Permutation Estimation for Crowdsourcing

joint work with Alexandra Carpentier and Nicolas Verzelen

Maximilian Graf, Universität Potsdam, 11 March 2025
supported by the DFG research unit FOR 5381

Part 1: Theoretical Overview

Motivating Example

Crowdsourcing: “the activity of getting information or help for a project or a task from a large number of people, typically using the internet” — `www.oxfordlearnersdictionaries.com`

Motivating Example

Crowdsourcing: “the activity of getting information or help for a project or a task from a large number of people, typically using the internet” — `www.oxfordlearnersdictionaries.com`

- group of experts works on several tasks
- for each expert/task: observe success or failure

Motivating Example

Crowdsourcing: “the activity of getting information or help for a project or a task from a large number of people, typically using the internet” — `www.oxfordlearnersdictionaries.com`

- group of experts works on several tasks
- for each expert/task: observe success or failure

Questions:

Can we sort experts by quality and tasks by difficulty?

Can we recover the probability of experts succeeding on tasks?

Mathematical Model

- n: number of experts, d: number of tasks
- Y: $n \times d$ observation matrix
  $M = \mathbb{E}[Y] \in [0,1]^{n \times d}$

Mathematical Model

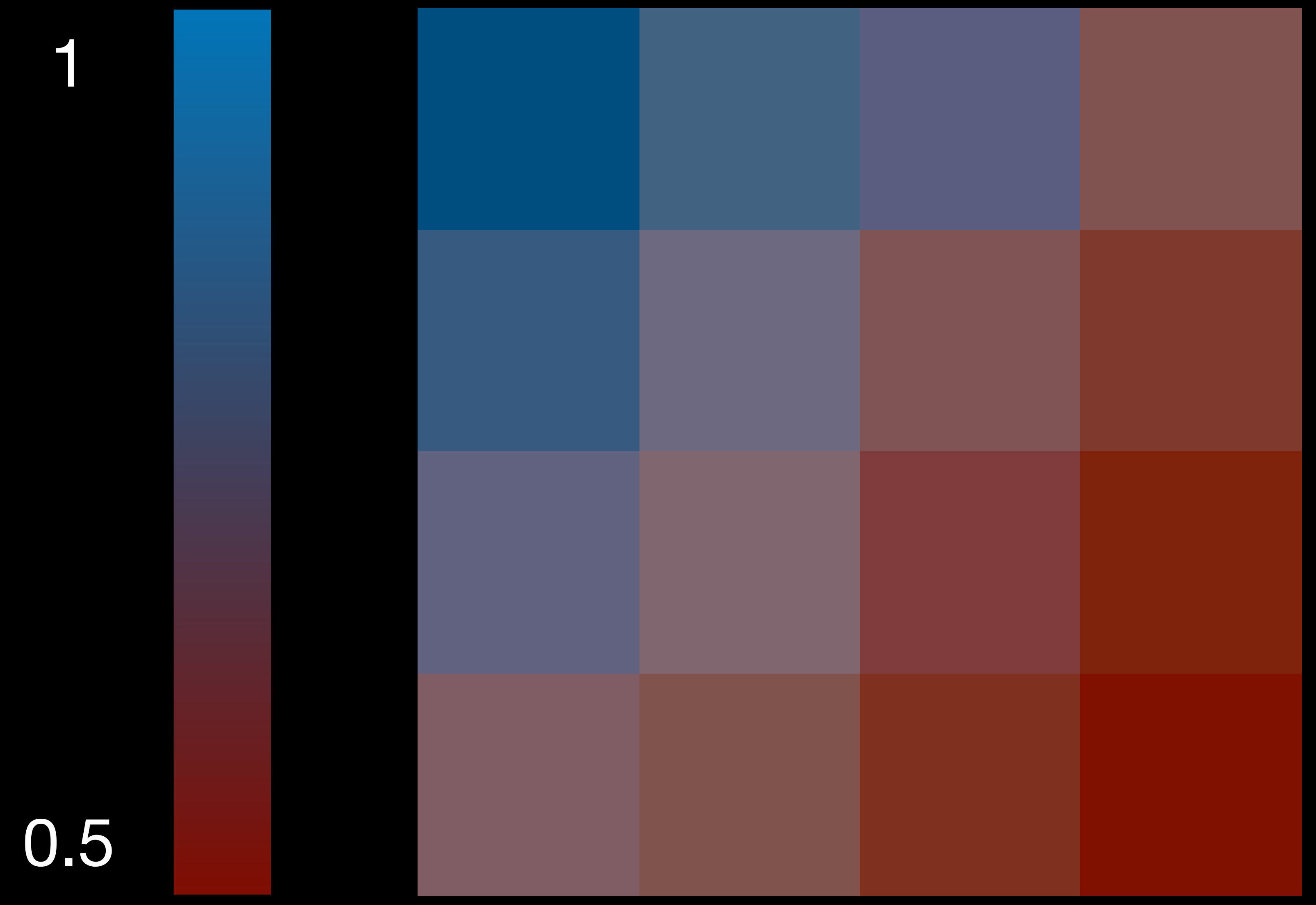
- n: number of experts, d: number of tasks
- Y: $n \times d$ observation matrix
  $M = \mathbb{E}[Y] \in [0,1]^{n \times d}$

Assumptions:

- 1-sub-Gaussian noise W_{ij} such that $Y_{ij} = M_{ij} + W_{ij}$
- rankings π of experts and η of tasks exist such that
  $M_{ij} = N_{\pi(i)\eta(j)}$ for a bi-isotonic matrix N
- N bi-isotonic: each row and column is decreasing

