Permutation Estimation for Crowdsourcing

Maximilian Graf – Universität Potsdam – ✉ graf9@uni-potsdam.de

supported by the Deutsche Forschungsgemeinschaft (DFG) FOR 5381

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

- 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?

Existing Results

- most work on estimating M and reconstruction error $\mathbb{E}\left[\left\|\widehat{M}-M\right\|_F^2\right]$
- least squares approach yields (up to polylog) minimax optimal rate $n \vee d$
- no polynomial time estimator known that is optimal!

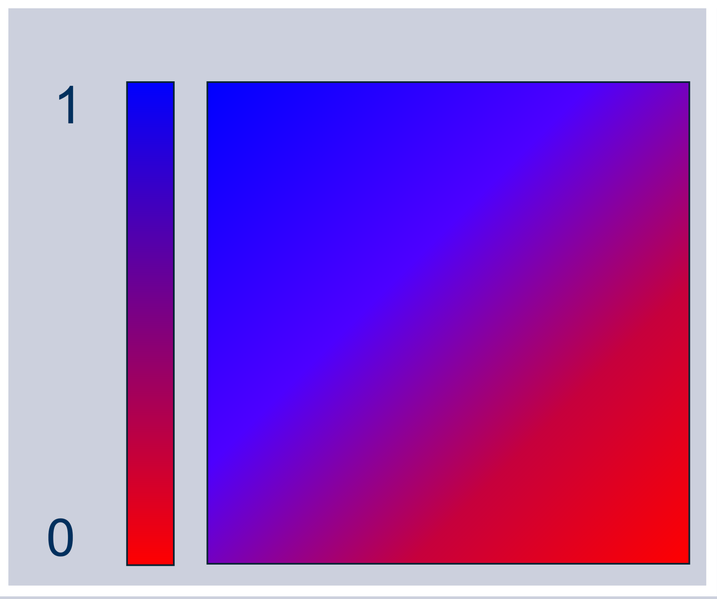
Key Idea:
efficient estimation of rankings π and η allows efficient estimation of M

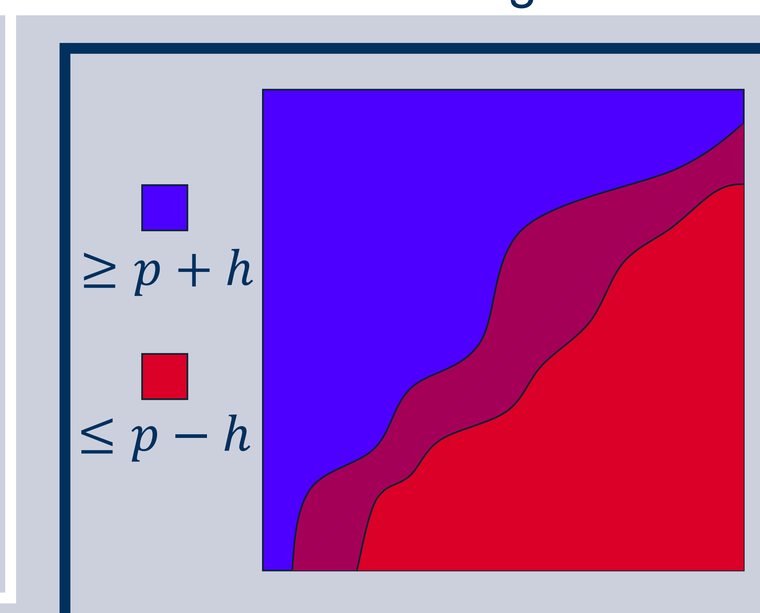
Gallery: Algorithmic Ideas of our Permutation Estimator `SoHLoB`

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:
- 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



