Sum-of-Squares Approach
for
Robust Mean Estimation

Pravesh Kothari
Princeton/IAS

Sum-of-Squares Approach
for
Parameter Estimation Problems

Pravesh Kothari
Princeton/IAS

Based on joint works with Adam Klivans, Raghu Meka, David Steurer and Jacob Steinhardt.

Machine Learning

DATA

STRUCTURE

- *documents*
- *music*
- *social network*

Learning →

- *topics*
- *genres*
- *communities*

Parameter Estimation

DATA

STRUCTURE

"estimation"

$x_1, x_2, \ldots, x_m \in \mathbb{R}^d$ → $\Theta \in \mathbb{R}^p$

iid samples from $M(\Theta, \ldots)$

$M(\Theta, \ldots)$

"generation"

Parameter Estimation

DATA

STRUCTURE

$x_1, x_2, \ldots, x_m \in \mathbb{R}^d$ —"estimation"→ $\Theta \in \mathbb{R}^p$

Machine Learning

mixture models, topic models, independent component analysis, principal component analysis, compressive sensing, matrix completion, regression, ***robust*** versions,…

Parameter Estimation

DATA

STRUCTURE

"estimation"

$x_1, x_2, \ldots, x_m \in \mathbb{R}^d \longrightarrow \Theta \in \mathbb{R}^p$

Machine Learning

Cryptography security of pseudorandom generators,…

Parameter Estimation

DATA

STRUCTURE

"estimation"

$x_1, x_2, \ldots, x_m \in \mathbb{R}^d$ → $\Theta \in \mathbb{R}^p$

Machine Learning

Cryptography

avg-case complexity planted clique, refuting random CSPs,…

Parameter Estimation

DATA → "estimation" → **STRUCTURE**

$x_1, x_2, \ldots, x_m \in \mathbb{R}^d$ → $\Theta \in \mathbb{R}^p$

SAMPLE COMPLEXITY

how much data is required for recovering Θ ?

COMPUTATIONAL COMPLEXITY

is there an efficient algorithm for recovering Θ ?

CO

ca

SUM-OF-SQUARES METHOD

a unified approach for parameter estimation

SoS for Parameter Estimation

ROBUST STATISTICS

MOMENT ESTIMATION [K-Steurer'18]

CLUSTERING MIXTURE MODELS [Hopkins-Li'18],[K-Steinhardt'18]

REGRESSION [Klivans-K-Meka'18]

SPARSE RECOVERY [Klivans-Karmalkar-K'18]

SoS for Parameter Estimation

MACHINE LEARNING

MOMENT ESTIMATION	[K-Steurer'18]
CLUSTERING MIXTURE MODELS	[Hopkins-Li'18],[K-Steinhardt'18]
REGRESSION	[Klivans-K-Meka'18]
SPARSE RECOVERY	[Klivans-Karmalkar-K'18]
TENSOR COMPLETION	[Barak-Moitra'15, Potechin-Steurer'16]
TENSOR PCA	[Hopkins-Shi-Steurer'15]
TENSOR DECOMPOSITION	[**Barak-Kelner-Steurer'14**, Ge-Ma'15, Ma-Shi-Steurer'16,]
DICTIONARY LEARNING	

SoS for Parameter Estimation

MACHINE LEARNING

MOMENT ESTIMATION [K-Steurer'18]

CLUSTERING MIXTURE MODELS [Hopkins-Li'18],[K-Steinhardt'18]

REGRESSION [Klivans-K-Meka'18]

SPARSE RECOVERY [Klivans-Karmalkar-K'18]

TENSOR COMPLETION [Barak-Moitra'15, Potechin-Steurer'16]

TENSOR PCA [Hopkins-Shi-Steurer'15]

TENSOR DECOMPOSITION
DICTIONARY LEARNING [**Barak-Kelner-Steurer'14**, Ge-Ma'15, Ma-Shi-Steurer'16,]

COMP. VS STAT. COMPLEXITY GAPS

RANDOM CSPS [Allen-O'Donnell-Witmer'15, [Barak-Chan-K'15] [K-Mori-O'Donnell-Witmer'17]

PLANTED CLIQUE [Barak-Hopkins-Kelner-K-Moitra-Potechin'16]

SPARSE PCA
TENSOR PCA [Hopkins-K-Potechin-Raghavendra-Schramm-Steurer'17]

Know Thy Hammer

Upshots

- Single blueprint for parameter estimation.