Mathematical Model

$N \in [0,1]^{n\times d}$ bi-isotonic: $N \in \mathbb{C}_{\mathrm{Biso}}$



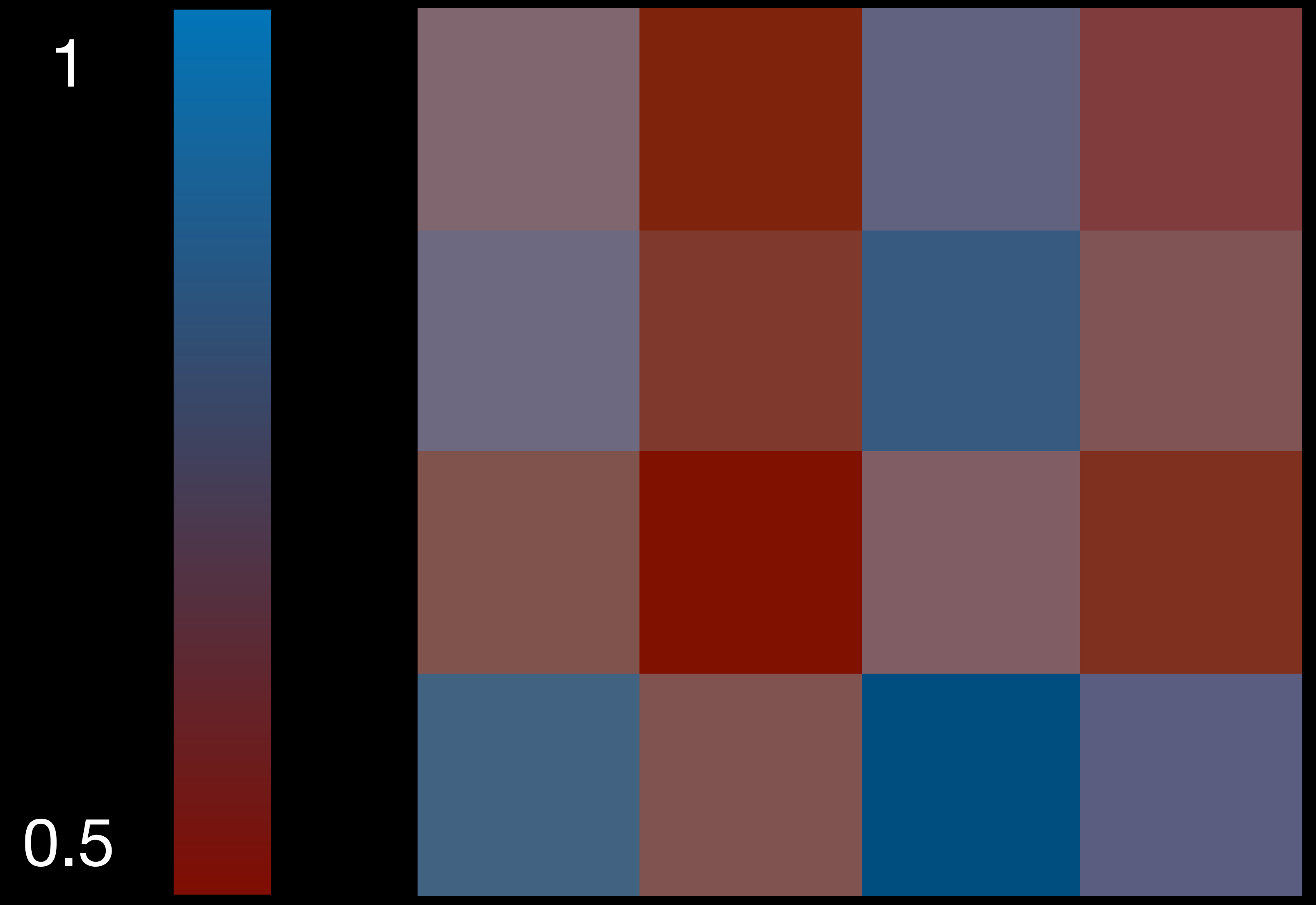
- n: number of experts, d: number of tasks
- Y: $n \times d$ observation matrix
  $M = \mathbb{E}[Y] \in [0,1]^{n\times d}$

Assumptions:

- 1-sub-Gaussian noise W_{ij} such that $Y_{ij} = M_{ij} + W_{ij}$
- rankings π of experts and η of tasks exist such that $M_{ij} = N_{\pi(i)\eta(j)}$ for a bi-isotonic matrix N
- N bi-isotonic: each row and column is decreasing

Mathematical Model

$M \in [0,1]^{n\times d}$ bi-isotonic up to row-/column-permutation: $M \in \mathbb{C}_{\mathrm{Biso}}^{\mathrm{Perm}}$



- n: number of experts, d: number of tasks
- Y: $n \times d$ observation matrix
  $M = \mathbb{E}[Y] \in [0,1]^{n\times d}$

Assumptions:

- 1-sub-Gaussian noise W_{ij} such that $Y_{ij} = M_{ij} + W_{ij}$
- rankings π of experts and η of tasks exist such that $M_{ij} = N_{\pi(i)\eta(j)}$ for a bi-isotonic matrix N
- N bi-isotonic: each row and column is decreasing

