

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n\times d} \]
Statistical Models

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
**Statistical Models**

**Observation Model**

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
- \(M_{ik} \in [0, 1]\) for all \(i, k\)
Statistical Models

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Parametric Models for \(M\):

- Questions Equally Difficult \(\sim M_{ik} = a_i \approx [Dawid and Skene, 1979]\)
- Ability/Difficulty \(\sim M_{ik} = \phi(\alpha_i - \beta_k) \approx [Bradley and Terry, 1952]\)
Statistical Models

**Observation Model**

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

**Parametric Models for \(M\):**

- Questions Equally Difficult \(\sim M_{ik} = a_i \approx \) [Dawid and Skene, 1979]
- Ability/Difficulty \(\sim M_{ik} = \phi(\alpha_i - \beta_k) \approx \) [Bradley and Terry, 1952]

**Non-Parametric Models for \(M\) \(\approx\) [Mao et al., 2018]

- Increasing Rows: \(M_{i,k} \leq M_{i,k+1}\)
Statistical Models

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Parametric Models for \(M\):

- Questions Equally Difficult \(\sim M_{ik} = a_i \approx [\text{Dawid and Skene, 1979}]\)
- Ability/Difficulty \(\sim M_{ik} = \phi(\alpha_i - \beta_k) \approx [\text{Bradley and Terry, 1952}]\)

Non-Parametric Models for \(M \approx [\text{Mao et al., 2018}]\)

- Increasing Rows: \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns up to permutation \(\pi^*\) of rows:
  \(M_{\pi^*(i),k} \leq M_{\pi^*(i+1),k}\)
Statistical Models

### Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian (e.g. Bernoulli)
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

### Parametric Models for \(M\):

- Questions Equally Difficult \(\sim M_{ik} = a_i \approx [\text{Dawid and Skene, 1979}]\)
- Ability/Difficulty \(\sim M_{ik} = \phi(\alpha_i - \beta_k) \approx [\text{Bradley and Terry, 1952}]\)

### Non-Parametric Models for \(M\) \(\approx [\text{Mao et al., 2018}]\)

- Increasing Rows: \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns up to permutation \(\pi^*\) of rows:
  \[ M_{\pi^*(i),k} \leq M_{\pi^*(i+1),k} \]
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

White = 0; Black = 1

Matrix \(M_{\pi^*}\). (isotonic).
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

White = 0 ; Black = 1

Matrix \(M_{\pi^*}\). (isotonic).
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]
- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):
- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

White = 0 ; Black = 1

Matrix \(M\) (isotonic up to a permutation of experts).
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns \textbf{for an unknown permutation} \(\pi^*\)

White = 0 ; Black = 1

Matrix \(Y\) (\(M\) in noise).
Non Parametric Model

Observation Model

\[ Y = M + \epsilon \in \mathbb{R}^{n \times d} \]

- \((\epsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

Aim

Estimation of \(\pi^*\).
Example with $n, d = 150, M \in [0, 1]$
Example with $n, d = 150, M \in [0.25, 0.75]$
Example with $n, d = 150$, $M \in [0.4, 0.6]$
Bi-isotonic $M$ - Other representation

Each line $M_{i,:}$ represents an expert $i$
Bi-isotonic $M$ - Other representation

Each line $M_{i,*}$ represents an expert $i$
Bi-isotonic $M$ - Other representation

Each line $M_{i,:}$ represents an expert $i$
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

Error Measures

Permutation loss

For an estimator \(\hat{\pi}\) of \(\pi^*\)

\[
\text{Perm-Loss} := \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2 = \sum_{i=1}^{n} \sum_{k=1}^{d} (M_{\pi(i),k} - M_{\pi^*(i),k})^2
\]
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

Error Measures

Permutation loss

For an estimator \(\hat{\pi}\) of \(\pi^*\)

\[
\text{Perm-Loss} := \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2
\]

If the two lines are misclassified:

\[
\text{Perm-Loss} = 2rh^2
\]
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \( (\varepsilon_{ik}) \) i.i.d. subGaussian
- \( M_{ik} \in [0, 1] \) for all \( i, k \)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \( M_{i,k} \leq M_{i,k+1} \)
- Increasing Columns for an unknown permutation \( \pi^* \)

Error Measures

Permutation loss

For an estimator \( \hat{\pi} \) of \( \pi^* \)

\[ \text{Perm-Loss} := \| M_{\hat{\pi}} - M_{\pi^*} \|_F^2 \]

Estimation loss

For an estimator \( \hat{M} \) of \( M \)

\[ \text{Estim-Loss} := \| \hat{M} - M \|_F^2. \]
Non Parametric Model

Observation Model

\[ Y = M + \varepsilon \in \mathbb{R}^{n \times d} \]

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

Error Measures

Permutation loss

For an estimator \(\hat{\pi}\) of \(\pi^*\)

\[ \text{Perm-Loss} := \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2 \]

Estimation loss

For an estimator \(\hat{M}\) of \(M\)

\[ \text{Estim-Loss} := \|\hat{M} - M\|_F^2. \]

Aim

Estimation of \(\pi^*\).
## Non Parametric Model

### Observation Model

$$Y = M + \varepsilon \in \mathbb{R}^{n \times d}$$

- \((\varepsilon_{ik})\) i.i.d. subGaussian
- \(M_{ik} \in [0, 1]\) for all \(i, k\)

