Optimal Permutation Estimation in Crowd-Sourcing Problems

Emmanuel Pilliat

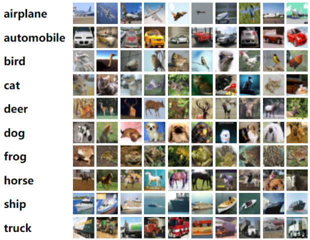
Université de Montpellier and INRAE

Based on a joint work with **Alexandra Carpentier** (Uni Potsdam) and **Nicolas Verzelen** (INRAE)

March, 21st 2023

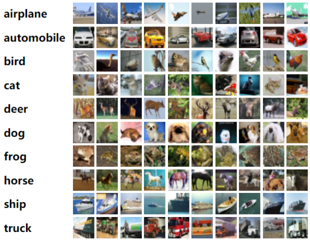
[arXiv:2211.04092] (2022)

Typical Dataset in Crowd-Sourcing



Cifar10H dataset: 10000 images, 10 labels.

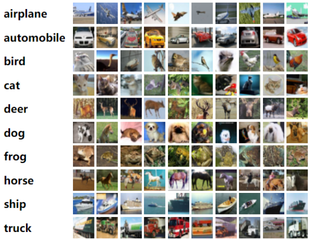
Typical Dataset in Crowd-Sourcing



Cifar10H dataset: 10000 images, 10 labels.

- Identification a worker: **annotator_id**

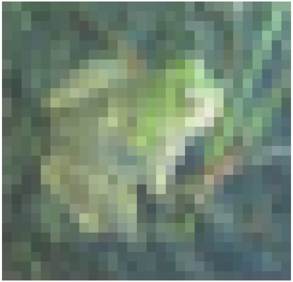
Typical Dataset in Crowd-Sourcing



Cifar10H dataset: 10000 images, 10 labels.

- Identification a worker: **annotator_id**
- Evaluation on a given image: **correct_guess**

Typical Dataset in Crowd-Sourcing



Frog (??)

- 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

$$
\begin{array}{cc}
 & \text{10 questions} \\
\text{4 experts} & \begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\end{array}
$$

0: Wrong answer   1: Correct answer

Example of Possible Data

$$
\begin{array}{cc}
 & \text{10 questions} \\
\text{4 experts} & \begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\end{array}
$$

0: Wrong answer 1: Correct answer

Bad Experts **Good Experts**

Example of Possible Data

$$
\begin{array}{cc}
 & \text{10 questions} \\
\text{4 experts} & \begin{pmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\end{array}
$$

0: Wrong answer   1: Correct answer

Hard Questions   **Easy Questions**

Example of Possible Data

$$
\begin{array}{cc}
 & \text{10 questions} \\
\text{4 experts} & \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}
\end{array}
$$

0: Wrong answer 1: Correct answer

This talk: **Ranking of Experts**

Example of Possible Data

$$
\begin{array}{cc}
 & \text{10 questions} \\
\text{4 experts} & \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}
\end{array}
$$

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:

$$Y_{ik} \sim \mathrm{Bern}\left(M_{ik}\right) \quad .$$

1: Correct 0: Wrong

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

Experts/Questions Setting

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

$$Y_{ik} \sim \mathrm{Bern}\left(M_{ik}\right) \quad .$$

expert i correct at question k

$$\Leftrightarrow Y_{ik} = 1 \quad .$$

1: Correct 0: Wrong

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

Experts/Questions Setting

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

$$Y_{ik} \sim \mathrm{Bern}\left(M_{ik}\right) \quad .$$

expert i correct at question k

$$\Leftrightarrow Y_{ik} = 1 \quad .$$

1: Correct   0: Wrong

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

- $M_{ik} = 1/2$: random choice of expert i at question k
- $M_{ik} = 1$: 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

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

- (ε_{ik}) i.i.d. subGaussian (e.g. Bernoulli)

Statistical Models

Observation Model

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

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

Statistical Models

Observation Model

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

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

Parametric Models for M:

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

Statistical Models

Observation Model

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

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

Parametric Models for M:

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

Non-Parametric Models for M ≈[Mao et al., 2018]

- Increasing Rows: $M_{i,k} \leq M_{i,k+1}$

Statistical Models

Observation Model

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

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

Parametric Models for M:

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

Non-Parametric Models for M   ≈[Mao et al., 2018]

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

Statistical Models

Observation Model

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

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

Parametric Models for M:

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

Non-Parametric Models for M ≈[Mao et al., 2018]

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

Non Parametric Model

Observation Model

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

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

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

White = 0 ; Black = 1

Matrix M_{π^*}. (isotonic).

Non Parametric Model

Observation Model

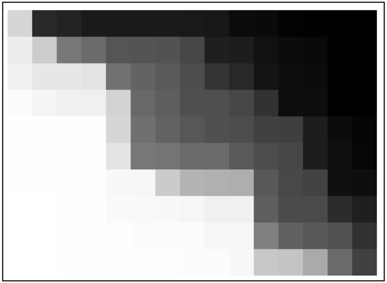
$Y = M + \varepsilon \in \mathbb{R}^{n \times d}$

- (ε_{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** π^*

White $= 0$; Black $= 1$



Matrix M_{π^*}. (isotonic).

Non Parametric Model

Observation Model

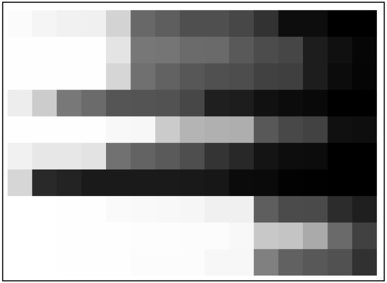
$Y = M + \varepsilon \in \mathbb{R}^{n \times d}$

- (ε_{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** π^*

White $= 0$; Black $= 1$



Matrix M (isotonic up to a permutation of experts).

Non Parametric Model

Observation Model

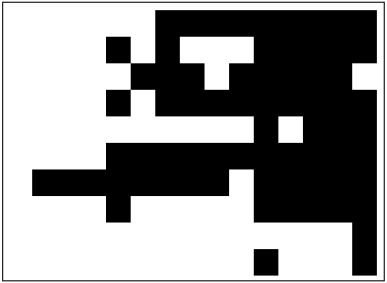
$Y = M + \varepsilon \in \mathbb{R}^{n \times d}$

- (ε_{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** π^*

White $= 0$; Black $= 1$



Matrix Y (M in noise).