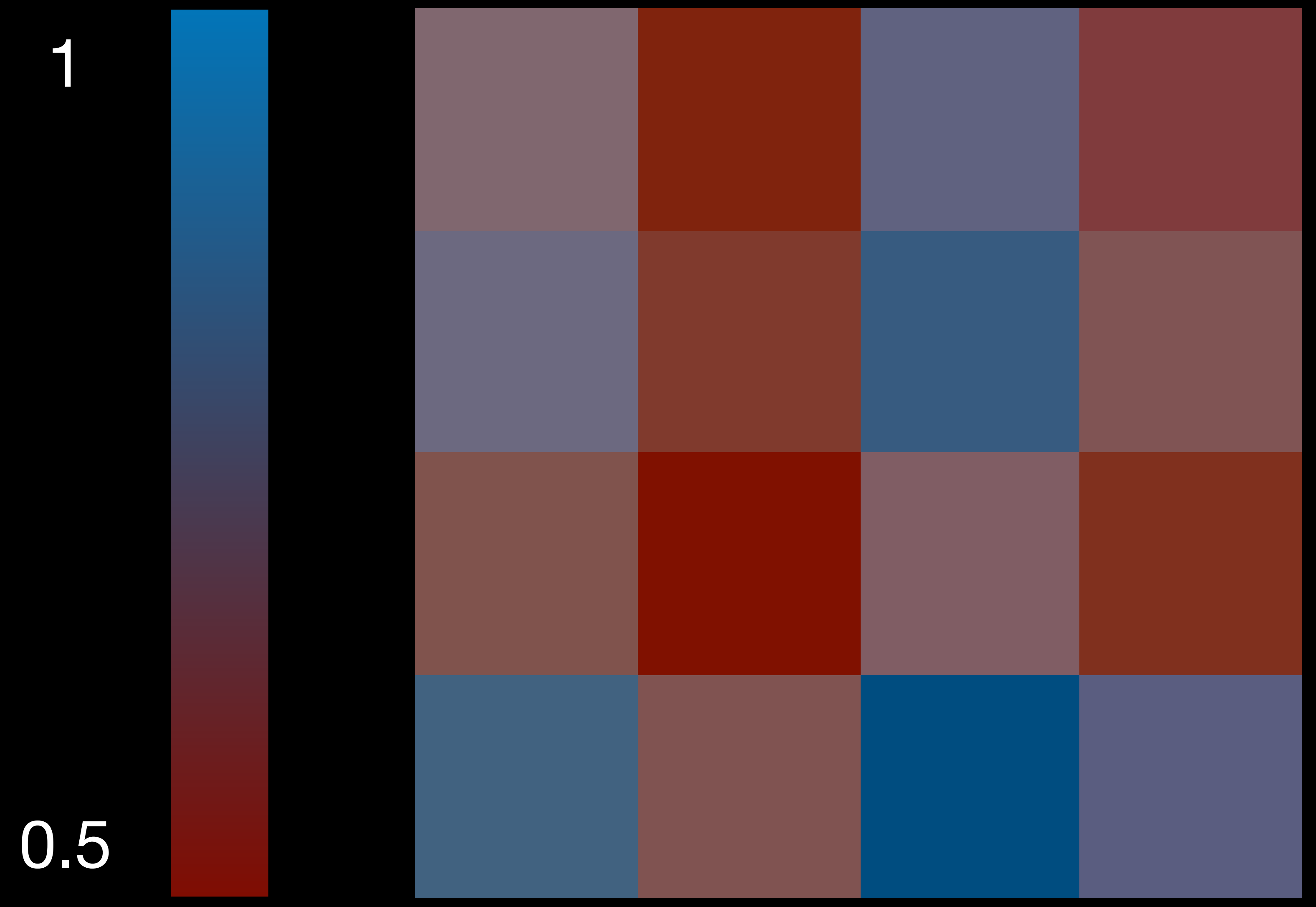
Mathematical Model

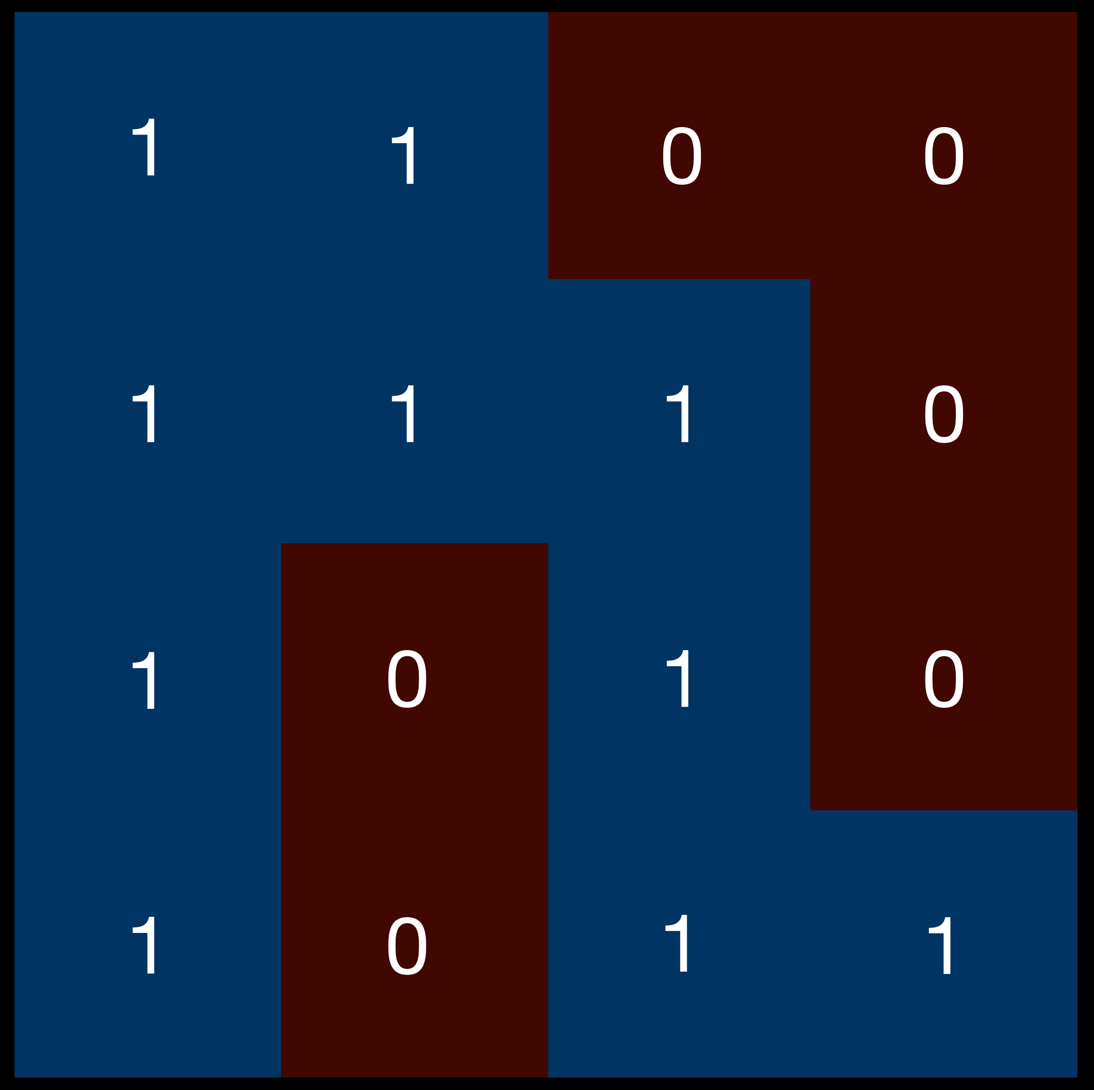
- n: number of experts, d: number of tasks
- Y: $n \times d$ observation matrix
  $M = \mathbb{E}[Y] \in [0,1]^{n \times d}$

Assumptions:

- 1-sub-Gaussian noise W_{ij} such that $Y_{ij} = M_{ij} + W_{ij}$
- rankings π of experts and η of tasks exist such that
  $M_{ij} = N_{\pi(i)\eta(j)}$ for a bi-isotonic matrix N
- N bi-isotonic: each row and column is decreasing

Y noisy version of M:
here $Y_{ij} = \mathrm{Bern}(M_{ij})$





Previous Work

Minimax Optimal Estimation [Mao et al., 2020]

- focus on estimating $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}$ and reconstruction error $\mathbb{E}\left[\|\hat{M}(Y) - M\|_F^2\right]$
- in Bernoulli model, have for any estimator $\hat{M}$ that
  $$\sup_{M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}} \mathbb{E}\left[\|\hat{M} - M\|_F^2\right] \geq c(n \vee d)$$
- least squares approach: $\hat{M}_{\text{LS}} \in \text{argmin}_{M' \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}} \|M' - Y\|_F^2$ yields
  $$\mathbb{E}\left[\|\hat{M}_{\text{LS}} - M\|_F^2\right] \leq c' \log(nd)^2 (n \vee d)$$

Previous Work

Minimax Optimal Estimation [Mao et al., 2020]

- focus on estimating $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}$ and reconstruction error $\mathbb{E}\left[\|\hat{M}(Y) - M\|_F^2\right]$
- in Bernoulli model, have for any estimator $\hat{M}$ that
$$\sup_{M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}} \mathbb{E}\left[\|\hat{M} - M\|_F^2\right] \geq c(n \vee d)$$
- least squares approach: $\hat{M}_{\text{LS}} \in \operatorname{argmin}_{M' \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}} \|M' - Y\|_F^2$ yields
$$\mathbb{E}\left[\|\hat{M}_{\text{LS}} - M\|_F^2\right] \leq c' \log(nd)^2 (n \vee d)$$
- **question: what about polynomial time estimators?**

Previous Work

Estimating the Permutation [Mao et al., 2020]