### Shape Constraints (Bi-isotonicity):

- Increasing Rows \(M_{i,k} \leq M_{i,k+1}\)
- Increasing Columns for an unknown permutation \(\pi^*\)

## Error Measures

### Permutation loss

For an estimator \(\hat{\pi}\) of \(\pi^*\)

$$\text{Perm-Loss} := \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2$$

### Estimation loss

For an estimator \(\hat{M}\) of \(M\)

$$\text{Estim-Loss} := \|\hat{M} - M\|_F^2.$$  

## Aim

Estimation of \(\pi^*\).
**Error Measures**

**Permutation loss**

For an estimator $\hat{\pi}$ of $\pi^*$

$$\text{Perm-Loss} := \|M_{\hat{\pi}} - M_{\pi^*}\|_F^2.$$  

**Estimation loss**

For an estimator $\hat{M}$ of $M$

$$\text{Estim-Loss} := \|\hat{M} - M\|_F^2.$$  

**Aim**

Estimation of $\pi^*$.  

**MiniMax-Risk**
## Error Measures

### Permutation loss
For an estimator $\hat{\pi}$ of $\pi^*$

$$\text{Perm-Loss} := \| M_{\hat{\pi}} - M_{\pi^*} \|_F^2.$$  

### Estimation loss
For an estimator $\hat{M}$ of $M$

$$\text{Estim-Loss} := \| \hat{M} - M \|_F^2.$$  

## MiniMax-Risk

### Max-Risk and MiniMax-Risk
If $\hat{\pi}$ is an estimator of $\pi^*$, we define

$$\text{Max-Perm}(\hat{\pi}) = \sup_{M,\pi^*} \mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2],$$

$$\text{MiniMax-Perm} = \inf_{\hat{\pi}} (\text{Max-Perm}(\hat{\pi}))$$

### Aim
Estimation of $\pi^*$.  

Error Measures

Permutation loss
For an estimator $\hat{\pi}$ of $\pi^*$

$$\text{Perm-Loss} := \| M_{\hat{\pi}} - M_{\pi^*} \|_F^2. $$

Estimation loss
For an estimator $\hat{M}$ of $M$

$$\text{Estim-Loss} := \| \hat{M} - M \|_F^2. $$

Aim
Estimation of $\pi^*$.

MiniMax-Risk

Max-Risk and MiniMax-Risk
If $\hat{\pi}$ is an estimator of $\pi^*$, we define

$$\text{Max-Perm}(\hat{\pi}) = \sup_{M, \pi^*} \mathbb{E}[\| M_{\hat{\pi}} - M_{\pi^*} \|_F^2],$$

$$\text{MiniMax-Perm} = \inf_{\hat{\pi}} (\text{Max-Perm}(\hat{\pi})).$$

Define similarly $\text{Max-Estim}$ and $\text{MiniMax-Estim}$ for estimation of $M$ with $\hat{M}$. 
Other Ranking and Permutation Problems

Related rectangular problems:

- **Two permutations** [Mao et al., 2018, Shah et al., 2019]
  
  $M$ is bi-isotonic up to permutations $\pi^*$ and $\sigma^*$ of rows and columns. 
  
  **Objective**: ranking the experts and the questions.
Other Ranking and Permutation Problems

Related rectangular problems:

- Two permutations [Mao et al., 2018, Shah et al., 2019]  
  $M$ is bi-isotonic up to permutations $\pi^*$ and $\sigma^*$ of rows and columns.  
  Objective: ranking the experts and the questions.

- Column isotony [Flammarion et al., 2019]
Other Ranking and Permutation Problems

Related rectangular problems:

- **Two permutations** [Mao et al., 2018, Shah et al., 2019]
  $M$ is bi-isotonic up to permutations $\pi^*$ and $\sigma^*$ of rows and columns.
  **Objective**: ranking the experts and the questions.

- **Column isotony** [Flammarion et al., 2019]

**Ranking players in a tournament**: $M$ is a $n \times n$ matrix with symmetries.

- **Non-parametric Models** SST [Shah et al., 2016]
Other Ranking and Permutation Problems

Related rectangular problems:

- **Two permutations** [Mao et al., 2018, Shah et al., 2019]
  \( M \) is bi-isotonic up to permutations \( \pi^* \) and \( \sigma^* \) of rows and columns.
  **Objective**: ranking the experts and the questions.

- **Column isotony** [Flammarion et al., 2019]

Ranking players in a tournament: \( M \) is a \( n \times n \) matrix with symmetries.

- **Non-parametric Models** SST [Shah et al., 2016]

- **Parametric Models**:
  Bradley-Luce-Terry (e.g. [Chen et al., 2019, Chen et al., 2022])
Other Ranking and Permutation Problems

Related rectangular problems:

- **Two permutations** [Mao et al., 2018, Shah et al., 2019]  
  $M$ is bi-isotonic up to permutations $\pi^*$ and $\sigma^*$ of rows and columns. 
  **Objective**: ranking the experts and the questions.

