Algorithmic Questions in High-Dimensional Robust Statistics

Ilias Diakonikolas (USC -> UW Madison)

IMA, June 2019
Can we develop learning algorithms that are robust to a constant fraction of corruptions in the data?
RELIABLE/ADVERSARIAL/SECURE ML

Syrian Hackers Compromise @AP

• Bots influenced US, other elections [Marwick & Lewis ‘17]
• Presidential debates, #MacronLeaks
• Affect trending topics

$136 Billion drop
DATA POISONING

Fake Reviews [Mayzlin et al. ‘14]

Recommender Systems:

[Li et al. ‘16]

Crowdsourcing:

[Wang et al. ‘14]

Malware/spam:

[Nelson et al. ‘08]
OUTLIER DETECTION

• High-dimensional datasets tend to be inherently noisy.

Biological Datasets: POPRES project, HGDP datasets

[November et al., Nature’08];
[Rosenberg et al., Science’02];
[Li et al., Science’08];
[Paschou et al., Medical Genetics’10]

• Outliers: either interesting or can contaminate statistical analysis
The Statistical Learning Problem

- **Input:** sample generated by a **probabilistic model** with unknown $\theta^*$
- **Goal:** estimate parameters $\theta$ so that $\theta \approx \theta^*$

**Question 1:** Is there an **efficient** learning algorithm?

**Question 2:** Are there **tradeoffs** between these criteria?

**Main performance criteria:**
- Sample size
- Running time
- Robustness
ROBUSTNESS IN A GENERATIVE MODEL

Contamination Model:
Let $\mathcal{F}$ be a family of probabilistic models. We say that a set of $N$ samples is $\epsilon$-corrupted from $\mathcal{F}$ if it is generated as follows:

- $N$ samples are drawn from an unknown $F \in \mathcal{F}$
- An omniscient adversary inspects these samples and changes arbitrarily an $\epsilon$-fraction of them.

cf. Huber’s contamination model [1964]
**EXAMPLE: PARAMETER ESTIMATION**

Given samples from an unknown distribution:

- e.g., a 1-D Gaussian $\mathcal{N}(\mu, \sigma^2)$

how do we accurately estimate its parameters?

**empirical mean:**

$$\frac{1}{N} \sum_{i=1}^{N} X_i \rightarrow \mu$$

**empirical variance:**

$$\frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2 \rightarrow \sigma^2$$
The maximum likelihood estimator is asymptotically efficient (1910-1920)

What about errors in the model itself? (1960)
“Robust Estimation of a Location Parameter”
What estimators behave well in a **neighborhood** around the model?
**ROBUST PARAMETER ESTIMATION**

Given *corrupted* samples from a 1-D Gaussian:

\[ \mathcal{N}(\mu, \sigma^2) \]

can we accurately estimate its parameters?
Do the empirical mean and empirical variance work?

No!

A single corrupted sample can arbitrarily corrupt the estimates

But the **median** and **interquartile range** do work
Fact [Folklore]: Given a set $S$ of $N$ $\epsilon$-corrupted samples from a one-dimensional Gaussian $\mathcal{N}(\mu, \sigma^2)$ with high constant probability we have that:

$$|\hat{\mu} - \mu| \leq O\left(\epsilon + \sqrt{1/N}\right) \cdot \sigma$$

where $\hat{\mu} = \text{median}(S)$. 

What about robust estimation in high-dimensions?
**Gaussian Robust Mean Estimation**

**Robust Mean Estimation**: Given an \( \epsilon \) - corrupted set of samples from an *unknown mean*, identity covariance Gaussian \( \mathcal{N}(\mu, I) \) in \( d \) dimensions, recover \( \hat{\mu} \) with

\[
\|\hat{\mu} - \mu\|_2 = O(\epsilon).
\]

**Remark**: Optimal rate of convergence with \( N \) samples is

\[
O(\epsilon) + O\left(\sqrt{d/N}\right)
\]

[Tukey’75]
**Previous Approaches: Robust Mean Estimation**

<table>
<thead>
<tr>
<th>Unknown Mean</th>
<th>Error Guarantee</th>
<th>Running Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pruning</td>
<td>$\Theta(\epsilon \sqrt{d})$</td>
<td>$O(dN)$</td>
</tr>
<tr>
<td>Coordinate-wise Median</td>
<td>$\Theta(\epsilon \sqrt{d})$</td>
<td>$O(dN)$</td>
</tr>
<tr>
<td>Geometric Median</td>
<td>$\Theta(\epsilon \sqrt{d})$</td>
<td>$\text{poly}(d, N)$</td>
</tr>
<tr>
<td>Tukey Median</td>
<td>$\Theta(\epsilon)$</td>
<td>NP-Hard</td>
</tr>
<tr>
<td>Tournament</td>
<td>$\Theta(\epsilon)$</td>
<td>$N^{O(d)}$</td>
</tr>
</tbody>
</table>
All known estimators are either hard to compute or can tolerate a negligible fraction of corruptions.