Non Parametric Model

Observation Model

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

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

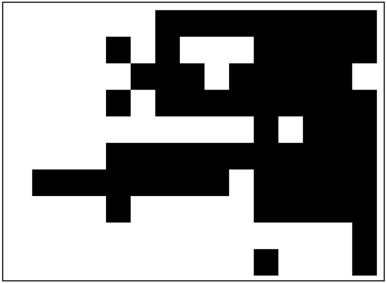
Shape Constraints (Bi-isotonicity):

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

Aim

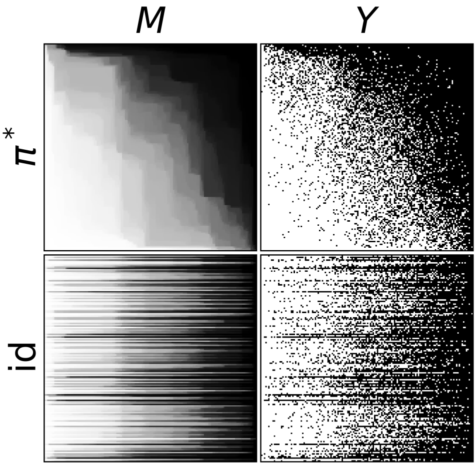
Estimation of π^*.

White $= 0$; Black $= 1$

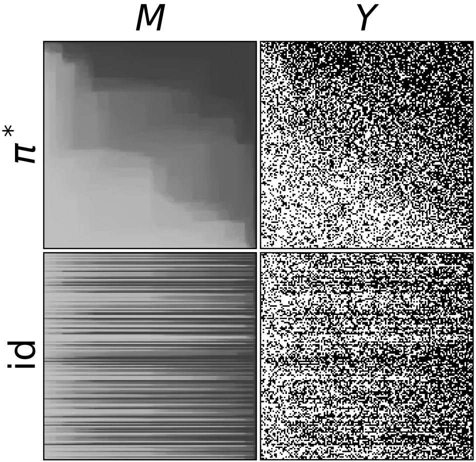


Matrix Y (M in noise).

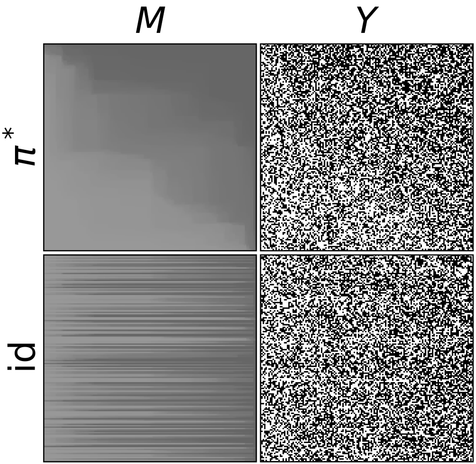
Example with $n, d = 150$, $M \in [0, 1]$



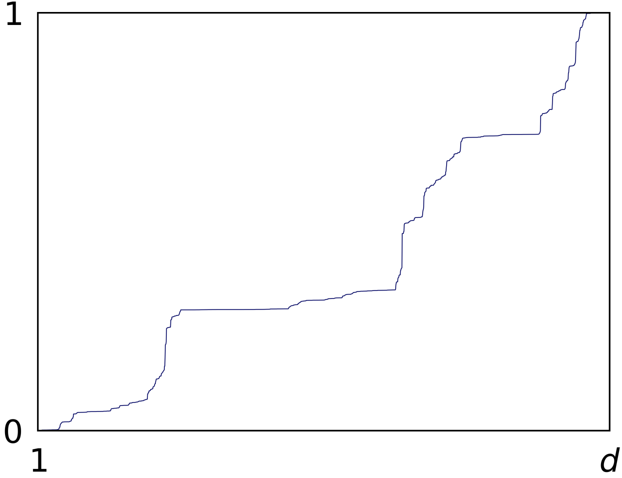
Example with $n, d = 150,\ M \in [0.25, 0.75]$



Example with $n, d = 150,\ M \in [0.4, 0.6]$

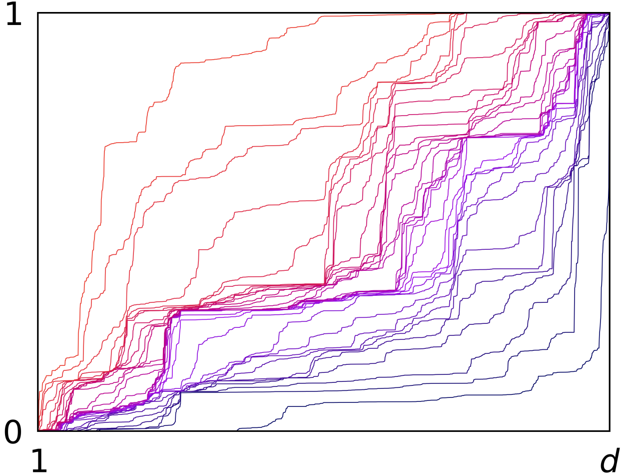


Bi-isotonic M - Other representation



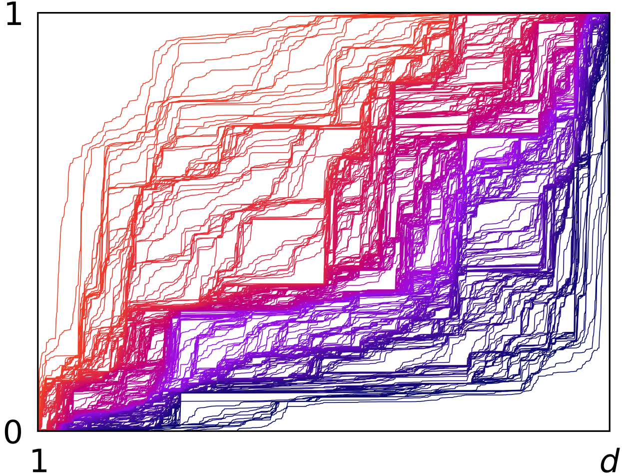
Each line $M_{i,\cdot}$ represents an expert i

Bi-isotonic M - Other representation



Each line $M_{i,\cdot}$ represents an expert i

Bi-isotonic M - Other representation



Each line $M_{i,\cdot}$ represents an expert i

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

Error Measures

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Non Parametric Model

Observation Model

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

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

Shape Constraints (Bi-isotonicity):

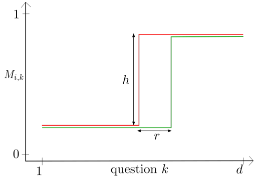
- Increasing Rows $M_{i,k} \leq M_{i,k+1}$
- Increasing Columns **for an unknown permutation** π^*

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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



If the two lines are misclassified: Perm-Loss $= 2rh^2$

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Aim

Estimation of π^*.