- **Column isotony** [Flammarion et al., 2019]

**Ranking players in a tournament**: $M$ is a $n \times n$ matrix with symmetries.

- **Non-parametric Models** SST [Shah et al., 2016]

- **Parametric Models**:  
  Bradley-Luce-Terry (e.g. [Chen et al., 2019, Chen et al., 2022])

**Short story**:

- No computational gap for *parametric models* (BLT, noisy sorting)

- Mostly unknown for *non-parametric* models: computational gaps were conjectured
Main questions

1. Is Estimating $\pi^*$ much easier than estimating $M$?
Main questions

1. Is Estimating $\pi^*$ much easier than estimating $M$?
2. Is there a computational-statistical gap?
Main questions

1. Is Estimating $\pi^*$ much easier than estimating $M$?
2. Is there a computational-statistical gap?

Our Contributions

- Control of MiniMax-Perm$(n, d)$
- A polynomial-time procedure achieves MiniMax-Perm$(n, d)$
Existing Methods

- Non-Polynomial Time Methods with Least Square
- Simple Global Average Comparison
- [Liu and Moitra, 2020] based on Hierarchical Clustering
Least Square on Bi-isotinic Matrix
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \{1, \ldots, n\}

- Mon be the set of all bi-isotonic matrix in [0, 1]

Least-square estimator

\[(\hat{M}^{LS}, \hat{\pi}^{LS}) = \arg \min_{\tilde{M} \in \text{Mon}, \tilde{\pi} \in \text{Perm}} \| \tilde{M}_{\tilde{\pi}} - Y \|_F^2]
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \{1, \ldots, n\}
- Mon be the set of all bi-isotonic matrix in [0, 1]

Least-square estimator

\[
(\hat{M}_{\text{LS}}, \hat{\pi}_{\text{LS}}) = \arg \min_{\tilde{M} \in \text{Mon}, \tilde{\pi} \in \text{Perm}} \| \tilde{M}_{\tilde{\pi}} - Y \|_F^2
\]

Matrix $M_{\pi^*}$, (bi-isotonic).
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \{1, \ldots, n\}
- Mon be the set of all bi-isotonic matrix in [0, 1]

**Least-square estimator**

\[
\begin{align*}
\left( \hat{M}^{LS}, \hat{\pi}^{LS} \right) &= \\
&= \arg\min_{\tilde{M} \in \text{Mon}, \tilde{\pi} \in \text{Perm}} \| \tilde{M}_{\tilde{\pi}} - Y \|_F^2
\end{align*}
\]

Matrix \( M \).
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \{1, \ldots, n\}
- Mon be the set of all bi-isotonic matrix in [0, 1]

**Least-square estimator**

\[
(\hat{M}^{LS}, \hat{\pi}^{LS}) = \arg \min_{\hat{M} \in \text{Mon}, \hat{\pi} \in \text{Perm}} \| \hat{M}_{\hat{\pi}} - Y \|_F^2
\]

Matrix Y.
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \(\{1, \ldots, n\}\)
- Mon be the set of all bi-isotonic matrix in \([0, 1]\)

**Least-square estimator**

\[
(\hat{M}^{LS}, \hat{\pi}^{LS}) = \arg\min_{\tilde{M} \in \text{Mon}, \tilde{\pi} \in \text{Perm}} \| \tilde{M}_{\tilde{\pi}} - Y \|_F^2
\]

No know polynomial-time method to compute \(Y_{\hat{\pi}^{LS}}\).
Non-Polynomial Time Method
[Mao et al., 2018]

- Perm the set of all permutation of \{1, \ldots, n\}
- Mon be the set of all bi-isotonic matrix in [0, 1]

Least-square estimator

\[
(\hat{M}^{LS}, \hat{\pi}^{LS}) = \arg\min_{\tilde{M} \in \text{Mon}, \tilde{\pi} \in \text{Perm}} \| \tilde{M}_{\tilde{\pi}} - Y \|_F^2
\]

Matrix \( \hat{M}_{\pi^*}^{LS} \).
Non-Polynomial Time Method [Mao et al., 2018]

<table>
<thead>
<tr>
<th>Least-square guarantees</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\hat{\pi}^{LS}, \hat{M}^{LS})$ satisfy -up to polylogs:</td>
</tr>
<tr>
<td>Max-Estim$(\hat{M}^{LS}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$</td>
</tr>
<tr>
<td>Max-Perm$(\hat{\pi}^{LS}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$</td>
</tr>
</tbody>
</table>
Non-Polynomial Time Method [Mao et al., 2018]

<table>
<thead>
<tr>
<th>Least-square guarantees</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\hat{\pi}^{LS}, \hat{M}^{LS})$ satisfy -up to polylogs:</td>
</tr>
<tr>
<td>Max-Estim$(\hat{M}^{LS}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$</td>
</tr>
<tr>
<td>Max-Perm$(\hat{\pi}^{LS}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$</td>
</tr>
</tbody>
</table>

**Entropy Arguments:**

- $n!$ permutations: $n \asymp \log(n!)$
Non-Polynomial Time Method [Mao et al., 2018]

Least-square guarantees

$$(\hat{\pi}^{\text{LS}}, \hat{M}^{\text{LS}})$$ satisfy -up to polylogs:

$$\text{Max-Estim}(\hat{M}^{\text{LS}}) \lesssim n \lor (\sqrt{nd} \land nd^{1/3})$$

$$\text{Max-Perm}(\hat{\pi}^{\text{LS}}) \lesssim n \lor (\sqrt{nd} \land nd^{1/3})$$

Entropy Arguments:

- $n!$ permutations: $n \asymp \log(n!)$
- Covering of bi-isotonic matrices: log-size $\asymp \sqrt{nd} \land nd^{1/3}$
Non-Polynomial Time Method [Mao et al., 2018]

Least-square guarantees

\( (\hat{\pi}^{LS}, \hat{M}^{LS}) \) satisfy -up to polylogs:

\[
\begin{align*}
\text{Max-Estim}(\hat{M}^{LS}) &\lesssim n \lor (\sqrt{nd} \wedge nd^{1/3}) \\
\text{Max-Perm}(\hat{\pi}^{LS}) &\lesssim n \lor (\sqrt{nd} \wedge nd^{1/3})
\end{align*}
\]

Entropy Arguments:

- \( n! \) permutations: \( n \approx \log(n!) \)
- Covering of bi-isotonic matrices: log-size \( \approx \sqrt{nd} \wedge nd^{1/3} \)

Remarks:

- MiniMax-Estim Optimal [Mao et al., 2018]
- not proven to be MiniMax-Perm Optimal
## Summary

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>??</td>
<td>??</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

But algo. not polynomial time.
Global Average Comparison

[Pananjady and Samworth, 2020,
Shah et al., 2019]

Matrix $M$. 
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:

- Compute expert $i$ average performances on all questions:

$$\overline{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik}$$

Matrix $Y$ ($M$ in noise).
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:

- Compute expert $i$ average performances on all questions:
  \[ \bar{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik} \]

- Rank experts according to their average: $\hat{\pi}^{\text{av}}$

Matrix $Y_{\hat{\pi}^{\text{av}}}$ ($M$ in noise).
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:
- Compute expert $i$ average performances on all questions:
  \[
  \bar{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik}
  \]
- Rank experts according to their average: $\hat{\pi}^{av}$

Guarantees on $\hat{\pi}^{av}$
Max-Perm($\hat{\pi}^{av}$) $\asymp n\sqrt{d}$. 

Matrix $Y_{\hat{\pi}^{av}}$ ($M$ in noise).
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:

- Compute expert $i$ average performances on all questions:
  $$\overline{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik}$$

- Rank experts according to their average: $\hat{\pi}^{av}$

Guarantees on $\hat{\pi}^{av}$

Max-Perm($\hat{\pi}^{av}$) $\asymp n \sqrt{d}$.

Idea of Proof

Perfect expert on $\sqrt{d}$ questions VS random:

$$M_{1,..} = (\underbrace{.5.5.5.5.5 \ldots 5.5}_d 111111111111)$$

$$M_{2,..} = (\underbrace{.5.5.5.5.5 \ldots 5.5}_d 5.5.5.5.5.5.5.5.5) \sim \sqrt{d}$$
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:
- Compute expert $i$ average performances on all questions:
  \[
  \overline{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik}
  \]
- Rank experts according to their average: $\hat{\pi}^{av}$

Guarantees on $\hat{\pi}^{av}$
Max-Perm($\hat{\pi}^{av}$) $\asymp n\sqrt{d}$.

Idea of Proof

Perfect expert on $\sqrt{d}$ questions VS random:

- $Y_{1,.} = (01101\ldots10\underbrace{11111111111}_{d})$
- $Y_{2,.} = (01000\ldots01\underbrace{10100101000}_{d}) \sim \sqrt{d}$

(Example of Observations)
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:

- Compute expert $i$ average performances on all questions:
  \[
  \bar{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik}
  \]

- Rank experts according to their average: $\hat{\pi}^{av}$

Guarantees on $\hat{\pi}^{av}$

\[
\text{Max-Perm}(\hat{\pi}^{av}) \preceq n\sqrt{d}.
\]

Idea of Proof

Perfect expert on $\sqrt{d}$ questions VS random:

\[
\begin{align*}
Y_{1,..} &= (01101 \ldots 10111111111) \\
Y_{2,..} &= (01000 \ldots 01101001010) \\
&\sim \sqrt{d}
\end{align*}
\]

1 and 2 cannot be distinguished with their average: \[
\text{Max-Perm}(\hat{\pi}^{av}) \preceq \sqrt{d}
\]
Global Average Comparison
[Pananjady and Samworth, 2020, Shah et al., 2019]

Method:

- Compute expert $i$ average performances on all questions:
  \[ \overline{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik} \]

- Rank experts according to their average: $\hat{\pi}^{av}$

Guarantees on $\hat{\pi}^{av}$

\[ \text{Max-Perm}(\hat{\pi}^{av}) \preceq n\sqrt{d}. \]

Idea of Proof

Perfect expert on $\sqrt{d}$ questions VS random:

\[
\begin{align*}
Y_1_{.,} &= (01101\ldots10111111111) \\
Y_2_{.,} &= (01000\ldots011010010100) \\
\text{\sim} \sqrt{d}
\end{align*}
\]

1 and 2 cannot be distinguished with their average: Max-Perm($\hat{\pi}^{av}$) \(\preceq \sqrt{d}\)

Lower Bound for $\hat{\pi}^{av}$: There exists $M$ s.t. Max-Perm($\hat{\pi}^{av}$) \(\geq n\sqrt{d}\)
**Global Average Comparison**  
[Pananjady and Samworth, 2020, Shah et al., 2019]

