Statistical and Computational Guarantees of Kernel Max-Sliced Wasserstein Distances

Jie Wang
Georgia Institute of Technology

Joint work with March Boedihardjo (Michigan State) and Yao Xie (Georgia Tech)

Runner Up of INFORMS Data Mining Best Student Paper Competition

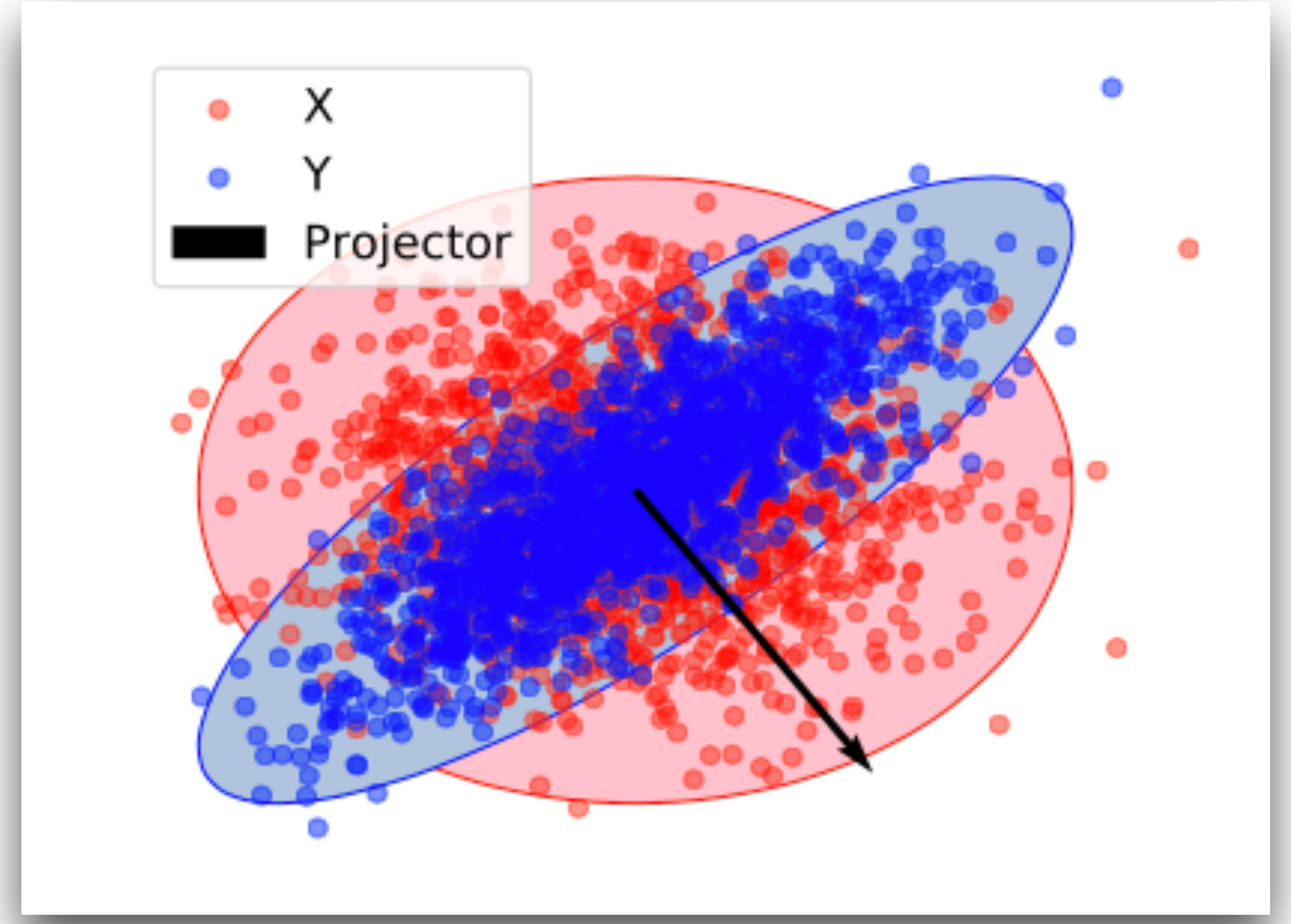
Background of Hypothesis Testing

- **Given**: *high-dimensional* samples from unknown distributions μ and ν

- **Goal**: determine whether μ and ν differ

- Project data linearly and follow by Wasserstein testing [1]

$$\mathbf{MS}(\mu, \nu) = \max_{f:\, f(x)=a^\top x} \mathbf{W}(f_\# \mu, f_\# \nu)$$