Is robust estimation algorithmically possible in high-dimensions?
“[…] Only simple algorithms (i.e., with a low degree of computational complexity) will survive the onslaught of huge data sets. This runs counter to recent developments in computational robust statistics. It appears to me that none of the above problems will be amenable to a treatment through theorems and proofs. They will have to be attacked by heuristics and judgment, and by alternative “what if” analyses.[…]”

**This Talk**

Robust estimation in high-dimensions is algorithmically possible!

- First computationally efficient robust estimators that can tolerate a *constant* fraction of corruptions.
- General methodology to detect outliers in high dimensions.

**Meta-Theorem (Informal):** Can obtain *dimension-independent* error guarantees, as long as good data has nice concentration.
[D-Kamath-Kane-Li-Moitra-Stewart, FOCS’16]

Can tolerate a **constant** fraction of corruptions:

- Mean and Covariance Estimation
- Mixtures of Spherical Gaussians, Mixtures of Balanced Product Distributions

[Lai-Rao-Vempala, FOCS’16]

Can tolerate a **mild sub-constant** (**inverse logarithmic**) fraction of corruptions:

- Mean and Covariance Estimation
- Independent Component Analysis, SVD
**THIS TALK: OUR CONTRIBUTIONS**

First computationally efficient robust estimators that can tolerate a *constant* fraction of corruptions.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>[D-Kamath-Kane-Li-Moitra-Stewart, FOCS’16]</td>
<td>Mean and Covariance Estimation</td>
</tr>
<tr>
<td>[D-Kamath-Kane-Li-Moitra-Stewart, ICML’17]</td>
<td></td>
</tr>
<tr>
<td>[D-Kamath-Kane-Li-Moitra-Stewart, SODA’18]</td>
<td></td>
</tr>
<tr>
<td>[Cheng-D-Kane-Stewart, NIPS’18]</td>
<td>Directed Graphical Models</td>
</tr>
<tr>
<td>[D-Kane-Stewart, FOCS’17]</td>
<td>Computational-Robustness Tradeoffs</td>
</tr>
<tr>
<td>[D-Kane-Stewart, STOC’18]</td>
<td>Malicious PAC Learning</td>
</tr>
<tr>
<td>[D-Kane-Stewart, STOC’18]</td>
<td>Clustering in Mixture Models</td>
</tr>
<tr>
<td>[D-Kong-Stewart, SODA’19]</td>
<td>Linear Regression</td>
</tr>
<tr>
<td>[D-Kamath-Kane-Li-Steinhardt-Stewart, ICML’19]</td>
<td>Stochastic Optimization</td>
</tr>
<tr>
<td>[Cheng-D-Ge, SODA’19; CDG+Woodruff, COLT’19]</td>
<td><em>Near-Linear</em> Time Algorithms</td>
</tr>
</tbody>
</table>
OUTLINE

Part I: Introduction
• Motivation
• Robust Statistics in Low and High Dimensions
• Our Contributions

Part II: Efficient Robust Estimation of Mean and Covariance
• Robust Mean Estimation in High-Dimensions
• General Recipe: Iterative Filtering
• Robust Covariance Estimation
• Empirical Results

Part III: Beyond Robust Statistics
• Extensions
• Clustering in Mixture Models
• Stochastic Optimization

Part IV: Summary and Future Directions
OUTLINE

Part I: Introduction
• Motivation
• Robust Statistics in Low and High Dimensions
• Our Contributions

Part II: Efficient Robust Estimation of Mean and Covariance
• Robust Mean Estimation in High-Dimensions
• General Recipe: Iterative Filtering
• Robust Covariance Estimation
• Empirical Results

Part III: Beyond Robust Statistics
• Extensions
• Clustering in Mixture Models
• Stochastic Optimization

Part IV: Summary and Future Directions
ROBUST MEAN ESTIMATION: GAUSSIAN CASE

Problem: Given data $x_1, \ldots, x_N \in \mathbb{R}^d$, of which $(1 - \epsilon)N$ come from some distribution $D$, estimate mean $\mu$ of $D$.