**Method:**

- Compute expert $i$ average performances on all questions:
  \[ \bar{Y}_i = \frac{1}{d} \sum_{k=1}^{d} Y_{ik} \]

- Rank experts according to their average: $\hat{\pi}^{av}$

**Guarantees on $\hat{\pi}^{av}$**

$\text{Max-Perm}(\hat{\pi}^{av}) \asymp n\sqrt{d}.$

---

**Idea of Proof**

**Perfect expert on $\sqrt{d}$ questions VS random:**

1. $Y_{1,.} = (01101\ldots10111111111)$
2. $Y_{2,.} = (01000\ldots011010010100)\sim \sqrt{d}$

1 and 2 cannot be distinguished with their average: $\text{Max-Perm}(\hat{\pi}^{av}) \asymp \sqrt{d}$

**Upper Bound:** For any $M, \pi^*$, $\text{Max-Perm}(\hat{\pi}^{av}) \lesssim n\sqrt{d}$
Summary

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>??</td>
<td>??</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
<tr>
<td>Global average (UB)</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
</tr>
</tbody>
</table>

Remarks:

- Algo. for rates in MiniMax-Estim and MiniMax-Perm **not in polynomial time**.
- One to one comparisons give UB but **sub-optimal** whenever $d \gtrsim 1$. 
CP and Hierarchical Clustering Based Algo.
[Liu and Moitra, 2020]

[Liu and Moitra, 2020] consider only the case \( d = n \), and provide a poly. time algo. returning \( \hat{\pi}(LM) \) such that

\[
\text{Max-Perm}(\hat{\pi}(LM)) \lesssim n.
\]
CP and Hierarchical Clustering Based Algo.
[Liu and Moitra, 2020]

[Liu and Moitra, 2020] consider only the case \(d = n\), and provide a poly. time algo. returning \(\hat{\pi}^{(LM)}\) such that

\[
\text{Max-Perm}(\hat{\pi}^{(LM)}) \lesssim n.
\]

One can push further their analysis for \(d \neq n\) and get \(n \lor d\) through this. Optimal for \(d = n\) in which case

\[
\text{MiniMax-Perm} \asymp n.
\]
CP and Hierarchical Clustering Based Algo.
[Liu and Moitra, 2020]

[Liu and Moitra, 2020] consider only the case \( d = n \), and provide a poly. time algo. returning \( \hat{\pi}^{(LM)} \) such that

\[
\text{Max-Perm}(\hat{\pi}^{(LM)}) \lesssim n.
\]

One can push further their analysis for \( d \neq n \) and get \( n \lor d \) through this. Optimal for \( d = n \) in which case

\[
\text{MiniMax-Perm} \asymp n.
\]
CP and Hierarchical Clustering Based Algo.

[Liu and Moitra, 2020] consider only the case $d = n$, and provide a poly. time algo. returning $\hat{\pi}(LM)$ such that

$$\text{Max-Perm}(\hat{\pi}(LM)) \lesssim n.$$ 

One can push further their analysis for $d \neq n$ and get $n \lor d$ through this. Optimal for $d = n$ in which case

$$\text{MiniMax-Perm} \asymp n.$$
Summary

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>??</td>
<td>??</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
<tr>
<td>Global average (UB)</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
</tr>
<tr>
<td>Ext. of LM (UB)</td>
<td>$d$</td>
<td>$d$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

Remarks:

- Poly. time algo of LM achieves MiniMax-Perm and MiniMax-Estim for $d = n$
- This algorithm can be analysed in a more refined way for $d \neq n$ - but not done in [Liu and Moitra, 2020].
Minimax and Poly. Time

Theorem [P., Carpentier, Verzelen, 2022]

Assume we have polylog samples.
There exists a estimator $\hat{\pi}$ of $\pi^*$ which is poly. time and minimax optimal

$$\mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2] \lesssim n \vee (n^{3/4}d^{1/4} \wedge nd^{1/6}) \asymp \text{MiniMax-Perm}.$$
Minimax and Poly. Time

Theorem [P., Carpentier, Verzelen, 2022]

Assume we have polylog samples.
There exists an estimator \( \hat{\pi} \) of \( \pi^* \) which is poly. time and minimax optimal

\[
\mathbb{E}[\| M_{\hat{\pi}} - M_{\pi^*} \|_F^2] \lesssim n \lor (n^{3/4}d^{1/4} \land nd^{1/6}) \preceq \text{MiniMax-Perm}.
\]

<table>
<thead>
<tr>
<th></th>
<th>( n \lesssim d^{1/3} )</th>
<th>( d^{1/3} \lesssim n \lesssim d )</th>
<th>( d \lesssim n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>( nd^{1/6} )</td>
<td>( n^{3/4}d^{1/4} )</td>
<td>( n )</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>( nd^{1/3} )</td>
<td>( \sqrt{nd} )</td>
<td>( n )</td>
</tr>
</tbody>
</table>
Minimax and Poly. Time

Theorem [P., Carpentier, Verzelen, 2022]

Assume we have polylog samples.
There exists an estimator $\hat{\pi}$ of $\pi^*$ which is poly. time and minimax optimal

$$\mathbb{E}[\|M_{\hat{\pi}} - M_{\pi^*}\|_F^2] \lesssim n \lor (n^{3/4}d^{1/4} \land nd^{1/6}) \lesssim \text{MiniMax-Perm}.$$ 

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>$nd^{1/6}$</td>
<td>$n^{3/4}d^{1/4}$</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

Can be combined with bi-isotonic regression to have a poly. time MiniMax-Estim algo!
Summary

Poly. time algo achieving the minimax rates:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>$n d^{1/6}$</td>
<td>$n^{3/4} d^{1/4}$</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$n d^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>
Uniform distance between two experts

Global average comparison is optimal:
Constant Perm-Risk - Confusion only if $h \lesssim 1/\sqrt{d}$. 
Localised distance between two experts

\[ \psi([d]) = \frac{1}{d} \sum_{i=1}^{d} Y_i \text{ achieves} \]

Perm-Risk \( \asymp \sqrt{d} \gg d^{1/6} \)
From Global to Local Averages

Global average good.

