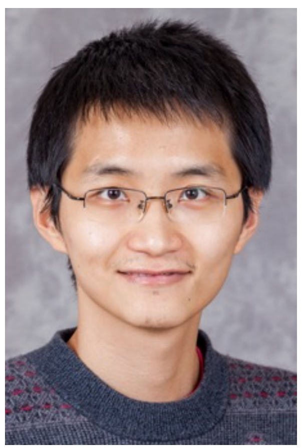
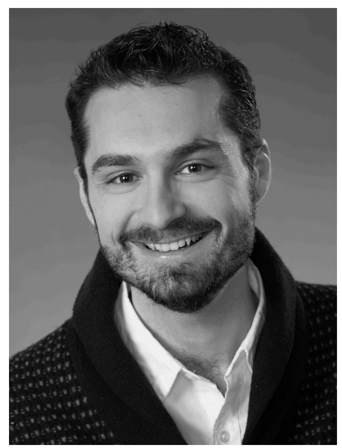
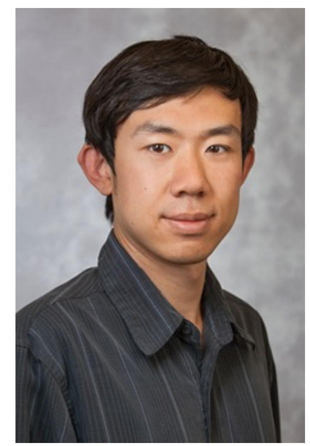
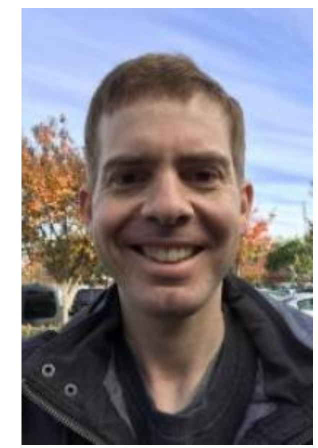
Faster Algorithms for High-Dimensional Robust Covariance Estimation