Theorem: Let $\epsilon < 1/2$. If $N = \Omega(d/\epsilon^2)$ and $D = \mathcal{N}(\mu, I)$, then can efficiently recover $\hat{\mu}$ with

$$||\hat{\mu} - \mu||_2 = O(\epsilon).$$

Error Guarantee Independent of $d$!

[D-Kamath-Kane-Li-Moitra-Stewart, SODA’18]
Robust Mean Estimation: Sub-Gaussian Case

Problem: Given data $x_1, \ldots, x_N \in \mathbb{R}^d$, of which $(1 - \epsilon)N$ come from some distribution $D$, estimate mean $\mu$ of $D$.

Theorem: Let $\epsilon < 1/2$. If $N = \Omega(d/\epsilon^2)$ and $D$ is sub-Gaussian with identity covariance, then can efficiently recover $\hat{\mu}$ with

$$||\hat{\mu} - \mu||_2 = O(\epsilon \sqrt{\log(1/\epsilon)}) .$$

Information-theoretically optimal error, even in one-dimension.

[D-Kamath-Kane-Li-Moitra-Stewart, FOCS’16, ICML’17]
**Robust Mean Estimation: General Case**

**Problem:** Given data $x_1, \ldots, x_N \in \mathbb{R}^d$, of which $(1 - \epsilon)N$ come from some distribution $D$, estimate mean $\mu$ of $D$.

**Theorem:** Let $\epsilon < 1/2$. If $N = \Omega(d/\epsilon)$, and $D$ has covariance $\Sigma \preceq \sigma^2 \cdot I$, then can efficiently recover $\hat{\mu}$ with

$$\|\hat{\mu} - \mu\|_2 = O(\sigma \cdot \sqrt{\epsilon}).$$

- Sample-optimal, even without corruptions.
- Information-theoretically optimal error, even in one-dimension.

[D-Kamath-Kane-Li-Moitra-Stewart, ICML’17]
Outlier Detection?
Naïve Outlier Removal

Gaussian Annulus Theorem: $\Pr_{X \sim \mathcal{N}(\mu, I)} \left[ \|X\|_2^2 - d > t \right] \leq 2e^{-\Omega \left( \min \left\{ \frac{t^2}{d}, t \right\} \right)}$
**GENERAL RECIPE: ITERATIVE FILTERING**

**Iterative Two-Step Procedure:**

**Step #1:** Find certificate of robustness of “standard” estimator

**Step #2:** If certificate is violated, detect and remove outliers
Iterate on “cleaner” dataset.

General recipe that works for fairly general settings.

Let’s see how this works for robust mean estimation.
CERTIFICATE OF ROBUSTNESS

Detect when the empirical estimator *may* be compromised

\[ \hat{\mu} \triangleq \frac{1}{N} \sum_{i=1}^{N} X_i \]

- \( \bullet \) = uncorrupted
- \( \bullet \) = corrupted

There is a direction of large (> 1) variance
Idea #1: If the empirical covariance is “close to what it should be”, then the empirical mean works.
Key Lemma: Let $X_1, X_2, \ldots, X_N$ be an $\epsilon$-corrupted set of samples from $\mathcal{N}(\mu, I)$ and $N = \Omega(d/\epsilon^2)$, then for

\begin{align*}
\text{(1)} \quad \hat{\mu} &\triangleq \frac{1}{N} \sum_{i=1}^{N} X_i \\
\text{(2)} \quad \hat{\Sigma} &\triangleq \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu})(X_i - \hat{\mu})^T
\end{align*}

with high probability we have:

$$
\|\hat{\Sigma}\|_2 \leq 1 + O(\epsilon \log(1/\epsilon)) \quad \Rightarrow \quad \|\hat{\mu} - \mu\|_2 \leq O(\epsilon \sqrt{\log(1/\epsilon)})
$$

**Take-away:** An adversary needs to corrupt the second empirical moment in order to corrupt the first empirical moment.
**Key Lemma:** Let $X_1, X_2, \ldots, X_N$ be an $\epsilon$-corrupted set of samples from $\mathcal{N}(\mu, I)$ and $N = \Omega(d/\epsilon^2)$, then for

\begin{align*}
(1) & \quad \hat{\mu} \triangleq \frac{1}{N} \sum_{i=1}^{N} X_i \\
(2) & \quad \hat{\Sigma} \triangleq \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu})(X_i - \hat{\mu})^T
\end{align*}

with high probability we have:

$$
\|\hat{\Sigma}\|_2 \leq 1 + \delta \quad \Rightarrow \quad \|\hat{\mu} - \mu\|_2 \leq O(\sqrt{\delta \epsilon} + \epsilon \sqrt{\log(1/\epsilon)})
$$