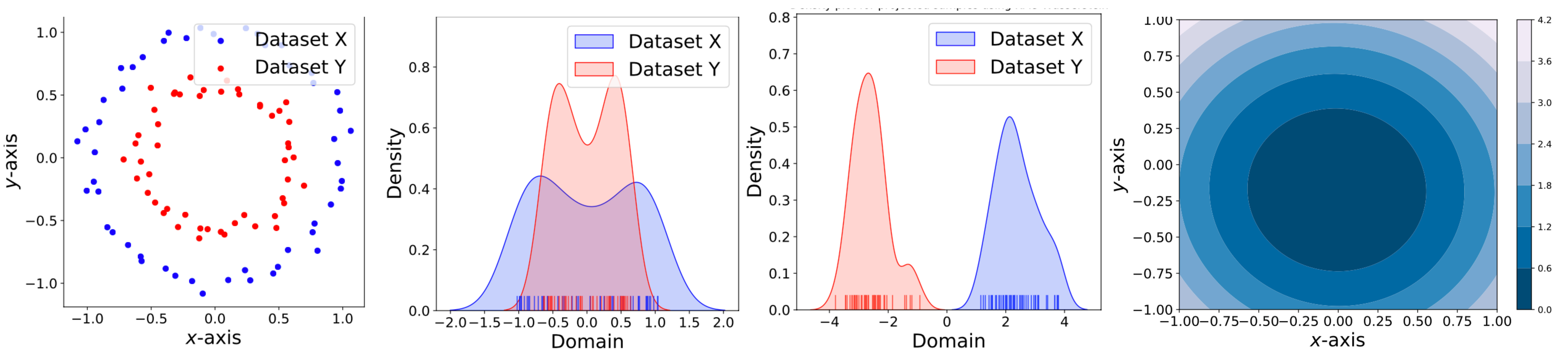


Wang, Gao, and Xie. Two-sample test using projected wasserstein distance. ISIT-21

Testing with Nonlinear Dimensionality Reduction

- Use **nonlinear** operator modeled by kernel method

$$\mathbf{KMS}(\mu, \nu) = \max_{f \in \mathscr{H}:\ \|f\|_{\mathscr{H}} \leq 1} \mathbf{W}(f_{\#}\mu, f_{\#}\nu)$$



(a) Scatter plot of data

(b) Density of projected samples using MS

(c) Density of projected samples using KMS

(d) Heatmap of estimated nonlinear projector

Wang, Gao, and Xie. Two-sample Test with Kernel Projected Wasserstein Distance. AISTATS-22

Kernel Max-Sliced Wasserstein Distance

$$\textbf{KMS}(\mu,\nu) = \max_{f \in \mathcal{H}:\ \|f\|_{\mathcal{H}} \leq 1} \mathbf{W}(f_{\#}\mu, f_{\#}\nu)$$

- $\mathcal{H}$: reproducing kernel Hilbert space (RKHS) dependent on $k(\cdot,\cdot): \mathbb{R}^{D\times D} \to \mathbb{R}^1$
- $f: \mathbb{R}^D \to \mathbb{R}^1$: nonlinear projector

1. Statistical guarantees of KMS
2. Computational guarantees of KMS
3. Practical Applications

Wang, March, and Xie. Statistical and Computational Guarantees of Kernel Max-Sliced Wasserstein Distances. ICML-25

Finite-Sample Guarantees

$$\mathbf{KMS}(\mu, \nu) = \max_{f \in \mathcal{H}:\ \|f\|_{\mathcal{H}} \le 1} \mathbf{W}(f_{\#}\mu, f_{\#}\nu)$$