Global average bad $\rightarrow$ need to localise.
From Global to Local Averages

Global average good.

Global average bad → need to localise.

Idea:

- Estimate by a change point (CP) method windows where any of the two experts changes by more than $h$. 
From Global to Local Averages

Global average good.

Global average bad → need to localise.

Idea:

- Estimate by a change point (CP) method windows where any of the two experts changes by more than $h$.
- Compute local average on these windows.
From Global to Local Averages

Global average good.

Global average bad $\rightarrow$ need to localise.

Idea:

- Estimate by a change point (CP) method windows where any of the two experts changes by more than $h$.
- Compute local average on these windows.

[Liu and Moitra, 2020] introduced this idea of localisation with CP - in a different context and regime.
Toward a Worst Case Scenario

\[ \frac{1}{6} \] is optimal for two experts: \( \text{MiniMax-Perm} \approx \frac{1}{6} \).

For any \( n \) (UB): \( \text{MiniMax-Perm} \precsim nd^{1/6} \).
Toward a Worst Case Scenario

**Idea:**

- A CP of size $h$ can be detected on a window of $1/h^2$ questions.
Toward a Worst Case Scenario

Idea:

- A CP of size $h$ can be detected on a window of $1/h^2$ questions.
- At most $1/h$ of these CP, since $M \in [0, 1]$. 

![Diagram showing a step function with a window of questions and the number of CP detected.](image)
Toward a Worst Case Scenario

**Idea:**

- A CP of size $h$ can be detected on a window of $1/h^2$ questions.

- At most $1/h$ of these CP, since $M \in [0, 1]$.

- If they are indistinguishable at scale $h$:

\[
\|M_1 - M_2\|_2^2 \leq h \|M_1 - M_2\|_1 \\
\leq h \sqrt{\frac{1}{h^2} \frac{1}{h}} \wedge d \\
\leq d^{1/6} .
\]
Toward a Worst Case Scenario

**Idea:**

- A CP of size $h$ can be detected on a window of $1/h^2$ questions.
- At most $1/h$ of these CP, since $M \in [0, 1]$.
- If they are indistinguishable at scale $h$:
  \[
  \| M_1. - M_2. \|_2^2 \leq h \| M_1. - M_2. \|_1 \leq h \sqrt{\frac{1}{h^2} \frac{1}{h}} \wedge d \leq d^{1/6}.
  \]
- $d^{1/6}$ is optimal for two experts: $\text{MiniMax-Perm} \asymp d^{1/6}$.
Toward a Worst Case Scenario

Idea:

- A CP of size $h$ can be detected on a window of $1/h^2$ questions.

- At most $1/h$ of these CP, since $M \in [0, 1]$.

- If they are indistinguishable at scale $h$:

  $$\|M_1. - M_2.\|_2^2 \leq h \|M_1. - M_2.\|_1 \leq h \sqrt{\frac{1}{h^2}} \wedge d \leq d^{1/6}.$$ 

- $d^{1/6}$ is optimal for two experts: $\text{MiniMax-Perm} \asymp d^{1/6}$.

- For any $n$ (UB): $\text{MiniMax-Perm} \lesssim nd^{1/6}$.
Poly. time algo achieving the minimax rates:

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>$nd^{1/6}$</td>
<td>$n^{3/4}d^{1/4}$</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
<tr>
<td>Global average (UB)</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
</tr>
<tr>
<td>Ext. of LM (UB)</td>
<td>$d$</td>
<td>$d$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

Ext. of LM (UB) extends [Liu and Moitra, 2020] to $d \neq n$


## Summary

Poly. time algo achieving the minimax rates:

<table>
<thead>
<tr>
<th></th>
<th>( n \lesssim d^{1/3} )</th>
<th>( d^{1/3} \lesssim n \lesssim d )</th>
<th>( d \lesssim n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>( nd^{1/6} )</td>
<td>( n^{3/4} d^{1/4} )</td>
<td>( n )</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>( nd^{1/3} )</td>
<td>( \sqrt{nd} )</td>
<td>( n )</td>
</tr>
<tr>
<td>Global average (UB)</td>
<td>( n\sqrt{d} )</td>
<td>( n\sqrt{d} )</td>
<td>( n\sqrt{d} )</td>
</tr>
<tr>
<td>Ext. of LM (UB)</td>
<td>( d )</td>
<td>( d )</td>
<td>( n )</td>
</tr>
</tbody>
</table>

Ext. of LM (UB) extends [Liu and Moitra, 2020] to \( d \neq n \)
Introduction
Overview of Existing Methods
Minimax and Poly. Time Algo.
Hierarchical Clustering