Non Parametric Model

Observation Model

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

- (ε_{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** π^*

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Aim

Estimation of π^*.

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Aim

Estimation of π^*.

MiniMax-Risk

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Aim

Estimation of π^*.

MiniMax-Risk

Max-Risk and MiniMax-Risk

If $\hat{\pi}$ is an estimator of π^*, we define

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

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

Error Measures

Permutation loss

For an estimator $\hat{\pi}$ of π^*

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

Estimation loss

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

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

Aim

Estimation of π^*.

MiniMax-Risk

Max-Risk and MiniMax-Risk

If $\hat{\pi}$ is an estimator of π^*, we define

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

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

Define similarly Max-Estim and 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 π^* and σ^* 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 π^* and σ^* 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 π^* and σ^* 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 π^* and σ^* 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 π^* and σ^* 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 there a **computational-statistical gap**?

Main questions

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

Main questions

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

Our Contributions

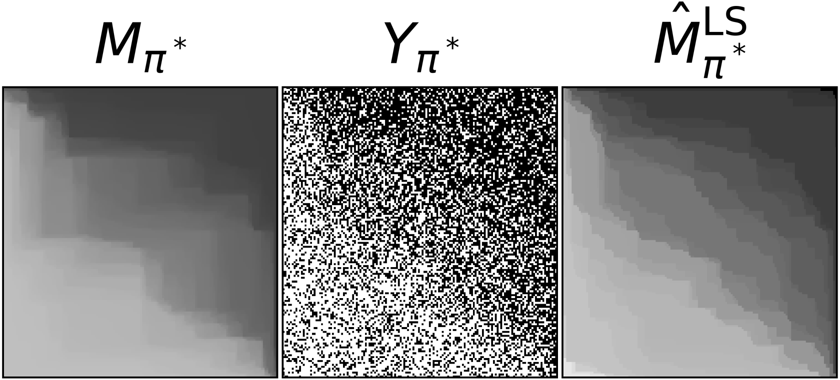
For all n, d:

- Control of MiniMax-Perm
- A polynomial-time procedure achieves MiniMax-Perm

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

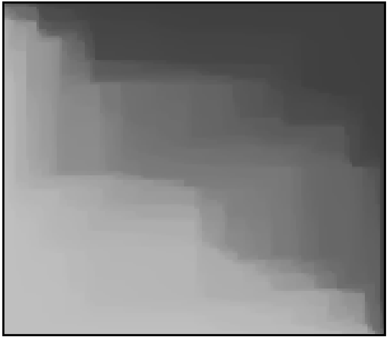
$$(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}}) = \underset{\widetilde{M} \in \mathrm{Mon}, \widetilde{\pi} \in \mathrm{Perm}}{\arg\min} \|\widetilde{M}_{\widetilde{\pi}} - Y\|_F^2$$

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

$$(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}}) = \underset{\widetilde{M}\in\mathrm{Mon},\widetilde{\pi}\in\mathrm{Perm}}{\arg\min} \|\widetilde{M}_{\widetilde{\pi}} - Y\|_F^2$$



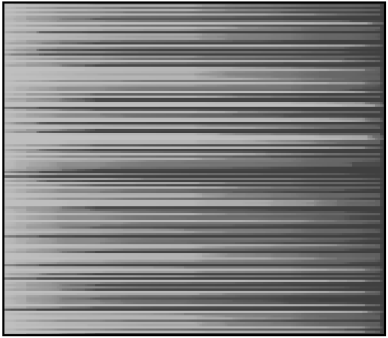
Matrix $M_{\pi^*,.}$ (bi-isotonic).

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

$$(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}}) = \underset{\widetilde{M} \in \mathrm{Mon}, \widetilde{\pi} \in \mathrm{Perm}}{\arg\min} \|\widetilde{M}_{\widetilde{\pi}} - Y\|_F^2$$



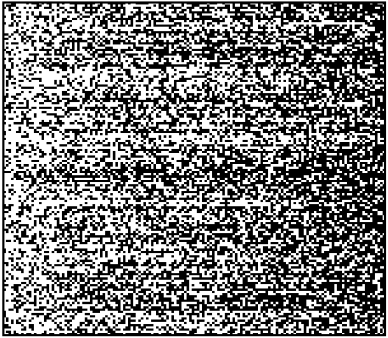
Matrix M.

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

$$(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}}) = \underset{\widetilde{M} \in \mathrm{Mon}, \widetilde{\pi} \in \mathrm{Perm}}{\arg\min} \|\widetilde{M}_{\widetilde{\pi}} - Y\|_F^2$$



Matrix Y.

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

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

Least-square estimator

$$(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}}) = \operatorname*{arg\,min}_{\widetilde{M} \in \mathrm{Mon}, \widetilde{\pi} \in \mathrm{Perm}} \|\widetilde{M}_{\widetilde{\pi}} - Y\|_F^2$$

No know polynomial-time method to compute $(\hat{M}^{\mathrm{LS}}, \hat{\pi}^{\mathrm{LS}})$

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

Least-square guarantees

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

$$\text{Max-Estim}(\hat{M}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$
$$\text{Max-Perm}(\hat{\pi}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$

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

Least-square guarantees

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

$$\text{Max-Estim}(\hat{M}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$
$$\text{Max-Perm}(\hat{\pi}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$

Entropy Arguments:

- $n!$ permutations: $n \asymp \log(n!)$

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

Least-square guarantees

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

$$\text{Max-Estim}(\hat{M}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$
$$\text{Max-Perm}(\hat{\pi}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$

Entropy Arguments:

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

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

Least-square guarantees

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

$$\text{Max-Estim}(\hat{M}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$
$$\text{Max-Perm}(\hat{\pi}^{\mathrm{LS}}) \lesssim n \vee (\sqrt{nd} \wedge nd^{1/3})$$

Entropy Arguments:

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

Remarks:

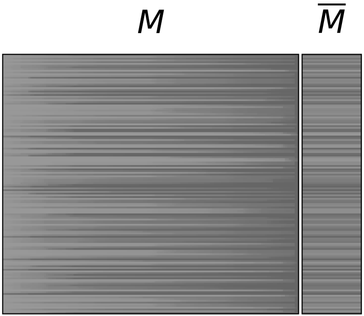
- MiniMax-Estim Optimal [Mao et al., 2018]
- not proven to be MiniMax-Perm Optimal

Summary

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	??	??	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n

But algo. not polynomial time.