- meta algorithm:
  - first estimate π and η with $(M_{\pi^{-1}(i)\eta^{-1}(j)})_{i,j} \in \mathbb{C}_{\mathrm{Biso}}$ by $\hat{\pi}$ and $\hat{\eta}$
  - then $(\hat{M}_{ij}) = (\hat{N}_{\hat{\pi}(i)\hat{\eta}(j)})$ with $\hat{N} \in \mathrm{argmin}_{N' \in \mathbb{C}_{\mathrm{Biso}}} \|N' - (Y_{\hat{\pi}^{-1}(i)\hat{\eta}^{-1}(j)})\|_F^2$

Previous Work

Estimating the Permutation [Mao et al., 2020]

- meta algorithm:
  - first estimate π and η with $(M_{\pi^{-1}(i)\eta^{-1}(j)})_{i,j} \in \mathbb{C}_{\mathrm{Biso}}$ by $\hat{\pi}$ and $\hat{\eta}$
  - then $(\hat{M}_{ij}) = (\hat{N}_{\hat{\pi}(i)\hat{\eta}(j)})$ with $\hat{N} \in \operatorname{argmin}_{N' \in \mathbb{C}_{\mathrm{Biso}}} \|N' - (Y_{\hat{\pi}^{-1}(i)\hat{\eta}^{-1}(j)})\|_F^2$
- error decomposition $\mathbb{E}\left[\|\hat{M} - M\|_F^2\right] \leq c(\mathscr{L} + \mathscr{P})$
  - $\mathscr{L}$ corresponds to risk of least squares estimation in $\mathbb{C}_{\mathrm{Biso}}$
  - $\mathscr{P} := \mathbb{E}\left[\|(M_{\pi^{-1}(i)\eta^{-1}(j)}) - (M_{\hat{\pi}^{-1}(i)\eta^{-1}(j)})\|_F^2\right] + \mathbb{E}\left[\|(M_{\pi^{-1}(i)\eta^{-1}(j)}) - (M_{\pi^{-1}(i)\hat{\eta}^{-1}(j)})\|_F^2\right]$

Our Contribution

Optimality in Special Cases [G., Carpentier, Verzelen, 2024+]

- for $p, h \in [0,1]$, estimate $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}} \cap \{p-h, p+h\}^{n \times d}$ and in particular the unknown, underlying rankings π and η

Our Contribution

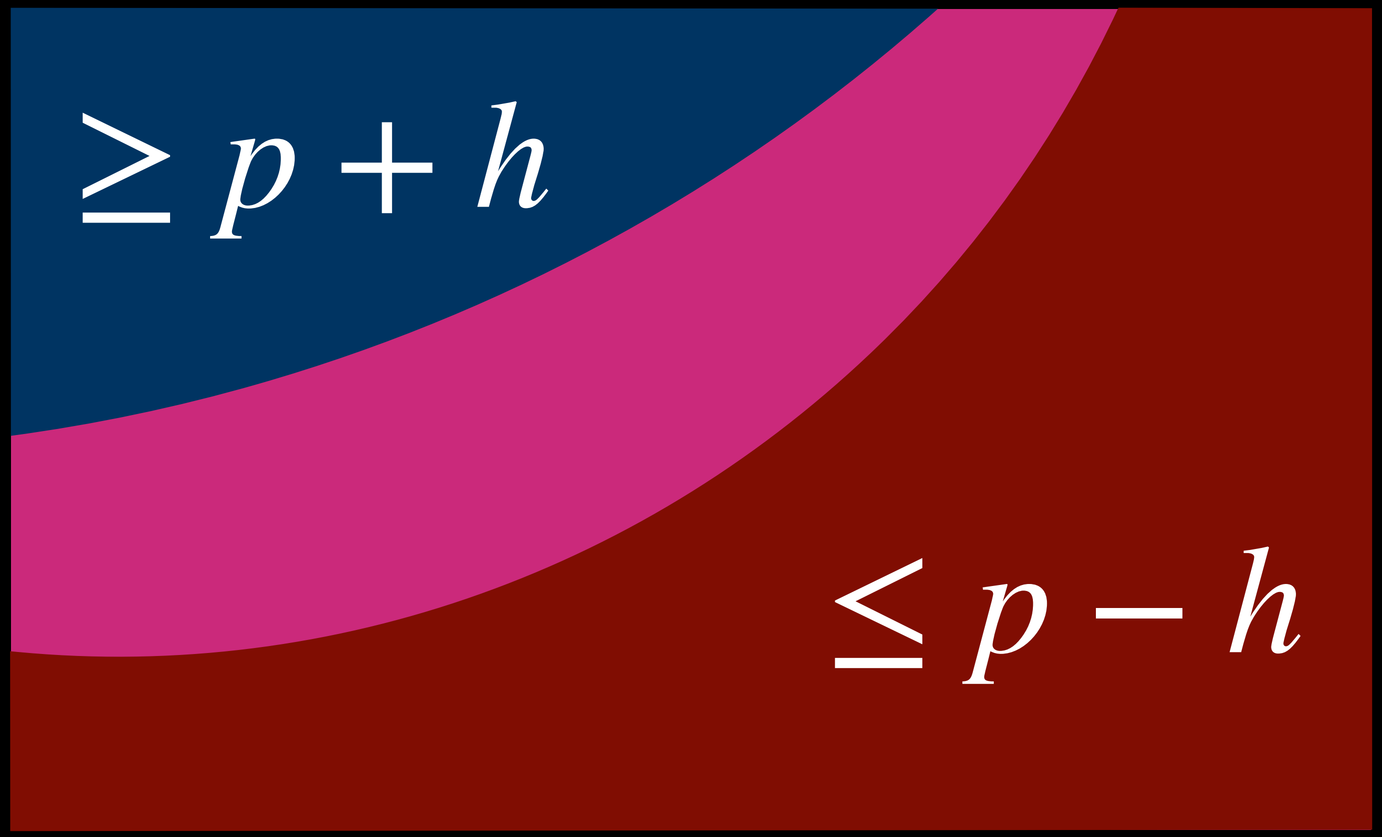
Optimality in Special Cases [G., Carpentier, Verzelen, 2024+]

- for $p, h \in [0,1]$, estimate $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}} \cap \{p-h, p+h\}^{n \times d}$ and in particular the unknown, underlying rankings π and η
- **main result:** polynomial time algorithm that yields $(\hat{\pi}, \hat{\eta})$ such that $\mathscr{P} \leq c \left(\log(nd)^{5/2}(n \vee d) \wedge ndh^2 \right)$

Our Contribution

Optimality in Special Cases [G., Carpentier, Verzelen, 2024+]

- for $p, h \in [0,1]$, estimate $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}} \cap \{p-h, p+h\}^{n\times d}$ and in particular the unknown, underlying rankings π and η
- **main result:** polynomial time algorithm that yields $(\hat{\pi}, \hat{\eta})$ such that $\mathscr{P} \leq c\left(\log(nd)^{5/2}(n \vee d) \wedge ndh^2\right)$
- remark: algorithm designed for the more general problem of level set estimation