**Take-away:** An adversary needs to corrupt the second empirical moment in order to corrupt the first empirical moment.
Idea #2: Removing *any* small constant fraction of good points does not move the empirical mean and covariance by much.
Idea #3: Iteratively “remove outliers” in order to “fix” the empirical covariance.
FILTERING (I)

Either output empirical mean, or remove many outliers.

**Filtering Approach:** Suppose that:

\[ \| \hat{\Sigma} \|_2 \geq 1 + \Omega(\epsilon \log(1/\epsilon)) \]

Let \( v^* \) be the direction of maximum variance.

cf. [Klivans-Long-Servedio’09]
FILTERING (II)

Either output empirical mean, or remove many outliers.

Filtering Approach: Suppose that:

\[ ||\hat{\Sigma}||_2 \geq 1 + \Omega(\epsilon \log(1/\epsilon)) \]

Let \( \nu \) be the direction of maximum variance.

- Project all the points on the direction of \( \nu \).
- Find a threshold \( T \) such that
  \[ \Pr_{x \sim \mathcal{U}\mathcal{S}} [ |\nu \cdot x - \text{median}(\nu \cdot x)| > T ] \geq 3e^{-T^2/2} \]
- Throw away all points such that
  \[ |\nu \cdot x - \text{median}(\nu \cdot x)| > T \]
- Iterate on new dataset.
FILTERING (III)

Either output empirical mean, or remove many outliers.

**Filtering Approach:** Suppose that:

\[ \| \hat{\Sigma} \|_2 \geq 1 + \Omega(\varepsilon \log(1/\varepsilon)) \]

**Claim:** We filter out more corrupted than good points.

After a number of iterations, we have removed all *consequential* corrupted points.

Eventually the empirical mean works
**SUMMARY: ROBUST MEAN ESTIMATION**

**Certificate for Robustness:**

“Spectral norm of empirical covariance is what it should be.”

**Exploiting the Certificate:**

- Check if certificate is satisfied.
- If violated, find “subspace” where behavior of outliers different than behavior of inliers.
- Use it to detect and remove outliers.
- Iterate on “cleaner” dataset.
**ROBUST COVARIANCE ESTIMATION**

**Problem:** Given data $x_1, \ldots, x_N \in \mathbb{R}^d$, of which $(1 - \epsilon)N$ come from some distribution $D$, estimate covariance $\Sigma$ of $D$.

**Theorem:** Let $\epsilon < 1/2$. If $N = \Omega(d^2 / \epsilon^2)$, then can efficiently recover $\hat{\Sigma}$ such that

$$\| \Sigma^{-1/2}(\hat{\Sigma} - \Sigma)\Sigma^{-1/2} \|_F = f(\epsilon),$$

where $f$ depends on the concentration of $D$.

**Main Idea:** Filtering using *fourth-order moment tensors*
EXPERIMENTS

Being Robust (in High Dimensions) Can Be Practical
D., Kamath, Kane, Li, Moitra, Stewart, ICML’17
SYNTHETIC EXPERIMENTS: UNKNOWN MEAN

Error rates on synthetic data (unknown mean):

\[ \mathcal{N}(\mu, I) \ + \ 10\% \ noise \]
SYNTHETIC EXPERIMENTS: UNKNOWN MEAN

Error rates on synthetic data (unknown mean):

\[ \text{Excess } \ell_2 \text{ error} \]

\[ \text{Dimension} \]

- Filtering
- LRV Mean
- Sample mean w/ noise
- Pruning
- RANSAC
- Geometric Median
SYNTHETIC EXPERIMENTS: UNKNOWN COVARIANCE (I)

Error rates on synthetic data (unknown covariance, isotropic):

\[ \mathcal{N}(0, \Sigma) + 10\% \text{ noise} \]

\[ \downarrow \]

close to identity
SYNTHETIC EXPERIMENTS: UNKNOWN COVARIANCE (I)

Error rates on synthetic data (unknown covariance, isotropic):