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



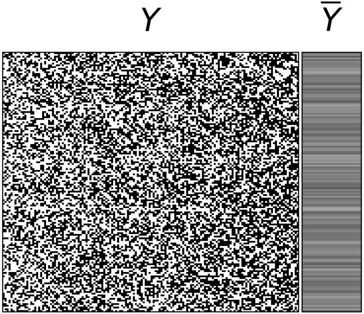
Matrix M.

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

Method:

- Compute expert i average performances on all questions:

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

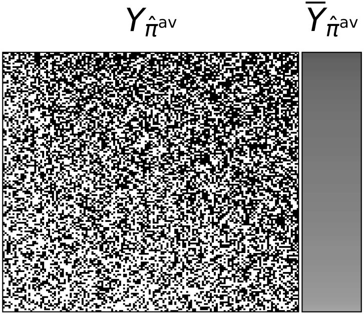


Matrix Y (M in noise).

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

Method:

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



Matrix $Y_{\hat{\pi}^{\mathrm{av}}}$ (M in noise).

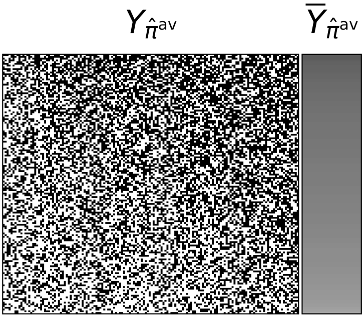
Global Average Comparison [Pananjady and Samworth, 2020, Shah et al., 2019]

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: $\hat{\pi}^{\mathrm{av}}$

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

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



Matrix $Y_{\hat{\pi}^{\mathrm{av}}}$ (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: $\hat{\pi}^{\mathrm{av}}$

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

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

Idea of Proof

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

$$M_{1,\cdot} = (.5 .5 .5 .5 .5 \dots .5 .5 \,\underbrace{1 1 1 1 1 1 1 1 1 1}\,)$$

$$M_{2,\cdot} = (.5 .5 .5 .5 .5 \dots .5 .5 \,\underbrace{.5 .5 .5 .5 .5 .5 .5}_{\sim\sqrt{d}}\,)$$

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: $\hat{\pi}^{\mathrm{av}}$

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

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

Idea of Proof

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

$$Y_{1,.} = (01101\ldots10\,\underbrace{1111111111})$$

$$Y_{2,.} = (01000\ldots01\,\underbrace{1010010100}_{\sim\sqrt{d}})$$

(Example of Observations)

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

Method:

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

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

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

Idea of Proof

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

$$Y_{1,\cdot} = (01101\ldots10\,\underbrace{1111111111})$$

$$Y_{2,\cdot} = (01000\ldots01\,\underbrace{1010010100}_{\sim\sqrt{d}})$$

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

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: $\hat{\pi}^{\mathrm{av}}$

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

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

Idea of Proof

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

$$Y_{1,.} = (01101\dots10\,\underbrace{1111111111})$$

$$Y_{2,.} = (01000\dots01\,\underbrace{1010010100}_{\sim\sqrt{d}})$$

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

- **Lower Bound** for $\hat{\pi}^{\mathrm{av}}$: There exists M s.t. $\text{Max-Perm}(\hat{\pi}^{\mathrm{av}}) \gtrsim n\sqrt{d}$

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: $\hat{\pi}^{\mathrm{av}}$

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

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

Idea of Proof

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

$$Y_{1,\cdot} = (01101\dots10\,\underbrace{1111111111})$$

$$Y_{2,\cdot} = (01000\dots01\,\underbrace{1010010100}_{\sim\sqrt{d}})$$

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

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

Summary

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	??	??	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n
Global average (UB)	$n\sqrt{d}$	$n\sqrt{d}$	$n\sqrt{d}$

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 $n \vee d$ through this. **Optimal** for $d = n$ in which case

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

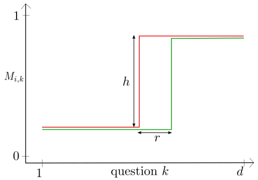
CP and Hierarchical Clustering Based Algo. [Liu and Moitra, 2020]

[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 $n \vee d$ through this. **Optimal** for $d = n$ in which case

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



Localisation through CP detection.

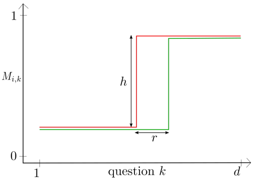
CP and Hierarchical Clustering Based Algo. [Liu and Moitra, 2020]

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

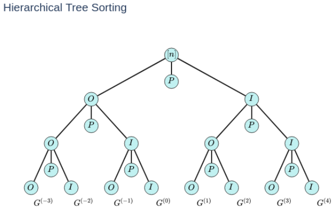
$$\text{Max-Perm}(\hat{\pi}^{(LM)}) \lesssim n.$$

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

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



Localisation through CP detection.



Hierarchical clustering.

Summary

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	??	??	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n
Global average (UB)	$n\sqrt{d}$	$n\sqrt{d}$	$n\sqrt{d}$
Ext. of LM (UB)	d	d	n

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] - accepted in AOS

Assume we have polylog samples.
There exists a estimator $\hat{\pi}$ of π^* **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] - accepted in AOS

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

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

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n

Minimax and Poly. Time

Theorem [P., Carpentier, Verzelen, 2022] - accepted in AOS

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

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

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n

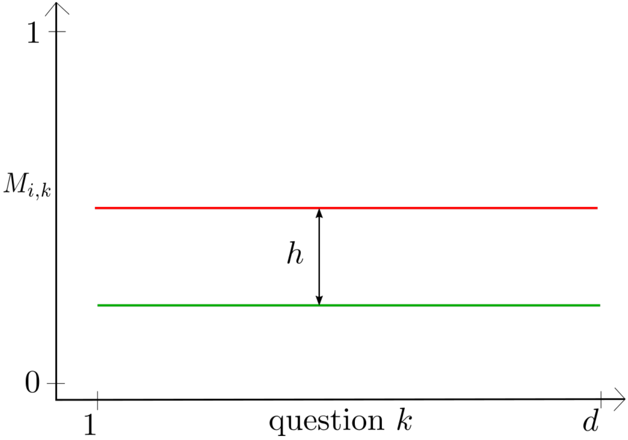
Can be combined with bi-isotonic regression to have a **poly. time MiniMax-Estim algo**!