Summary of Part 1

- conjectured computational-statistical gaps for estimating $M \in \mathbb{C}_{\text{Biso}}^{\text{Perm}}$
- least-squares estimator with rate $\log(nd)^2(n \vee d)$, but polynomial-time?
- polynomial-time estimators so far:
  - rate $\log(nd)^2(n \vee d)(n \wedge d)^{1/4}$ achieved by [Mao et al., 2020]
  - improved rate $n^{7/6+o(1)}$ in the case $n = d$ by [Liu and Moitra, 2020]
- our algorithm is optimal for special instances of the problem (and for the problem of level set estimation)

Part 2: Algorithmic Ideas for Expert Ranking

Row Sums

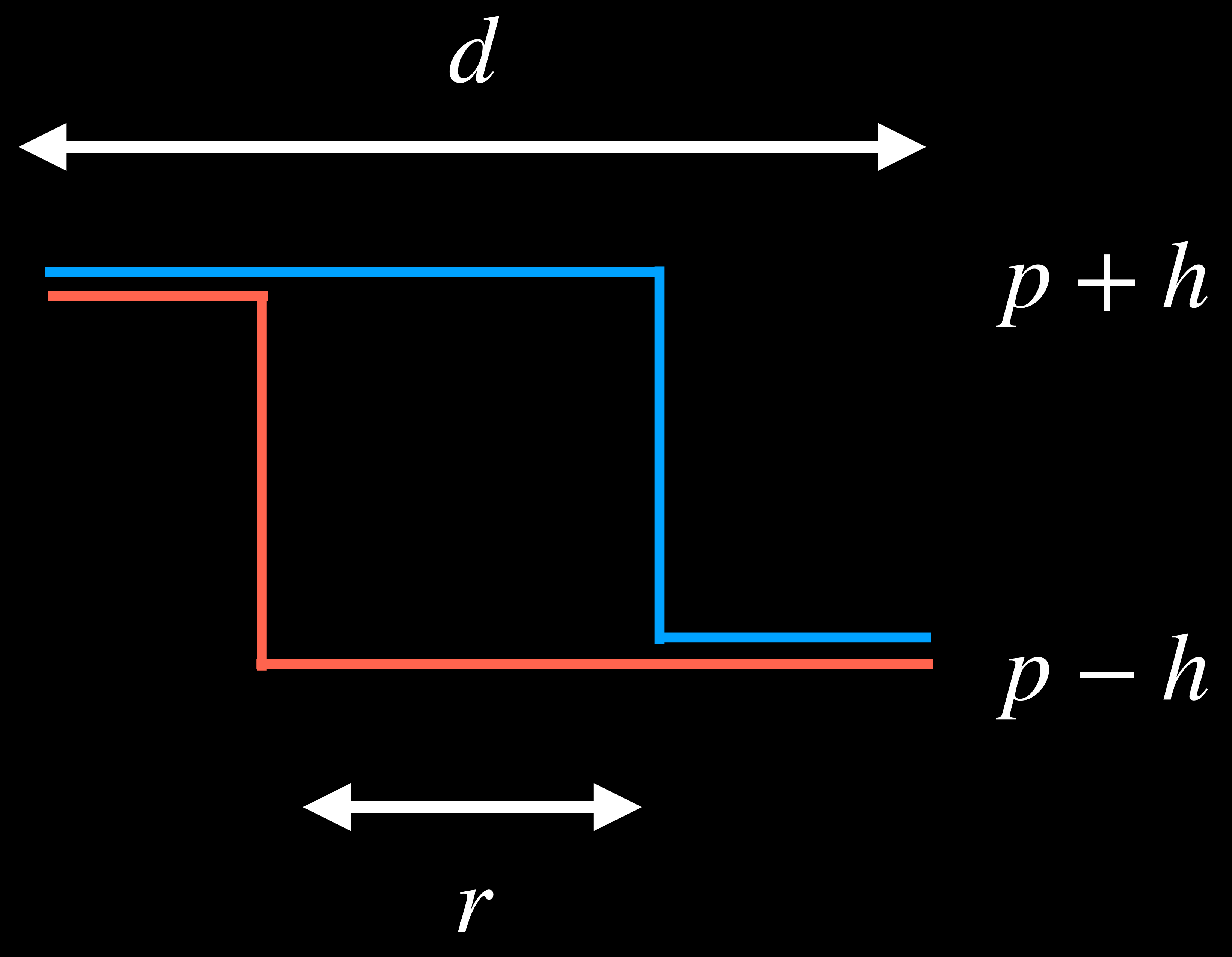
A Simple Global Approach

- recall:
  - π, η unknown permutations
  - $M_{ij} = N_{\pi(i)\eta(j)} \in \{p-h, p+h\}$ with $N \in \mathbb{C}_{\text{Biso}}$
  - $Y_{ij} = M_{ij} + W_{ij}$ with W_{ij} 1-sub-Gaussian
- bi-isotonicity of N implies
  $\pi(i) \leq \pi(i') \Rightarrow M_{ij} \geq M_{i'j} \qquad \forall j \in [d]$
- sort according to $y_i := \sum_{j=1}^{d} Y_{ij}$

Row Sums

A Simple Global Approach

- recall:
  - π, η unknown permutations
  - $M_{ij} = N_{\pi(i)\eta(j)} \in \{p-h, p+h\}$ with $N \in \mathbb{C}_{\text{Biso}}$
  - $Y_{ij} = M_{ij} + W_{ij}$ with W_{ij} 1-sub-Gaussian
- bi-isotonicity of N implies
  $\pi(i) \le \pi(i') \Rightarrow M_{ij} \ge M_{i'j} \qquad \forall j \in [d]$
- sort according to $y_i := \sum_{j=1}^{d} Y_{ij}$