Beyond [Liu and Moitra, 2020] for $d \neq n$
Hierarchical Clustering
Beyond [Liu and Moitra, 2020] for $d \neq n$
Hierarchical Clustering

Beyond [Liu and Moitra, 2020] for $d \neq n$
Hierarchical Clustering
Beyond [Liu and Moitra, 2020] for $d \neq n$
Hierarchical Clustering
Beyond [Liu and Moitra, 2020] for $d \neq n$
Hierarchical Clustering
Beyond [Liu and Moitra, 2020] for $d \neq n$
Worst Case for a Group $G^{(0)}$  
($n \gg d^{1/3}$)

In $G^{(0)}$, an expert is either in $L$ or in $U$.  

\[ G^{(4)} \]
\[ G^{(3)} \]
\[ G^{(2)} \]
\[ G^{(1)} \]
\[ G^{(0)} \]
\[ G^{(-1)} \]
\[ G^{(-2)} \]
\[ G^{(-3)} \]
Worst Case for a Group $G^{(0)}$ 
($n \gg d^{1/3}$)

In $G^{(0)}$, an expert is either in $L$ or in $U$.

After Aggregation

\[
\begin{pmatrix}
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
\end{pmatrix}
\]

1: above the mean ($U$)  
-1: below the mean ($L$)
Worst Case for a Group $G^{(0)}$ 
($n \gg d^{1/3}$)

In $G^{(0)}$, an expert is either in $L$ or in $U$.

After Aggregation

\[
\begin{pmatrix}
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\]

1: above the mean ($U$) 
-1: below the mean ($L$)

**Rank one matrix $\sim$ (PCA):**
1st left singular vector: better clustering than local averages in some regimes
Beyond [Liu and Moitra, 2020] for $d \neq n$

The corresponding Max-Perm is upper bounded by

$$n \lor (n^{2/3} d^{1/3}) .$$
Beyond [Liu and Moitra, 2020] for $d \neq n$

The corresponding Max-Perm is upper bounded by

$$n \lor (n^{2/3}d^{1/3}) .$$

- Better than (UB) of [Liu and Moitra, 2020] (CP + PCA) - Improvement when $d < n$:

$$n \lor d \gg n \lor (n^{2/3}d^{1/3}) .$$
Beyond [Liu and Moitra, 2020] for $d \neq n$

The corresponding Max-Perm is upper bounded by

$$n \lor (n^{2/3}d^{1/3})$$

- Better than (UB) of [Liu and Moitra, 2020] (CP + PCA) -
  Improvement when $d < n$:

$$n \lor d \gg n \lor (n^{2/3}d^{1/3})$$

- But not Optimal!

$$n \lor (n^{2/3}d^{1/3}) \gg n \lor (n^{3/4}d^{1/4})$$
Summary

Poly. time algo achieving the minimax rates:

<table>
<thead>
<tr>
<th>Method</th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>$nd^{1/6}$</td>
<td>$n^{3/4}d^{1/4}$</td>
<td>$n$</td>
</tr>
<tr>
<td>MiniMax-Estim</td>
<td>$nd^{1/3}$</td>
<td>$\sqrt{nd}$</td>
<td>$n$</td>
</tr>
<tr>
<td>Global average (UB)</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
<td>$n\sqrt{d}$</td>
</tr>
<tr>
<td>Ext. of LM (UB)</td>
<td>$d$</td>
<td>$d$</td>
<td>$n$</td>
</tr>
<tr>
<td>Super ext. of LM</td>
<td>$nd^{1/6}$</td>
<td>$n^{2/3}d^{1/3}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

**Remark:** Super ext. of LM requires a lot of additional work w.r.t. [Liu and Moitra, 2020]
Ideas to achieve $n^{3/4}d^{1/4}$

From an oblivious Hierarchical Clustering
Ideas to achieve $n^{3/4}d^{1/4}$

To using the Memory of the Tree
Ideas to achieve $n^{3/4}d^{1/4}$

To using the Memory of the Tree

$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$
Two Types of Information

$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$
Two Types of Information

First Type

$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$

Removing regions where $G^{(0)}$ is sandwiched
Two Types of Information

\[ G^{(0)} \] is sandwiched between \( \mathcal{V}_- \) and \( \mathcal{V}_+ \)

Second Type

Better Change-Point Detection
Conclusion of the Method with Memory

Poly. time algo achieving the minimax rates:

<table>
<thead>
<tr>
<th></th>
<th>$n \lesssim d^{1/3}$</th>
<th>$d^{1/3} \lesssim n \lesssim d$</th>
<th>$d \lesssim n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiniMax-Perm</td>
<td>$nd^{1/6}$</td>
<td>$n^{3/4}d^{1/4}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>
Conclusion

For all $n, d$:

- The rate MiniMax-Perm which is of order $n \vee (n^{3/4} d^{1/4} \wedge nd^{1/6})$ (UB and LB).
- An associated poly.-time ranking method.
- Together with bi-isotonic regression, this provides a poly.-time method for Minimax-Estim.
- Related to [Liu and Moitra, 2020] but new concepts necessary for minimax rate (memory of the tree).
- Setting can be relaxed without problems to partial observations.