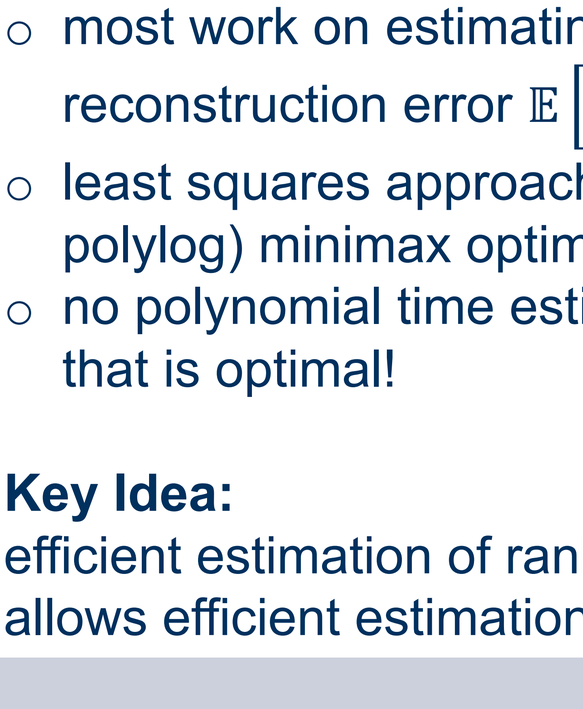


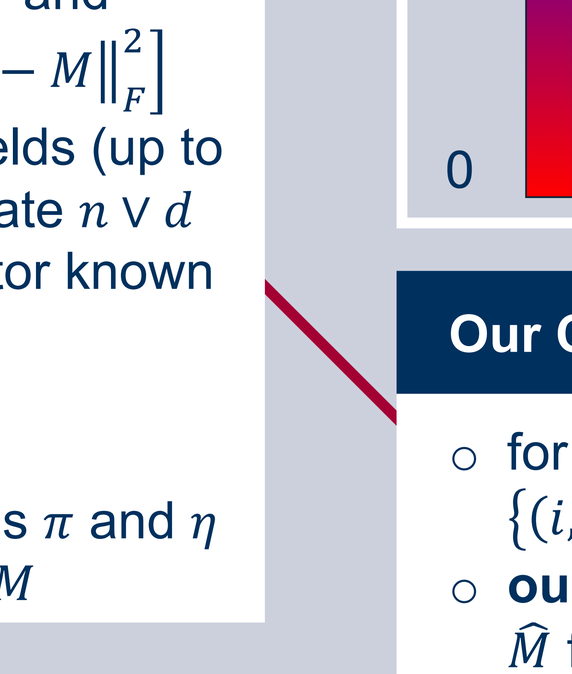
Our Contribution

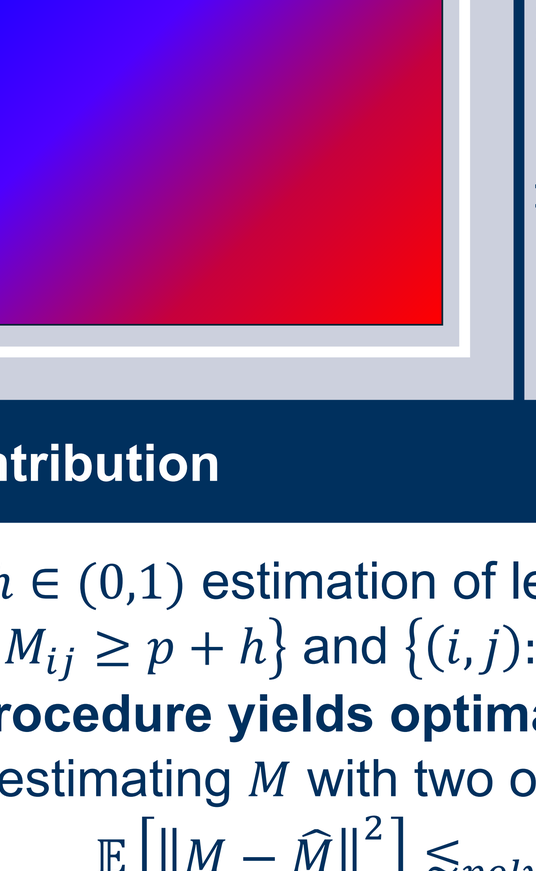
- for $p, h \in (0,1)$ estimation of level sets $\{(i,j): M_{ij} \geq p+h\}$ and $\{(i,j): M_{ij} \leq p-h\}$
- **our procedure yields optimal and efficient estimators** $\widehat{M}$ for estimating M with two or finitely many values s.t.

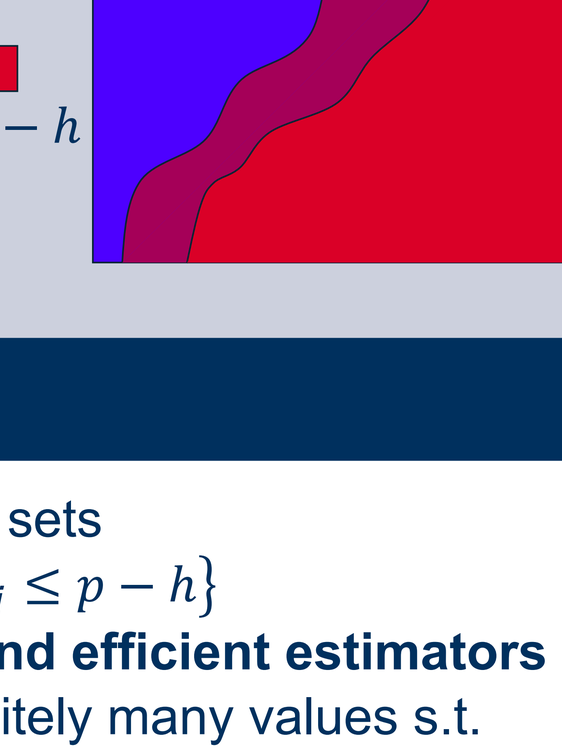
$$\mathbb{E}\left[\left\|M-\widehat{M}\right\|_F^2\right] \lesssim_{polylog} n \vee d$$

- applicable to tournament models (SST, noisy sorting)









Further Reading:
- Cheng Mao, Ashwin Pananjadyy and Martin Wainwright, 2020, Towards Optimal Estimation of Bivariate Isotonic Matrices with Unknown Permutations, *The Annals of Statistics*
- Allen Liu and Ankur Moitra, 2020, Better Algorithms for Estimating Non-Parametric Models in Crowd-Sourcing and Rank Aggregation, *Conference on Learning Theory*
- Alexandra Carpentier, Emmanuel Pilliat and Nicolas Verzelen, 2023, Optimal Permutation Estimation in Crowd-Sourcing Problems, *The Annals of Statistics*
- our preprint, soon on arXiv ☺