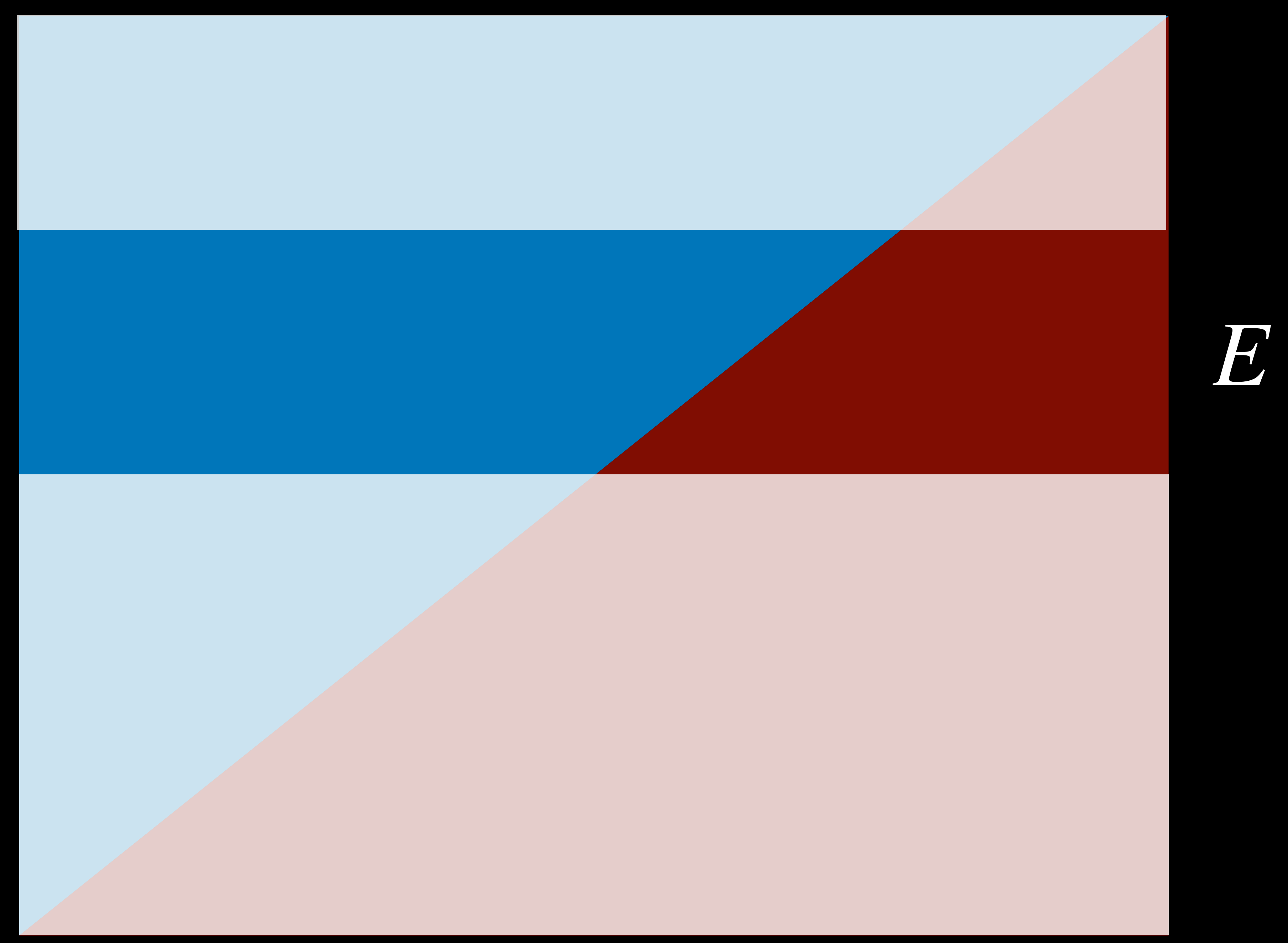


signal $2rh$ vs. noise of order $\sqrt{d}$

Partial Row Sums

Reducing the Noise by Reducing the Tasks

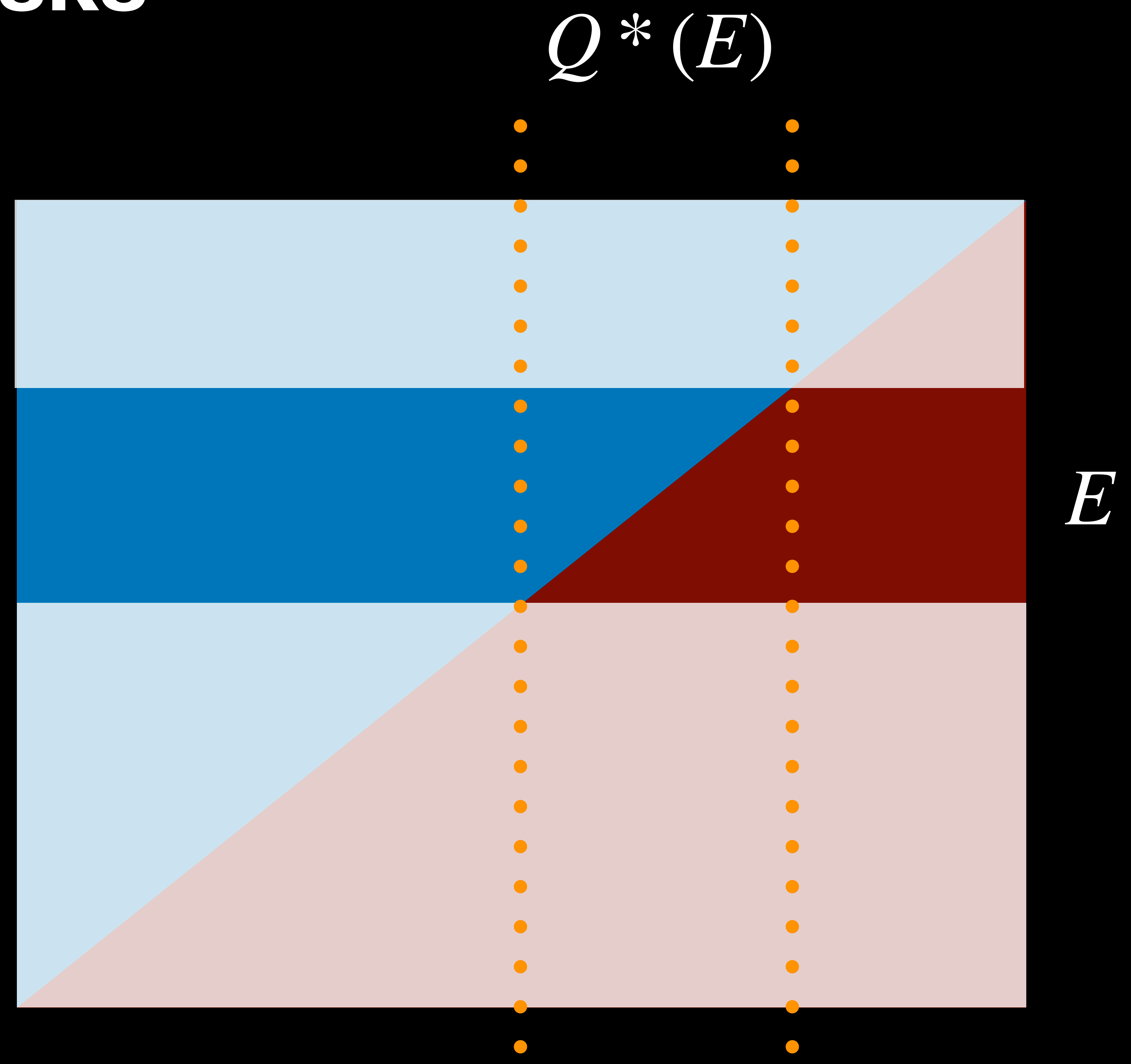
- focus on smaller group of experts $E \subseteq [n]$
- only some tasks $Q^*(E)$ relevant for comparison