Reference: [arXiv:2211.04092], accepted in AOS (2022)
Conclusion

For all $n, d$:

- The rate MiniMax-Perm which is of order $n \lor (n^{3/4}d^{1/4} \land nd^{1/6})$ (UB and LB).
- An associated poly.-time ranking method.
- Together with bi-isotonic regression, this provides a poly.-time method for Minimax-Estim.
- Related to [Liu and Moitra, 2020] but new concepts necessary for minimax rate (memory of the tree).
- Setting can be relaxed without problems to partial observations.

Reference: [arXiv:2211.04092], accepted in AOS (2022)

Research Directions:

- Unknown order on questions.
- Removing the isotonicity constraint on questions.
- Unknown answers: -observing labels instead of correctness.
(Isotonic)-$\pi^*$

- Isotonicity in experts for an unknown permutation $\pi^*$

- $M_{ik} \in [0, 1]$

- $(\varepsilon_{ik})$ independent and Subgaussian
## Overview of Existing Methods

### (Isotonic)-$\pi^*$
- Isotonicity in experts **for an unknown permutation** $\pi^*$
  - $M_{ik} \in [0, 1]$
  - $(\varepsilon_{ik})$ independent and Subgaussian

### (Bi-isotonic)-$\pi^*$
- Isotonicity in experts **for an unknown permutation** $\pi^*$
  - Isotonicity in questions: $M_{ik} \leq M_{(k+1)}$
  - $M_{ik} \in [0, 1]$
  - $(\varepsilon_{ik})$ independent and Subgaussian
(Isotonic)-$\pi^*$

- Isotonicity in experts for an unknown permutation $\pi^*$

- $M_{ik} \in [0, 1]$

- $(\varepsilon_{ik})$ independent and Subgaussian

(Bi-isotonic)-$\pi^*$

- Isotonicity in experts for an unknown permutation $\pi^*$

- Isotonicity in questions: $M_{ik} \leq M_{(k+1)}$

- $M_{ik} \in [0, 1]$

- $(\varepsilon_{ik})$ independent and Subgaussian
Introduction

Overview of Existing Methods

Minimax and Poly. Time Algo.

(Isotonic)-$\pi^*$

(Bi-isotonic)-$\pi^*$
(Isotonic)-$\pi^*$

(Bi-isotonic)-$\pi^*$
(Isotonic)-$\pi^*$

- Isotonicity in experts for an unknown permutation $\pi^*$

(Bi-isotonic)-$\pi^*$

- Isotonicity in experts for an unknown permutation $\pi^*$
- Isotonicity in questions: $M.k \leq M.(k+1)$
(Isotonic)-$\pi^*$
- Isotonicity in experts for an unknown permutation $\pi^*$

(Bi-isotonic)-($\pi^*$, $\sigma^*$)
- Isotonicity in experts for an unknown permutation $\pi^*$
- Isotonicity in questions for an unknown permutation $\sigma^*$

(Bi-isotonic)-$\pi^*$
- Isotonicity in experts for an unknown permutation $\pi^*$
- Isotonicity in questions: $M_k \leq M_{(k+1)}$
<table>
<thead>
<tr>
<th>(Isotonic)-$\pi^*$</th>
<th>(Bi-isotonic)-($\pi^<em>$, $\sigma^</em>$)</th>
<th>(Bi-isotonic)-$\pi^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ Isotonicity in experts for an unknown permutation $\pi^*$</td>
<td>▶ Isotonicity in experts for an unknown permutation $\pi^*$</td>
<td>▶ Isotonicity in experts for an unknown permutation $\pi^*$</td>
</tr>
<tr>
<td>▶ Isotonicity in questions for an unknown permutation $\sigma^*$</td>
<td>▶ Isotonicity in questions: $M.k \leq M.(k+1)$ (known permutation $\sigma^*$)</td>
<td>▶ Isotonicity in questions for an unknown permutation $\pi^*$</td>
</tr>
</tbody>
</table>
### Introduction

Overview of Existing Methods

Minimax and Poly. Time Algo.

<table>
<thead>
<tr>
<th>(Isotonic)-$\pi^*$</th>
<th>(Bi-isotonic)-($\pi^<em>$, $\sigma^</em>$)</th>
<th>(Bi-isotonic)-$\pi^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ Isotonicity in experts <strong>for an unknown permutation</strong> $\pi^*$</td>
<td>▶ Isotonicity in experts <strong>for an unknown permutation</strong> $\pi^*$</td>
<td>▶ Isotonicity in experts <strong>for an unknown permutation</strong> $\pi^*$</td>
</tr>
<tr>
<td>▶ Isotonicity in questions <strong>for an unknown permutation</strong> $\sigma^*$</td>
<td></td>
<td>▶ Isotonicity in questions: $M_k \leq M_{k+1}$ (known permutation $\sigma^*$)</td>
</tr>
</tbody>
</table>

Statistical difficulty:

$$(\text{Isotonic})-\pi^* \succ (\text{Bi-isotonic})-(\pi^*, \sigma^*) \succ (\text{Bi-isotonic})-\pi^*$$
References I

Rank analysis of incomplete block designs: I. the method of paired comparisons.

Partial recovery for top-k ranking: Optimality of mle and suboptimality of the spectral method.

Spectral method and regularized mle are both optimal for top-k ranking.
References II

Maximum likelihood estimation of observer error-rates using the em algorithm.

Optimal rates of statistical seriation.

Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation.
References III


References IV