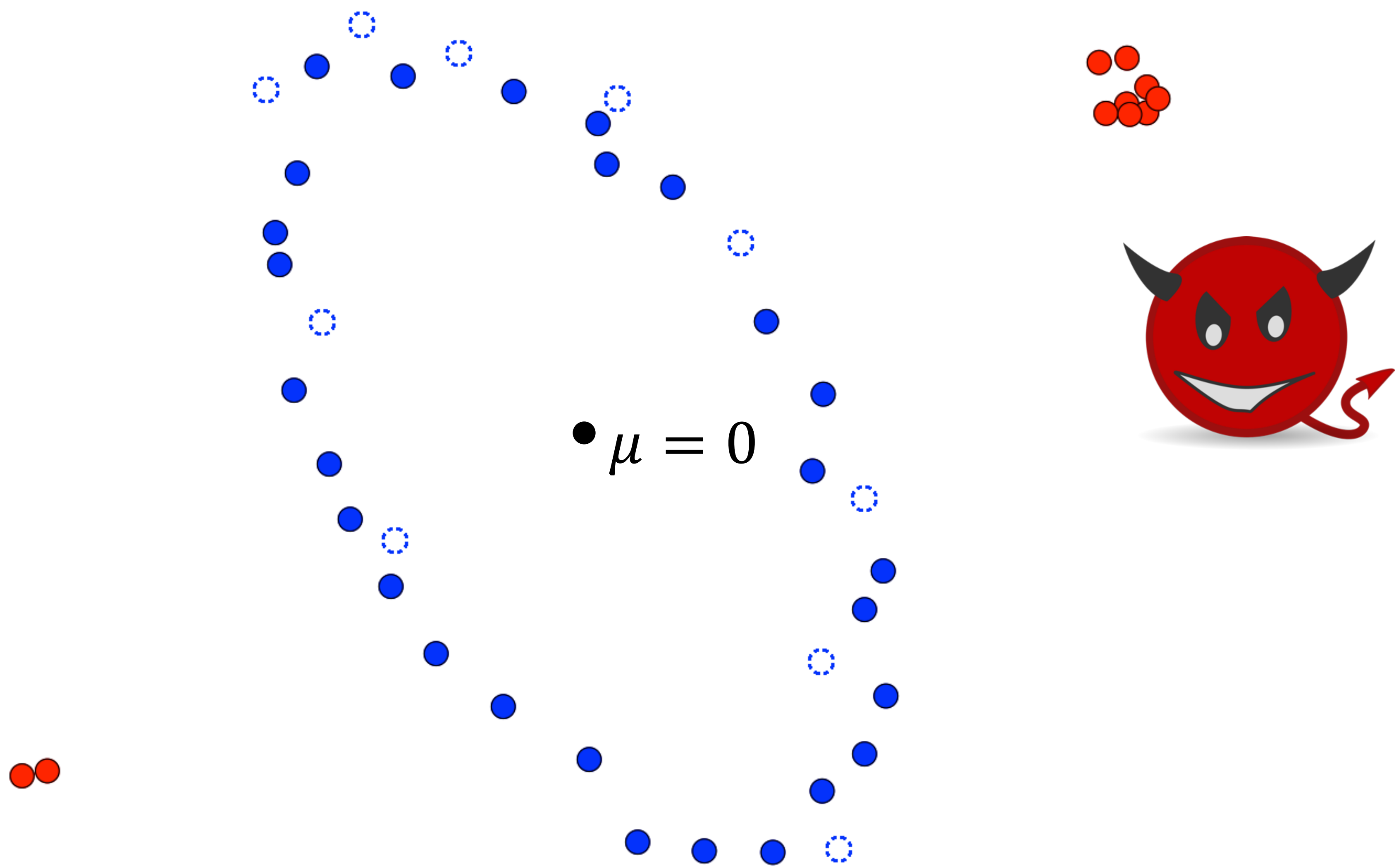
Yu Cheng
Duke → UIC

Ilias Diakonikolas
USC

Rong Ge
Duke

David Woodruff
CMU

Robust Covariance Estimation



Covariance Estimation

- Input: N samples $\{X_1, \dots, X_N\}$ drawn from $\mathcal{N}(0, \Sigma)$ on $\mathbb{R}^d$.
- Goal: Learn Σ.

$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} X_i X_i^\top$ works!

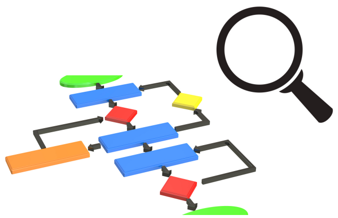
$\left\|\hat{\Sigma} - \Sigma\right\|_F \le \epsilon \|\Sigma\|_2$ when $N = \Omega(d^2/\epsilon^2)$.

Input size: $Nd = O(d^3/\epsilon^2)$.

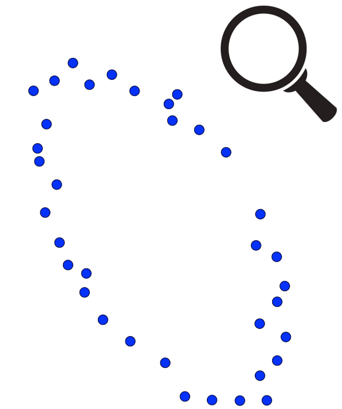
Runtime: (d, N, d)-matrix multiplication time $= O(d^{3.26}/\epsilon^2)$.

Robust Covariance Estimation

ϵ-Corruption:



specifies N.



N samples are drawn from $\mathcal{N}(0, \Sigma)$.

replaces ϵN samples with arbitrary points.

Goal: Compute $\hat{\Sigma} \approx \Sigma$ given an ϵ-corrupted set of N samples.

Previous Work

Algorithm	Error Guarantee
[Diakonikolas+ '16]	$\left\lVert \Sigma^{-1/2}\,\hat{\Sigma}\,\Sigma^{-1/2} - I \right\rVert_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right)$
[Lai+ '16]	$\left\lVert \hat{\Sigma} - \Sigma \right\rVert_F \le O\left(\sqrt{\epsilon \log d}\right) \lVert \Sigma \rVert_2$

Previous Work

Algorithm	Error Guarantee	Runtime
[Diakonikolas+ '16]	$\left\|\Sigma^{-1/2}\,\hat{\Sigma}\,\Sigma^{-1/2} - I\right\|_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right)$	$\Omega(d^5)$
[Lai+ '16]	$\left\|\hat{\Sigma} - \Sigma\right\|_F \le O\left(\sqrt{\epsilon \log d}\right) \|\Sigma\|_2$	$\Omega(d^{4.74})$