Theorem (Informal). **Assume** $k(x, x) \le A, \forall x$. With high probability,

$$\mathbf{KMS}(\mu, \hat{\mu}_n) = \mathcal{O}(n^{-1/2}).$$

- $\mathcal{O}(\cdot)$ hides constant depending on A

- **KMS** breaks the curse of dimensionality of Wasserstein distance
- Free of distribution assumptions
  - MS distance **[Sloan N et. al, 2022, Tianyi et. al, 2021]**: finite diameter of support
  - KMS distance **[Wang et. al, 2022]**: light-tailed distribution

Finite-Dimensional Reformulation

$$\mathrm{KMS}(\hat{\mu}_n, \hat{\nu}_n) = \max_{\omega \in \mathbb{R}^{2n}:\ \|\omega\|_2 = 1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} (M_{i,j}^\top \omega)^2 \right\}$$

- $M_{i,j} \in \mathbb{R}^{2n}$: concatenation of kernel valued on data points
- Non-concave quadratic optimization problem

Theorem. $\mathrm{KMS}(\hat{\mu}_n, \hat{\nu}_n)$ **is $\mathcal{NP}$-hard to compute**

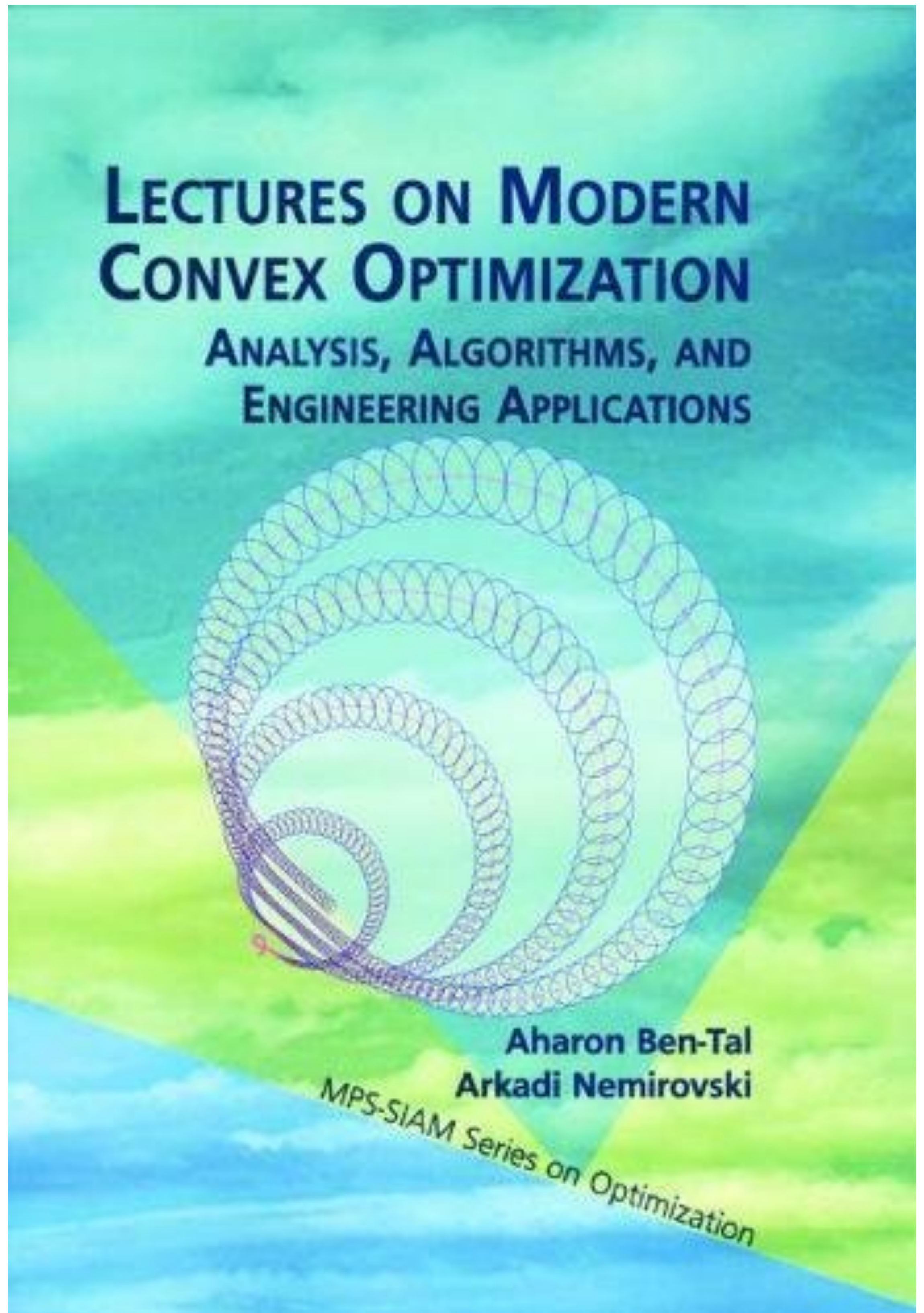
(KMS Problem) $\supseteq$ $\max_{\omega:\ \|\omega\|_2=1} \min_{i \in [n]} \omega^\top A_i A_i^\top \omega$ **(Fair-PCA with rank-1 data)** $\supseteq$ Give integers $a_1, \ldots, a_N$, determine whether binary variables $\{x_i\}_{i=1}^N \in \{-1, 1\}^N$ exist such that $\sum_{i=1}^N a_i x_i = 0$? **(Partition)**