- Filtering
- LRVCov
- Sample covariance w/ noise
- Pruning
- RANSAC
SYNTHETIC EXPERIMENTS: UNKNOWN COVARIANCE (II)

Error rates on synthetic data (unknown covariance):

\[ \mathcal{N}(0, \Sigma) + 10\% \text{ noise} \]

far from identity
SYNTHETIC EXPERIMENTS: UNKNOWN COVARIANCE (II)

Error rates on synthetic data (unknown covariance, anisotropic):

![Graphs showing error rates against dimension for different methods, including Filtering, LRVcov, Sample covariance w/ noise, Pruning, and RANSAC.](image)
REAL DATA EXPERIMENTS

[Novembre et al. ’08]: Take top two singular vectors of people x SNP matrix (POPRES)

“Genes Mirror Geography in Europe”
EXPERIMENTS: PRUNING PROJECTION

A comparison of error rate on semi-synthetic data:

[Diagram showing a scatter plot labeled 'Pruning Projection']

[Map of Europe with countries colored]
EXPERIMENTS: RANSAC PROJECTION

A comparison of error rate on semi-synthetic data:
EXPERIMENTS: ROBUST PCA (XCS)

A comparison of error rate on semi-synthetic data:
EXPERIMENTS: FILTER PROJECTION
OUTLINE

Part I: Introduction
• Motivation
• Robust Statistics in Low and High Dimensions
• Our Contributions

Part II: Efficient Robust Estimation of Mean and Covariance
• Robust Mean Estimation in High-Dimensions
• General Recipe: Iterative Filtering
• Robust Covariance Estimation
• Empirical Results

Part III: Beyond Robust Statistics
• Extensions
• Clustering in Mixture Models
• Stochastic Optimization

Part IV: Summary and Future Directions
Robust Unsupervised Learning

Robustly Learning Graphical Models [Cheng-D-Kane-Stewart’16]

Clustering in Mixture Models [D-Kane-Stewart’18]

Computational/Statistical-Robustness Tradeoffs [D-Kane-Stewart’17]
ROBUST UNSUPERVISED LEARNING

Clustering in Mixture Models
[D-Kane-Stewart’18]
# Clustering in Mixture Models

Under what assumptions can we disentangle mixture models?

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Separation</th>
<th>Robustness?</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spherical Gaussians</td>
<td>( k^{1/4} )</td>
<td>NO</td>
<td>[VW02]</td>
</tr>
<tr>
<td>Second Moments</td>
<td>( k^{1/2} )</td>
<td>NO</td>
<td>[AM05]</td>
</tr>
<tr>
<td></td>
<td>( k^{1/2} )</td>
<td>YES</td>
<td>[CSV’17, DKS’18]</td>
</tr>
<tr>
<td>Spherical Gaussians</td>
<td>( \sqrt{\log k} )</td>
<td>YES</td>
<td>[DKS’18, HL’18,KSS’18]</td>
</tr>
</tbody>
</table>
**Robustness As A Lens**

Under what assumptions can we disentangle mixture models?

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Separation</th>
<th>Robustness?</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second Moments</td>
<td>$k^{1/2}$</td>
<td>YES</td>
<td>[DKS’18]</td>
</tr>
<tr>
<td>Spherical Gaussians</td>
<td>$\sqrt{\log k}$</td>
<td>YES</td>
<td>[DKS’18]</td>
</tr>
</tbody>
</table>

**Main Idea:** Remove Outliers from a dataset when the *majority* of the points are corrupted (*List-Decodable Learning*).
**Robust Supervised Learning**

Malicious PAC Learning
[Klivans-Long-Servedio’10, Awasthi-Balcan-Long’14, D-Kane-Stewart’18]

Robust Linear Regression
[Klivans-Kothari-Meka’18, D-Kong-Stewart’18]

Stochastic Convex Optimization
[Prasad-Suggala-Balakrishnan-Ravikumar’18, D-Kamath-Kane-Li-Steinhardt-Stewart’18]
Robust Supervised Learning

Sever: A Robust Meta-Algorithm for Stochastic Optimization.
[D-Kamath-Kane-Li-Steinhardt-Stewart, ICML’19]
**ROBUST STOCHASTIC CONVEX OPTIMIZATION**

**Problem:** Given loss function $\ell(X, w)$ and $\epsilon$-corrupted samples from a distribution $\mathcal{D}$ over $X$, minimize $f(w) = \mathbb{E}_{X \sim \mathcal{D}}[\ell(X, w)]$