Previous Work

Algorithm	Error Guarantee	Runtime
[Diakonikolas+ '16]	$\left\Vert \Sigma^{-1/2}\,\hat{\Sigma}\,\Sigma^{-1/2} - I \right\Vert_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right)$	$\Omega(d^5)$
[Lai+ '16]	$\left\Vert \hat{\Sigma} - \Sigma \right\Vert_F \le O\left(\sqrt{\epsilon \log d}\right) \Vert\Sigma\Vert_2$	$\Omega(d^{4.74})$
Empirical covariance	✗	$O(d^{3.26})$

Motivating Question

Can we design robust estimators that are as efficient as their non-robust analogues?

[C Diakonikolas Ge '19]:

Robust mean estimation in time $\tilde{O}(Nd)/\mathrm{poly}(\epsilon)$.

Our Results

# of Samples	Error Guarantee	Runtime
$N = \widetilde{\Omega}\left(\frac{d^2}{\epsilon^2}\right)$	$\left\|\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - I\right\|_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right)$	$\frac{\tilde{O}(d^{3.26} \log \kappa)}{\mathrm{poly}(\epsilon)}$
	$\left\|\hat{\Sigma} - \Sigma\right\|_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right) \|\Sigma\|_2$	$\frac{\tilde{O}(d^{3.26})}{\mathrm{poly}(\epsilon)}$

We also provide evidence that $d^{3.26}$ runtime may be a bottleneck.

Naïve Approach

For $X \sim \mathcal{N}(0, \Sigma)$, we have $\mathrm{E}[XX^\top] = \Sigma$.

Robustly estimate the mean $Z = X \otimes X \in \mathbb{R}^{d^2}$.

Runtime: $\Omega(Nd^2) = \Omega(d^4)$ to write down all $(Z_i)_{i=1}^N$.

Challenges

1) Need to estimate the mean of $X \otimes X$ without computing the vectors explicitly.

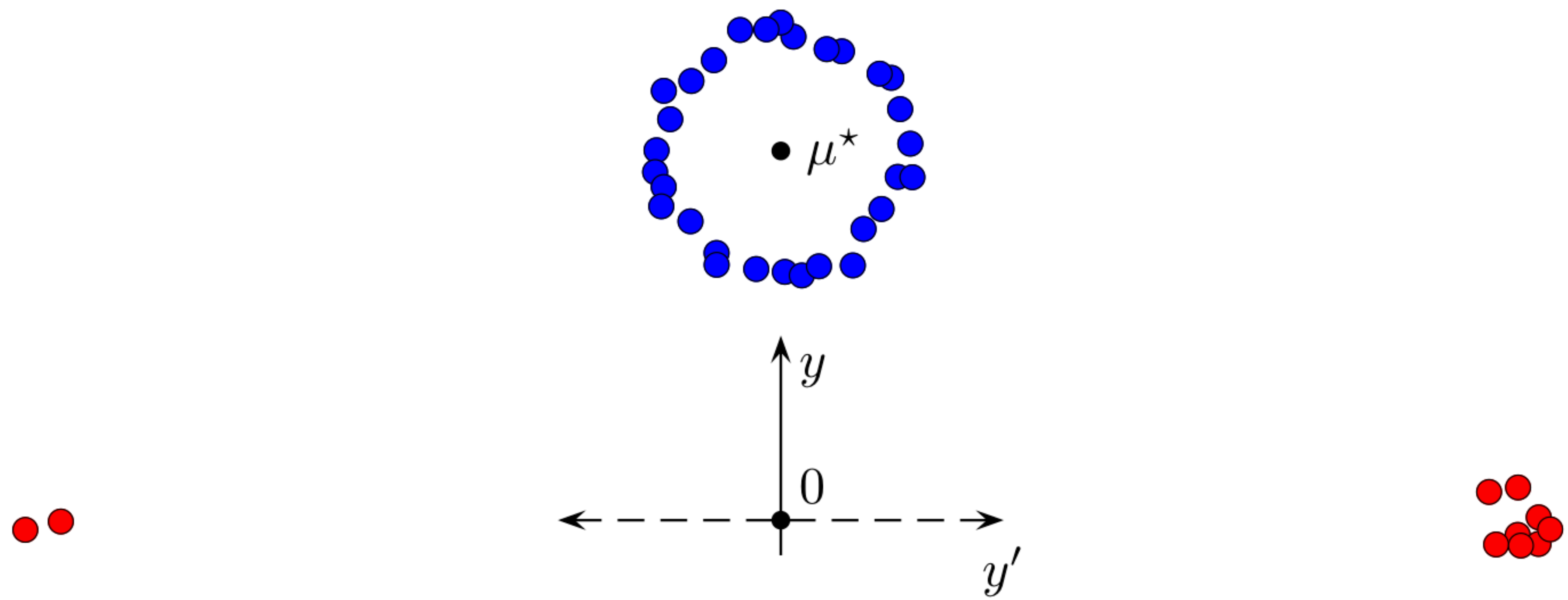
2) Robust mean estimation requires assumptions on the covariance.
   - $\mathbf{cov}[X \otimes X]$ is related to the 4th-order moments of $X \sim \mathcal{N}(0, \Sigma)$.
   - We made no assumptions on Σ.