Samadi, Samira, et al. NIPS2018

Finite-Dimensional Reformulation

$$\mathbf{KMS}(\hat{\mu}_n, \hat{\nu}_n) = \max_{\omega \in \mathbb{R}^{2n}:\ \|\omega\|_2 = 1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^\top, \omega\omega^\top \rangle \right\}$$

- **Approximation algorithm** using semidefinite relaxation (SDR):



$$S = \omega\omega^\top, \quad \omega \in \mathbb{R}^{2n}, \quad \|\omega\|_2 = 1$$

$\Updownarrow$

$$S \succeq 0, \quad \mathrm{Trace}(S) = 1, \quad \mathrm{rank}(S) = 1$$

Semidefinite Relaxation (SDR)

$$\mathbf{KMS}(P_n, Q_n) = \max_{S \succeq 0, \mathrm{Trace}(S)=1, \mathrm{rank}(S)=1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^\top, S \rangle \right\}$$

$$\leq \max_{S \succeq 0, \mathrm{Trace}(S)=1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^\top, S \rangle \right\}$$

Theorem (Informal). Stochastic gradient method with biased oracles solves SDR up to δ optimality gap with operational complexity

$$\tilde{\mathcal{O}}(n^3 \delta^{-3})$$

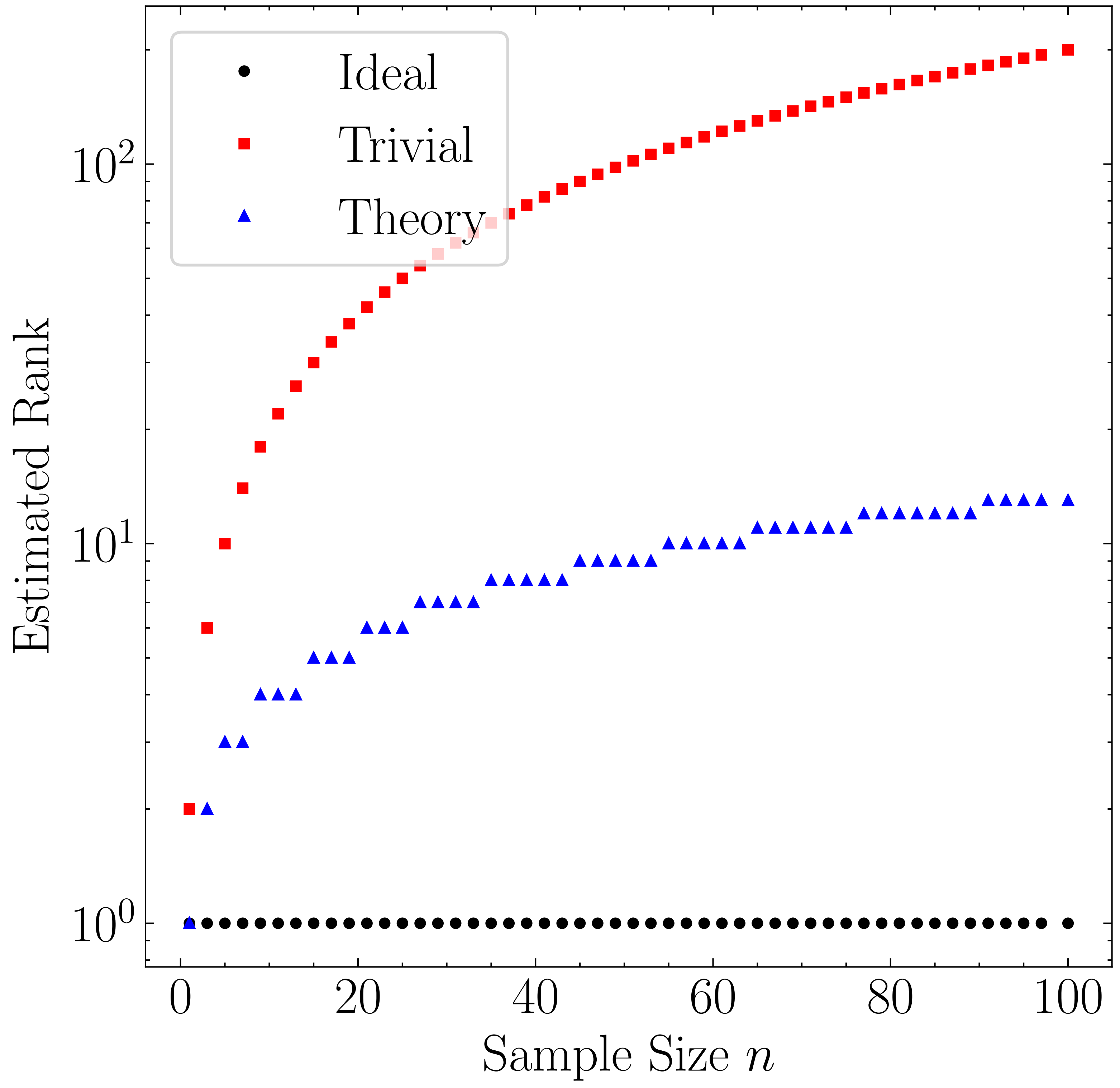
Quality of Semidefinite Relaxation