Summary

Poly. time algo achieving the minimax rates:

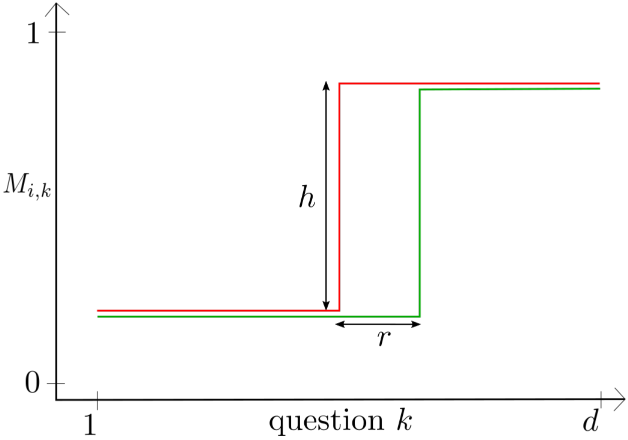
	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n

Uniform distance between two experts



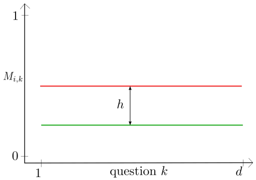
Global average comparison is optimal:
Constant Perm-Risk - Confusion only if $h \lesssim 1/\sqrt{d}$.

Localised distance between two experts

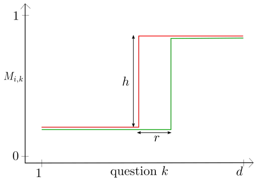


$\psi([d]) = \frac{1}{d}\sum_{i=1}^{d} Y_i$ achieves
Perm-Risk $\asymp \sqrt{d} \gg d^{1/6}$

From Global to Local Averages

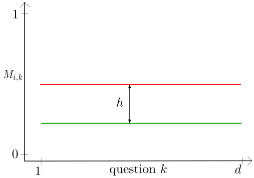


Global average good.

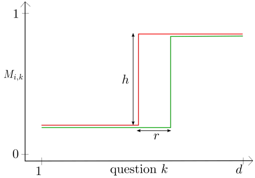


Global average bad → need to **localise**.

From Global to Local Averages



Global average good.

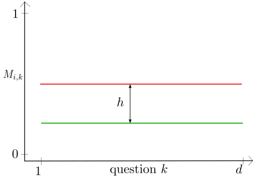


Global average bad $\rightarrow$ need to **localise**.

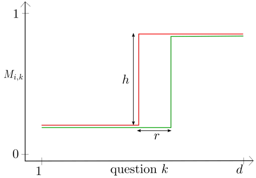
Idea:

- Estimate by a **change point** (CP) method **windows** where any of the two experts changes by more than h.

From Global to Local Averages



Global average good.

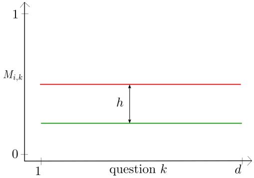


Global average bad → need to **localise**.

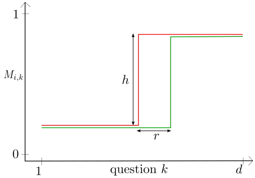
Idea:

- 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.



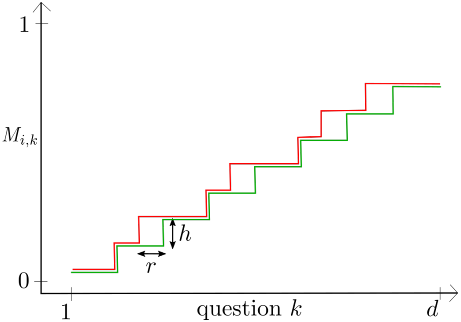
Global average bad → need to **localise**.

Idea:

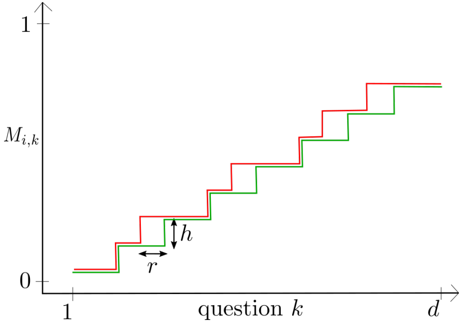
- Estimate by a **change point** (CP) method **windows** where any of the two experts changes by more than h.
- Compute **local average** on these **windows**.

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

Toward a Worst Case Scenario



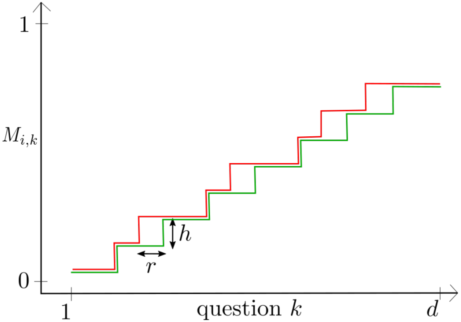
Toward a Worst Case Scenario



Idea:

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

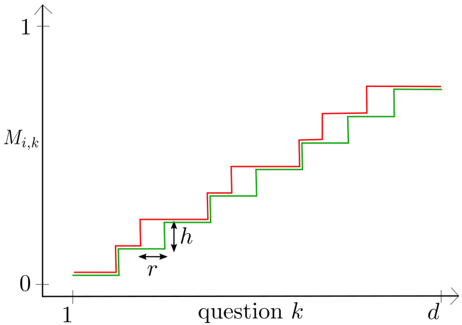
Toward a Worst Case Scenario



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 $M \in [0, 1]$

Toward a Worst Case Scenario

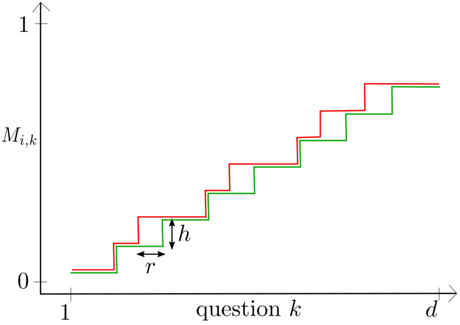


Idea:

- 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:

$$\begin{aligned}\|M_{1\cdot} - M_{2\cdot}\|_2^2 &\leq h\|M_{1\cdot} - M_{2\cdot}\|_1 \\ &\leq h\sqrt{\frac{1}{h^2}\frac{1}{h} \wedge d} \\ &\leq d^{1/6} \quad .\end{aligned}$$

Toward a Worst Case Scenario



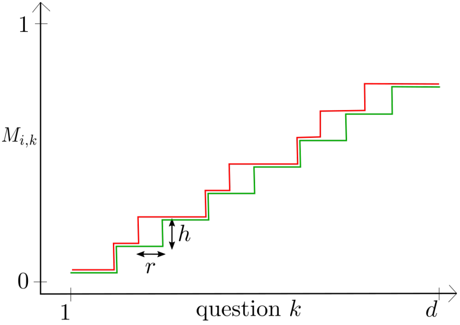
Idea:

- 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:

$$\begin{aligned}\|M_{1\cdot} - M_{2\cdot}\|_2^2 &\le h\|M_{1\cdot} - M_{2\cdot}\|_1 \\ &\le h\sqrt{\frac{1}{h^2}\frac{1}{h} \wedge d} \\ &\le d^{1/6} \ .\end{aligned}$$

- $d^{1/6}$ is optimal for two experts: 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:

$$\begin{aligned}\|M_{1\cdot} - M_{2\cdot}\|_2^2 &\leq h\|M_{1\cdot} - M_{2\cdot}\|_1\\ &\leq h\sqrt{\frac{1}{h^2}\frac{1}{h} \wedge d}\\ &\leq d^{1/6}\ .\end{aligned}$$

- $d^{1/6}$ is optimal for two experts: MiniMax-Perm $\asymp d^{1/6}$.
- For any n (UB): MiniMax-Perm $\lesssim nd^{1/6}$.

Summary

Poly. time algo achieving the minimax rates:

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n
Global average (UB)	$n\sqrt{d}$	$n\sqrt{d}$	$n\sqrt{d}$
Ext. of LM (UB)	d	d	n

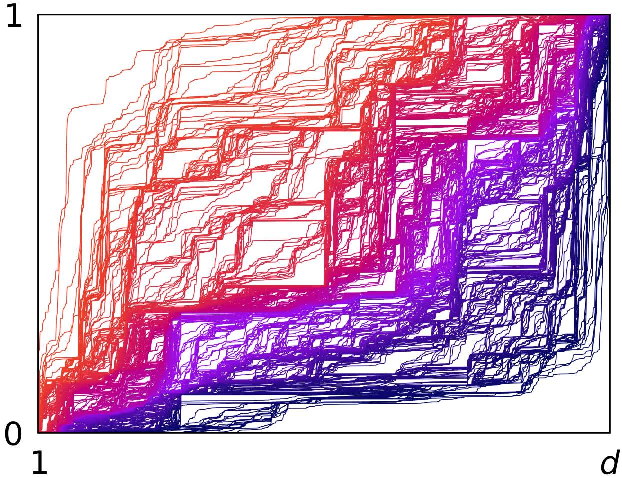
Ext. of LM (UB) extends [Liu and Moitra, 2020] to $d \neq n$

Summary

Poly. time algo achieving the minimax rates:

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n
Global average (UB)	$n\sqrt{d}$	$n\sqrt{d}$	$n\sqrt{d}$
Ext. of LM (UB)	d	d	n

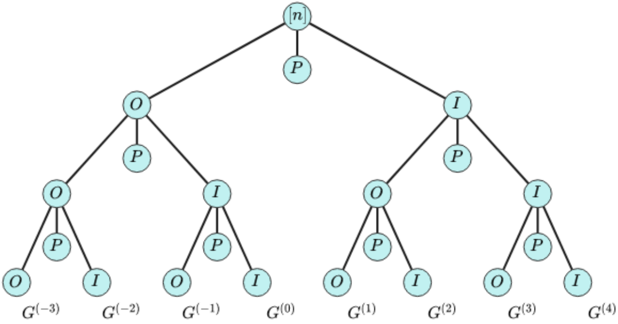
Ext. of LM (UB) extends [Liu and Moitra, 2020] to $d \neq n$