Partial Row Sums

Reducing the Noise by Reducing the Tasks

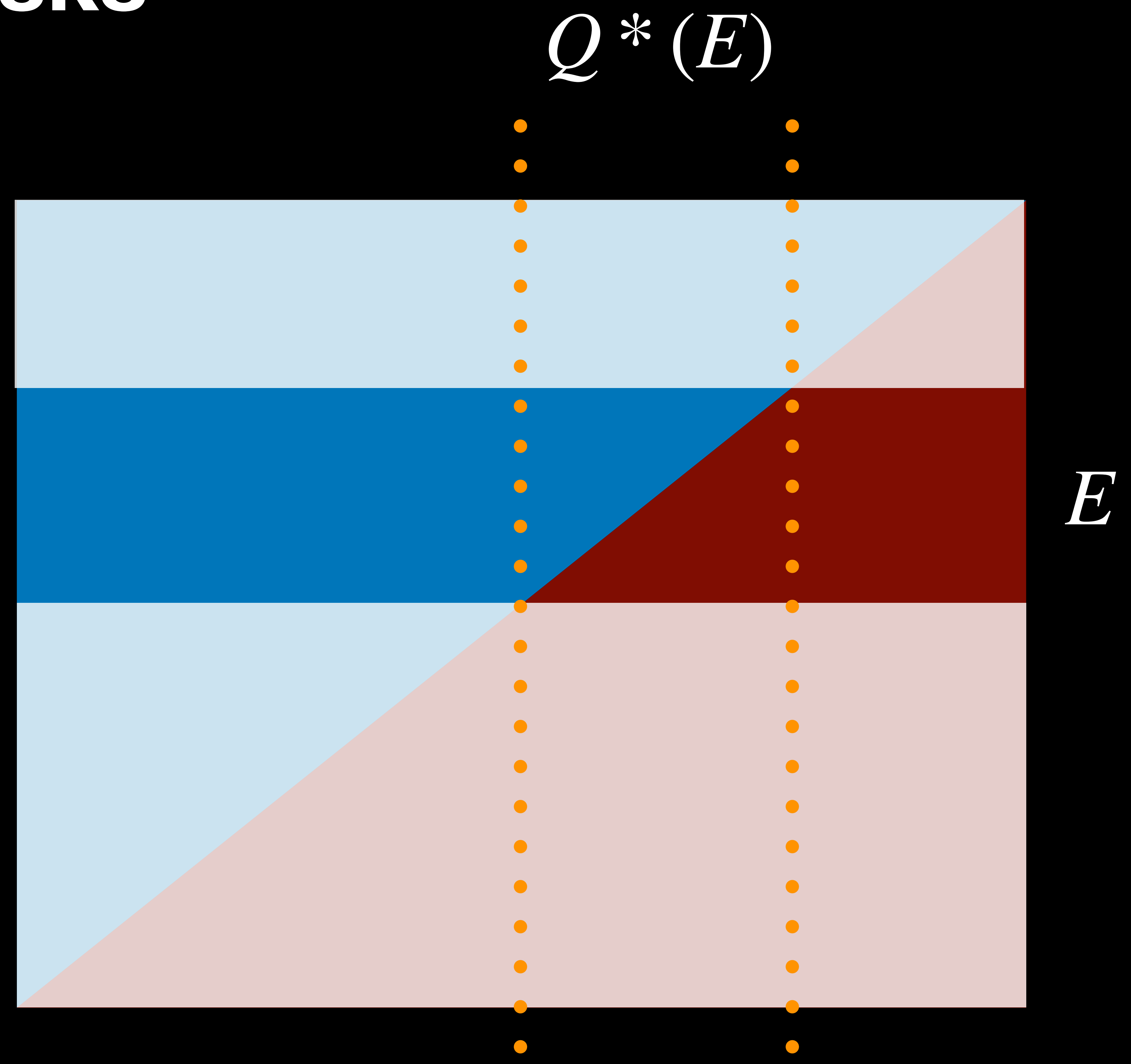
- focus on smaller group of experts $E \subseteq [n]$
- only some tasks $Q*(E)$ relevant for comparison
- noise of order $\sqrt{|Q*(E)|}$ instead of $\sqrt{d}$



Partial Row Sums

Reducing the Noise by Reducing the Tasks

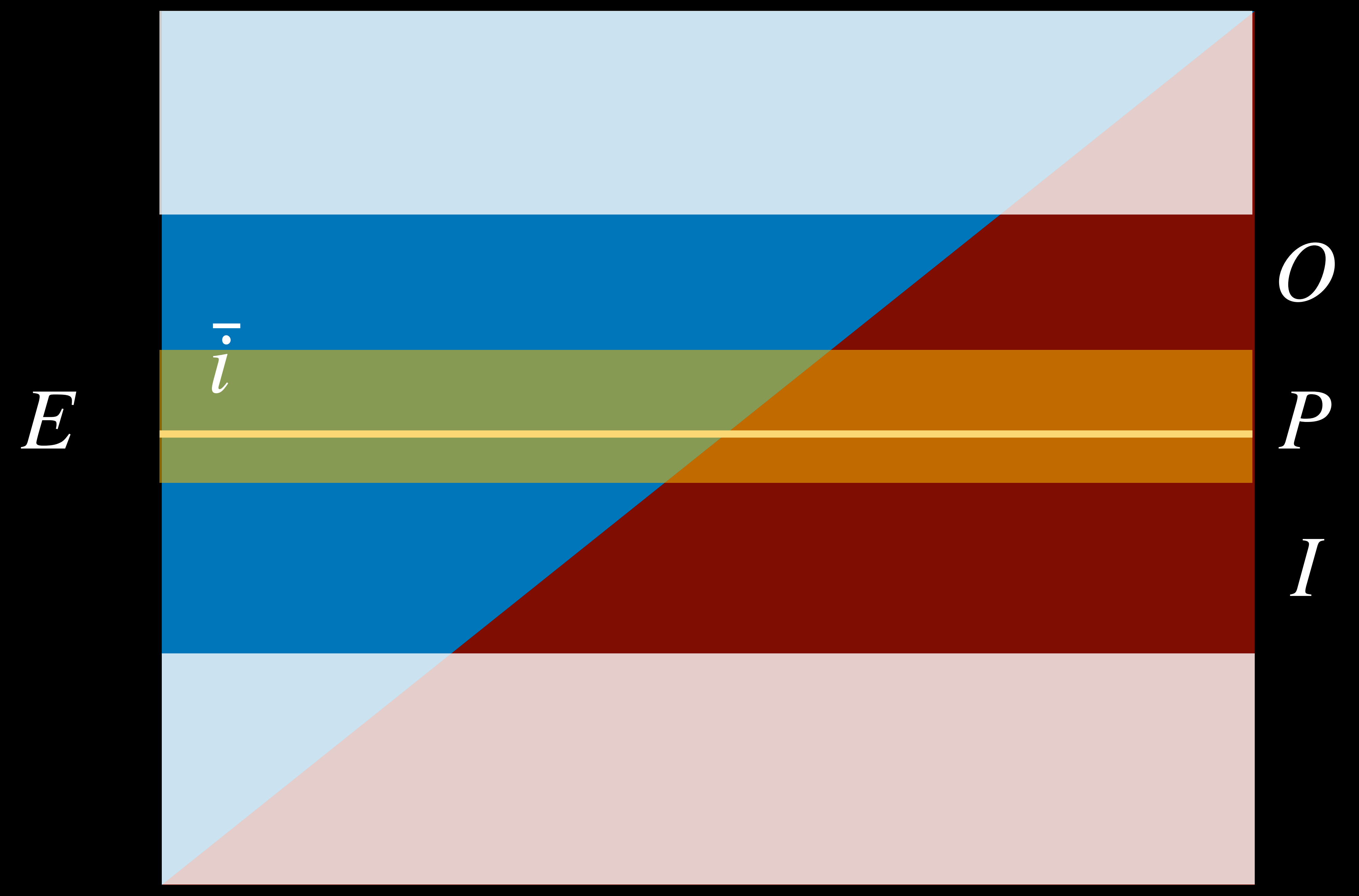
- focus on smaller group of experts $E \subseteq [n]$
- only some tasks $Q*(E)$ relevant for comparison
- noise of order $\sqrt{|Q*(E)|}$ instead of $\sqrt{d}$
- **problem:** cannot access $Q*(E)$ directly