- $$(\text{KMS}) = \max_{S \succeq 0, \text{Trace}(S)=1, \text{rank}(S)=1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^{\top}, S \rangle \right\}$$
- $$(\text{SDR}) = \max_{S \succeq 0, \text{Trace}(S)=1} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^{\top}, S \rangle \right\}$$

Smaller rank of optimal solution yields better performance

Theorem. There exists an optimal solution to (SDR) with rank $k \triangleq 1 + \left\lfloor \sqrt{2n + \frac{9}{4}} - \frac{3}{2} \right\rfloor$.

- $$(\text{KMS}) \leq (\text{SDR}) \leq \max_{S \succeq 0, \text{Trace}(S)=1, \text{rank}(S)=k} \left\{ \min_{\pi \in \Gamma_n} \sum_{i,j=1}^{n} \pi_{i,j} \langle M_{i,j} M_{i,j}^{\top}, S \rangle \right\}$$



Summary

- A novel **non-parametric** metric for comparing high-dimensional distributions
- **Sharp** finite-sample guarantees
- Computational Guarantees:
    - A. Non-concave quadratic maximization problem: **$\mathcal{NP}$-hard**
    - B. Approximation algorithm **with performance guarantees**:
- Practical Applications:
    1. High-dimensional Two-Sample Testing
    2. Change-Point Detection