**Difficulty:** Corrupted data can move the gradients.

**Theorem:** Suppose $\ell$ is convex and $\text{Cov}_{X \sim \mathcal{D}}[\nabla \ell(X, w)] \preceq \sigma^2 \cdot I$. Under mild assumptions on $\mathcal{D}$, can recover a point such that

$$f(\hat{w}) - \min_w f(w) \leq O(\sigma \sqrt{\epsilon}).$$

**Main Idea:** Filter at minimizer of empirical risk.
EXPERIMENTS: RIDGE REGRESSION

Regression: Synthetic data

Regression: Drug discovery data

Regression: Drug discovery data, attack targeted against SEVER

uncorrupted  l2  loss  gradientCentered  SEVER
SUBSEQUENT WORKS

- **Graphical Models** [Cheng-D-Kane-Stewart’16, D-Kane-Stewart’18]
- **Sparse models (e.g., sparse PCA, sparse regression)** [Balakrishan-Du-Li-Singh’17, Liu-Shen-Li-Caramanis’18, D-Karmalkar-Kane-Price’19]
- **List-Decodable Learning** [Charikar-Steinhardt-Valiant ’17, Meister-Valiant’18, D-Kane-Stewart’18]
- **Robust PAC Learning** [Klivans-Long-Servedio’10, Awasthi-Balcan-Long’14, D-Kane-Stewart’18]
- “**Robust estimation via SoS**” (higher moments, learning mixture models) [Hopkins-Li’18, Kothari-Steinhardt-Steurer’18, …]
- “**SoS Free**” learning of mixture models [D-Kane-Stewart’18]
- **Robust Regression** [Klivans-Kothari-Meka’18, D-Kong-Stewart’18, …]
- **Robust Stochastic Optimization** [Prasad-Suggala-Balakrishnan-Ravikumar’18, D-Kamath-Kane-Li-Steinhard-Stewart’18]
- …
OUTLINE

Part I: Introduction
• Motivation
• Robust Statistics in Low and High Dimensions
• Our Contributions

Part II: Efficient Robust Estimation of Mean and Covariance
• Robust Mean Estimation in High-Dimensions
• General Recipe: Iterative Filtering
• Robust Covariance Estimation
• Empirical Results

Part III: Beyond Robust Statistics
• Extensions
• Clustering in Mixture Models
• Stochastic Optimization

Part IV: Summary and Future Directions
SUMMARY AND CONCLUSIONS

• First computationally efficient robust estimators with dimension-independent error guarantees.

• General recipe for various high-dimensional problems.

• Lens of robustness resolves long-standing open problems.

• Practical applications in data analysis.
Concrete Open Problem:
Robustly Learn a Mixture of 2 *Arbitrary* Gaussians

- Pick your favorite high-dimensional learning problem for which a (non-robust) efficient algorithm is known.
- Make it robust!
FUTURE DIRECTIONS

General Algorithmic Theory of Robustness

How can we robustly learn rich representations of data, based on natural hypotheses about the structure in data?

Can we robustly test our hypotheses about structure in data before learning?

Concrete Challenges:
• Richer Families of Problems and Models
• Connections to Non-convex Optimization
• Relation to other Notions of Algorithmic Stability (Differential Privacy, Adaptive Data Analysis)
• Further applications (ML Security, Computer Vision)

Thank you!
Questions?
Related Materials:

- **Simons Institute Bootcamp Tutorial:**
  
  [www.iliasdiakonikolas.org/simons-tutorial-robust.html](http://www.iliasdiakonikolas.org/simons-tutorial-robust.html)

- **TTI-Chicago Summer Workshop Program**
  

- **Simons Institute, Foundations of Data Science Program**
  

(co-organized with Montanari, Candes, Vempala)
Upcoming Events:

- **STOC’19 Tutorial** (June 23, Phoenix)
  

- **ISIT’19 Tutorial** (July 7, Maison de la Mutualité, Paris)
  
Question: What is the effect of additive and subtractive corruptions?

Let's study the simplest possible example of $\mathcal{N}(\mu, 1)$.

**Subtractive** errors at rate $\epsilon$ can:
- Move the mean by at most $O(\epsilon \sqrt{\log(1/\epsilon)})$
- Increase the variance by $O(\epsilon)$ and decrease it by at most $O(\epsilon \log(1/\epsilon))$

**Additive** errors at rate $\epsilon$ can:
- Move the mean arbitrarily
- Increase the variance arbitrarily and decrease it by at most $O(\epsilon)$