1
0
1
d

Hierarchical Clustering

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

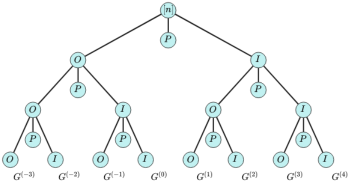
Hierarchical Tree Sorting



Hierarchical Clustering

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

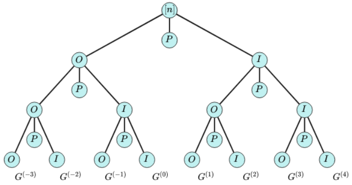
Hierarchical Tree Sorting

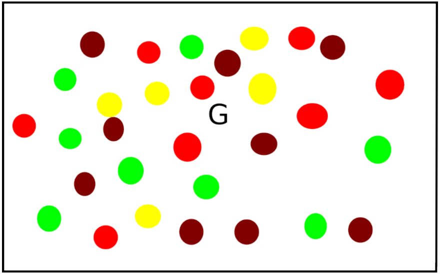


Hierarchical Clustering

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

Hierarchical Tree Sorting

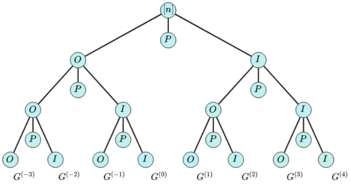


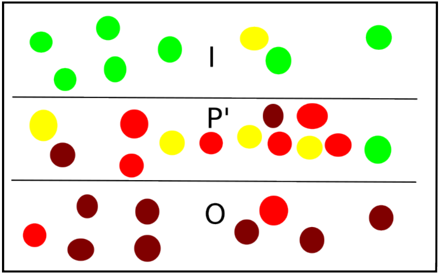


Hierarchical Clustering

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

Hierarchical Tree Sorting

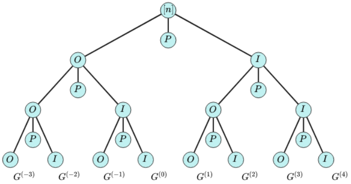


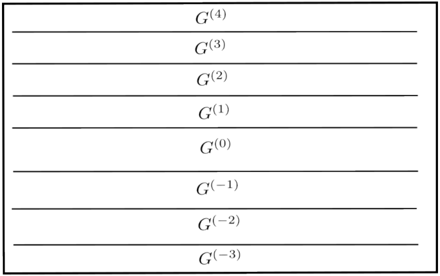


Hierarchical Clustering

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

Hierarchical Tree Sorting

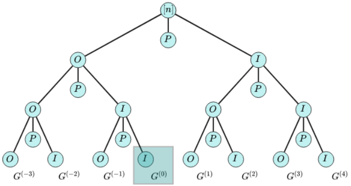


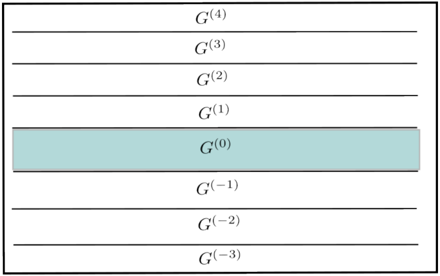


Hierarchical Clustering

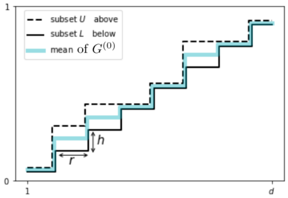
Beyond [Liu and Moitra, 2020] for $d \neq n$

Hierarchical Tree Sorting

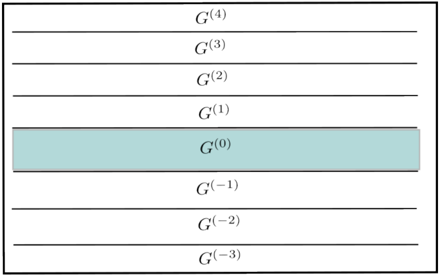




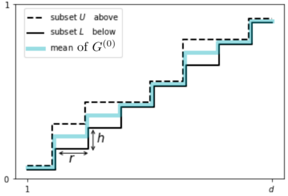
Worst Case for a Group $G^{(0)}$ $(n \gg d^{1/3})$



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



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



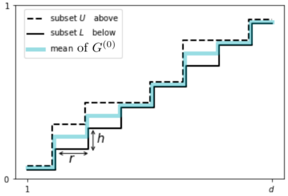
In $G^{(0)}$, an expert is either in U or in L.

After Aggregation

$$\frac{\sqrt{r}h}{2}\begin{pmatrix} 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

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 U or in L.

After Aggregation

$$\frac{\sqrt{r}h}{2}\begin{pmatrix} 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

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

Rank one matrix $\rightsquigarrow$ **(PCA)**:
1^{st} 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

$$n \vee \left(n^{2/3} d^{1/3}\right) \ .$$

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

The corresponding Max-Perm is upper bounded by

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

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

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

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

The corresponding Max-Perm is upper bounded by

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

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

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

- But not Optimal !

$$n \vee (n^{2/3}d^{1/3}) \gg n \vee (n^{3/4}d^{1/4}) \ .$$

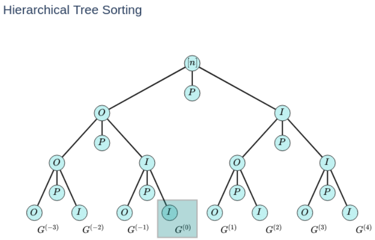
Summary

Poly. time algo achieving the minimax rates:

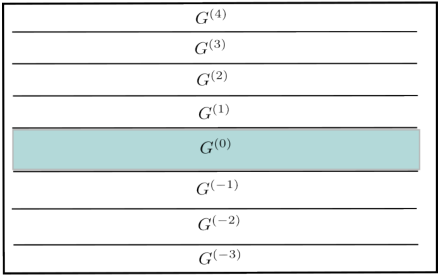
	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n
MiniMax-Estim	$nd^{1/3}$	$\sqrt{nd}$	n
Global average (UB)	$n\sqrt{d}$	$n\sqrt{d}$	$n\sqrt{d}$
Ext. of LM (UB)	d	d	n
Super ext. of LM	$nd^{1/6}$	$n^{2/3}d^{1/3}$	n

Remark: Super ext. of LM requires a lot of additional work w.r.t. [Liu and Moitra, 2020]

Ideas to achieve $n^{3/4}d^{1/4}$

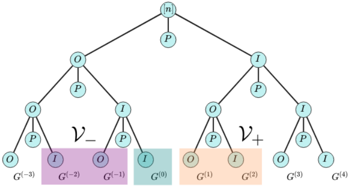


From an oblivious Hierarchical Clustering



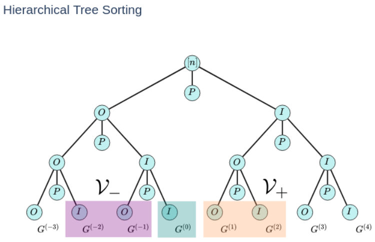
Ideas to achieve $n^{3/4}d^{1/4}$

Hierarchical Tree Sorting

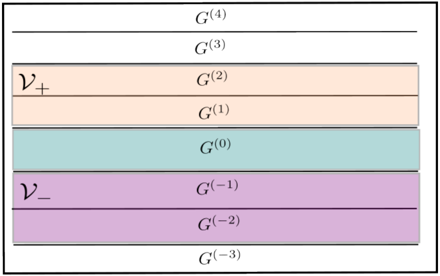


To using the Memory of the Tree

Ideas to achieve $n^{3/4}d^{1/4}$

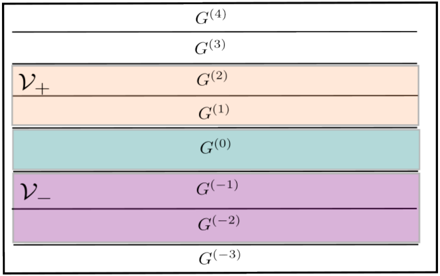


To using the Memory of the Tree



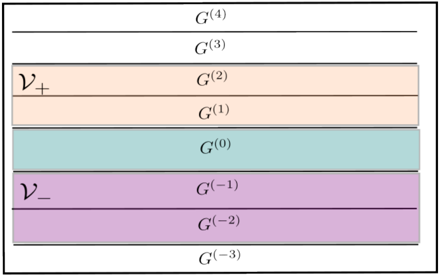
$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$