Challenges

1) Need to estimate the mean of $X \otimes X$ without computing the vectors explicitly.

2) Robust mean estimation requires assumptions on the covariance.
   - $\mathrm{cov}[X \otimes X]$ is related to the 4th-order moments of $X \sim \mathcal{N}(0, \Sigma)$.
   - We made no assumptions on Σ.

Robust Mean Estimation [C Diakonikolas Ge '19]



Find a direction y such that the variance along y is the largest after throwing away the farthest ϵN samples.

Faster Positive SDP Solvers for Tensor Input

$$\max_{x \geq 0} 1^\top x \quad \text{s.t. } \sum_{i=1}^{N} x_i A_i \preccurlyeq I.$$

$$\min_{Y \succcurlyeq 0} \operatorname{tr}(Y) \quad \text{s.t. } A_i \cdot Y \geq 1,\ \forall i.$$

$$A_i = Z_i Z_i^\top \in \mathbb{R}^{d^2 \times d^2}$$

$$Z_i = X_i \otimes X_i \in \mathbb{R}^{d^2}$$

We build on fast positive SDP solvers (e.g., [Peng+ ’16]).

We exploit the structure of $Z = X \otimes X$ to compute $(1 \pm \epsilon)$-approximate solutions in time $\tilde{O}(d^{3.26} + N)/\operatorname{poly}(\epsilon)$.

Hardness Results

It is not known how to approximate the empirical covariance matrix faster than (d, d^2, d)-matrix multiplication time.

Let $X \in \mathbb{R}^{N \times d}$ be the sample matrix.

We give a communication complexity lower bound: any oblivious sketching matrix S must have $\widetilde{\Omega}(N)$ rows if $X^\top S^\top S\, X \approx X^\top X$.

Hardness Results (cont.)

Consider an easier task of finding good weights w_i so that $\sum_{i=1}^{N} w_i X_i X_i^\top \approx \Sigma$.

We give a reduction to show that this easier problem is still at least as hard as some basic matrix computation question.

Open Problems

Can we design robust estimators that are as efficient as their non-robust analogues?

- [C Diakonikolas Ge '19] Mean estimation. ✓
- [This paper]: Covariance estimation. ✓
- Sparse mean
- Graphical models ?
- …

Open Problems

# of Samples	Error Guarantee	Runtime
$N = \widetilde{\Omega}\left(\frac{d^2}{\epsilon^2}\right)$	$\left\Vert \Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} - I \right\Vert_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right)$	$\frac{\tilde{O}(d^{3.26} \log \kappa)}{\mathrm{poly}(\epsilon)}$
	$\left\Vert \hat{\Sigma} - \Sigma \right\Vert_F \le O\left(\epsilon \log \frac{1}{\epsilon}\right) \Vert \Sigma \Vert_2$	$\frac{\tilde{O}(d^{3.26})}{\mathrm{poly}(\epsilon)}$

Recently, the ϵ-dependence in the running time was removed for robust mean estimation [Depersin Lecué '19] [Dong Hopkins Li '19].