Trisection

Refining sets of Experts based on Partial Row Sums

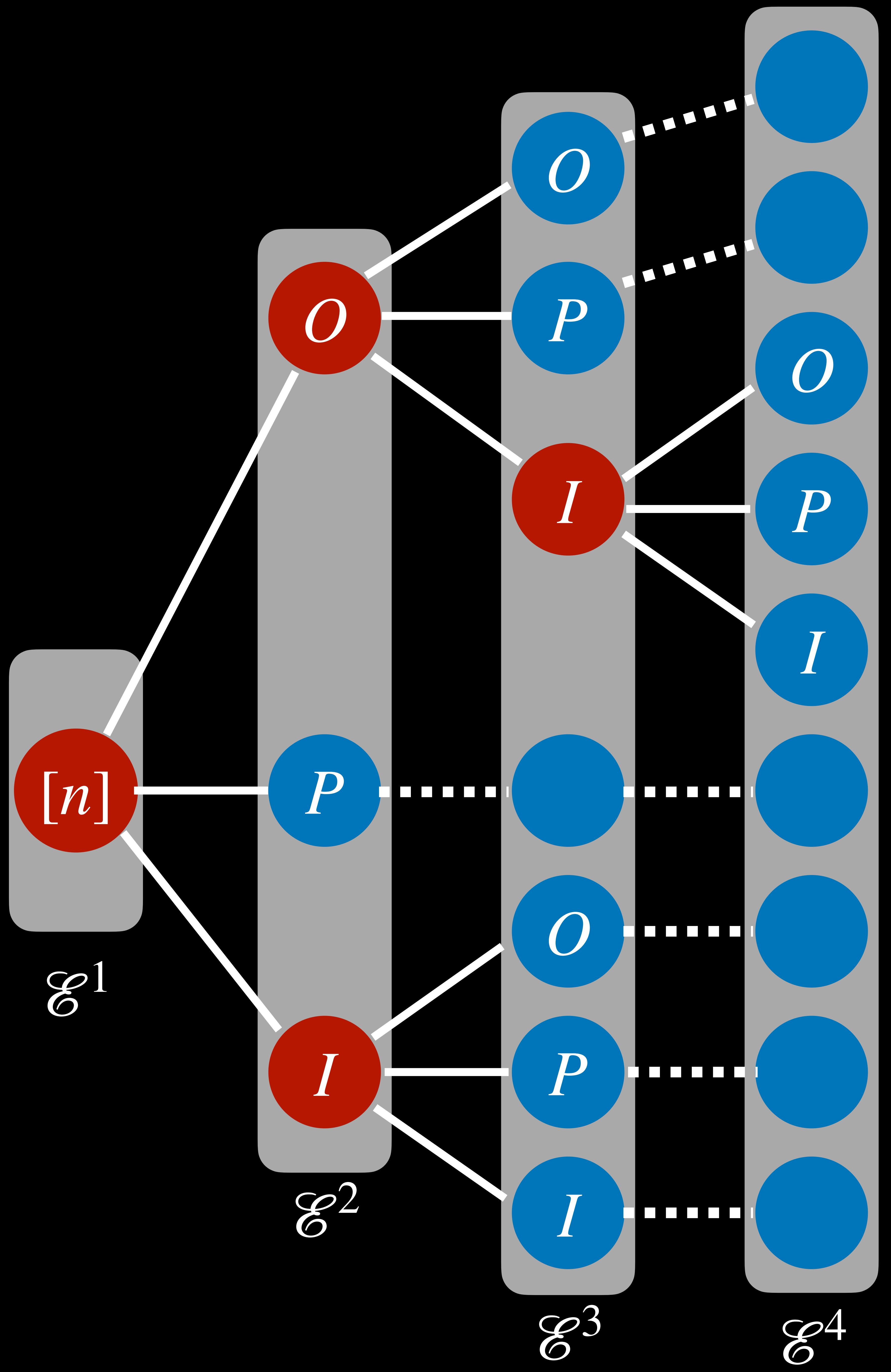
- we want to partition sets $E \subseteq [n]$ into trisection (O, P, I)
  - O: experts "better" than the median expert $\bar{i}$
  - P: experts we cannot distinguish from $\bar{i}$
  - I: experts "worse" than $\bar{i}$
- split based on partial row sums over sets related to $Q*(E)$



The Sorting Tree

Hierarchical Sorting Based on Trisections

- start by trisecting $[n]$
- inductively obtain partitions $\mathscr{E}$ of $[n]$
- trisect each $E \in \mathscr{E}$ until:
  - E is sufficiently small
  - $Q*(E)$ is sufficiently small
  - $E = P$ from some previous trisection (O, P, I)



Summary of Part 2

- sorting based on global row sums in general lacks precision
- hierarchical sorting allows us to reduce the global sorting problem into multiple local sorting problem
- for local sorting, less tasks are relevant, so we reduce the noise

References

- Cheng Mao, Ashwin Pananjady, and Martin J. Wainwright. "Towards optimal estimation of bivariate isotonic matrices with unknown permutations." The Annals of Statistics 48.6, 2020.
- Allen Liu, and Ankur Moitra. "Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation." Conference on Learning Theory. PMLR, 2020.
- M. G., Alexandra Carpentier, and Nicolas Verzelen. "Optimal level set estimation for non-parametric tournament and crowdsourcing problems." arXiv preprint arXiv:2408.15356, 2024.