Two Types of Information



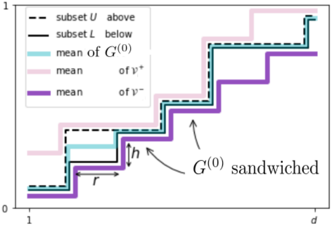
$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$

Two Types of Information



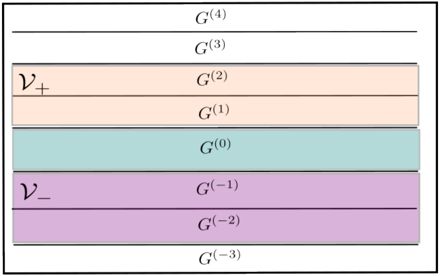
$G^{(0)}$ is sandwiched between $\mathcal{V}_-$ and $\mathcal{V}_+$

First Type



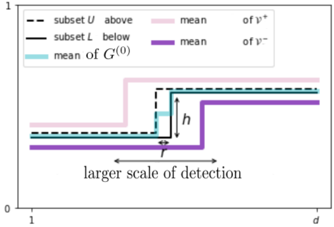
Removing regions where $G^{(0)}$ is sandwiched

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:

	$n \lesssim d^{1/3}$	$d^{1/3} \lesssim n \lesssim d$	$d \lesssim n$
MiniMax-Perm	$nd^{1/6}$	$n^{3/4}d^{1/4}$	n

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] (2022)

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] (2022)

Research Directions:

- Removing the isotonicity constraint on questions.
- Unknown answers: -observing labels instead of correctness.

References I

Bradley, R. A. and Terry, M. E. (1952).
Rank analysis of incomplete block designs: I. the method of paired comparisons.
Biometrika, 39(3/4):324–345.

Chen, P., Gao, C., and Zhang, A. Y. (2022).
Partial recovery for top-k ranking: Optimality of mle and suboptimality of the spectral method.
The Annals of Statistics, 50(3):1618–1652.

Chen, Y., Fan, J., Ma, C., and Wang, K. (2019).
Spectral method and regularized mle are both optimal for top-k ranking.
Annals of statistics, 47(4):2204.

References II

Dawid, A. P. and Skene, A. M. (1979).
Maximum likelihood estimation of observer error-rates using the em algorithm.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1):20–28.

Flammarion, N., Mao, C., and Rigollet, P. (2019).
Optimal rates of statistical seriation.

Liu, A. and Moitra, A. (2020).
Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation.
In Abernethy, J. and Agarwal, S., editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2780–2829. PMLR.

References III

Mao, C., Pananjady, A., and Wainwright, M. J. (2018).
Towards optimal estimation of bivariate isotonic matrices with unknown permutations.
arXiv preprint arXiv:1806.09544.

Pananjady, A. and Samworth, R. J. (2020).
Isotonic regression with unknown permutations: Statistics, computation, and adaptation.
arXiv preprint arXiv:2009.02609.

Shah, N., Balakrishnan, S., Guntuboyina, A., and Wainwright, M. (2016).
Stochastically transitive models for pairwise comparisons: Statistical and computational issues.
In *International Conference on Machine Learning*, pages 11–20. PMLR.

References IV

Shah, N. B., Balakrishnan, S., and Wainwright, M. J. (2019).
Feeling the bern: Adaptive estimators for bernoulli probabilities of pairwise comparisons.
IEEE Transactions on Information Theory, 65(8):4854–4874.

(Isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*

- $M_{ik} \in [0, 1]$
- (ε_{ik}) independent and Subgaussian

(Isotonic)-π^*

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

(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$
- $M_{ik} \in [0, 1]$
- (ε_{ik}) independent and Subgaussian

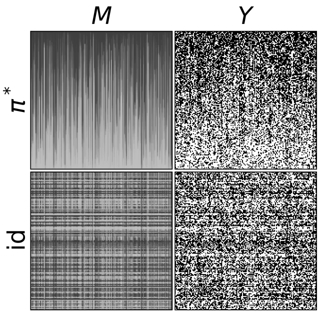
(Isotonic)-π^*

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

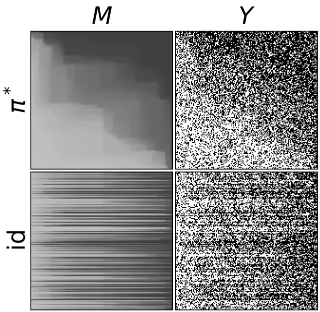
(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$
- $M_{ik} \in [0, 1]$
- (ε_{ik}) independent and Subgaussian

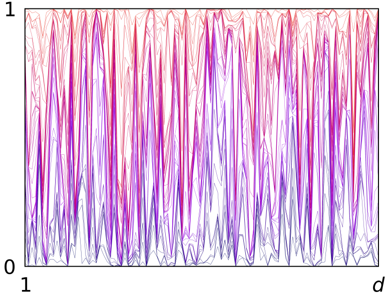
(Isotonic)-π^*



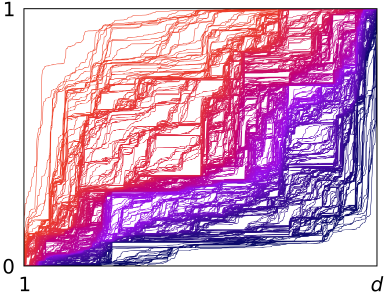
(Bi-isotonic)-π^*



(Isotonic)-π^*



(Bi-isotonic)-π^*



(Isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*

(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$

(Isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*

(Bi-isotonic)-(π^*, σ^*)

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions **for an unknown permutation** σ^*

(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$

(Isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*

(Bi-isotonic)-(π^*, σ^*)

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions **for an unknown permutation** σ^*

(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$ **(known permutation** σ^***)**

(Isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*

(Bi-isotonic)-(π^*, σ^*)

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions **for an unknown permutation** σ^*

(Bi-isotonic)-π^*

- Isotonicity in experts **for an unknown permutation** π^*
- Isotonicity in questions: $M_{\cdot k} \leq M_{\cdot (k+1)}$ **(known permutation** σ^***)**

Statistical difficulty:

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