  "identifiability to algorithm"

- general tools to prove optimal lower bounds

  "comp. vs stat. gaps"

Downsides

- theoretically efficient, practically slow

  "hammer not a scalpel"

can extract fast practical algorithms sometimes

[Hopkins-Schramm-Shi-Steurer'16],…

ask Sam!

Know Thy Hammer

Upshots

- Single blueprint for parameter estimation.

"identifiability to algorithm"

- general tools to prove optimal lower bounds

"comp. vs stat. gaps"

Downsides

- theoretically efficient, practically slow

"hammer not a scalpel"

can extract fast practical algorithms sometimes

Our Goal

- understand algorithmically exploitable structure in the problem
- uncover fundamental tradeoffs/barriers.

Today

Illustrate Sum-of-Squares Method for *Parameter Estimation*

Parameter Estimation Via SoS

"Simple" Identifiability Proof → **SoS** → **Efficient Algorithm**

Example: *Robust Moment Estimation* [**K**-Steurer'18]

focus on ***mean*** estimation

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

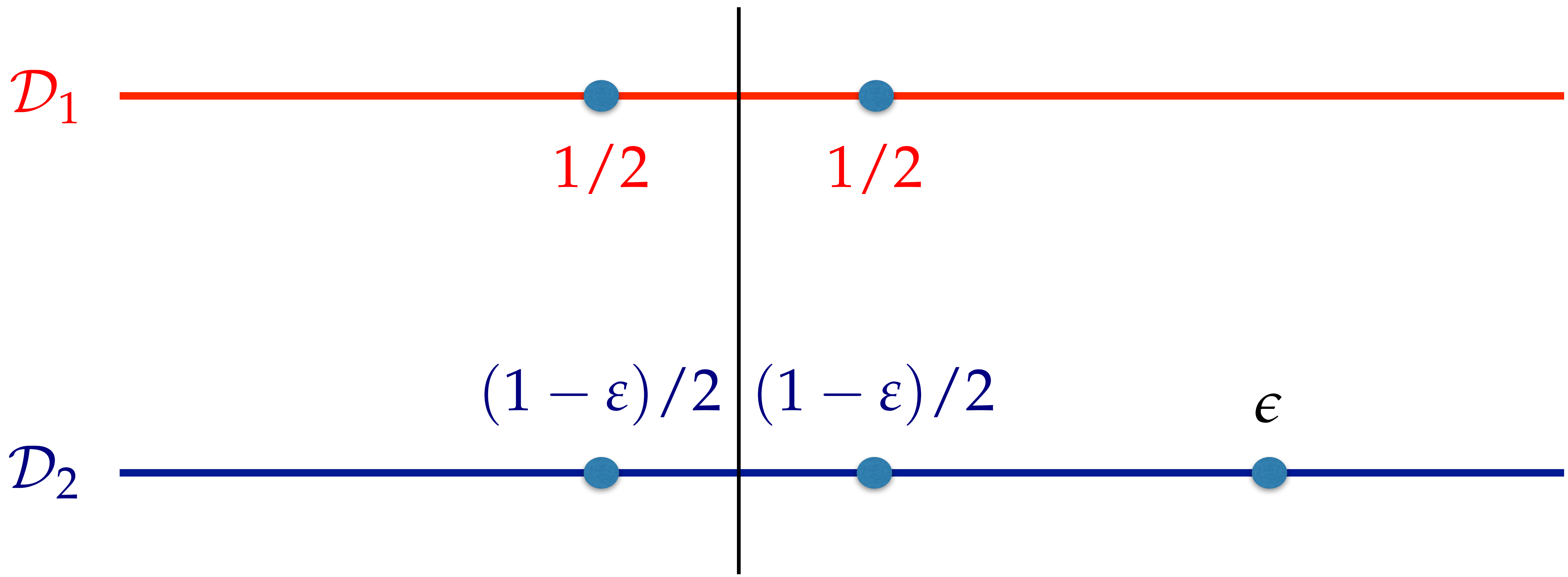
$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Is robust mean estimation possible?



Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Is robust mean estimation possible?

- cannot tell apart distributions ε-close in stat. distance.
- ε-close distributions can have ***arbitrarily*** differing means.

so info. theoretically impossible in general.

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Is robust mean estimation possible?

What we'll do: assume that $\mathcal{D}$ comes from a reasonable family where *tails do not strongly control the mean*.

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

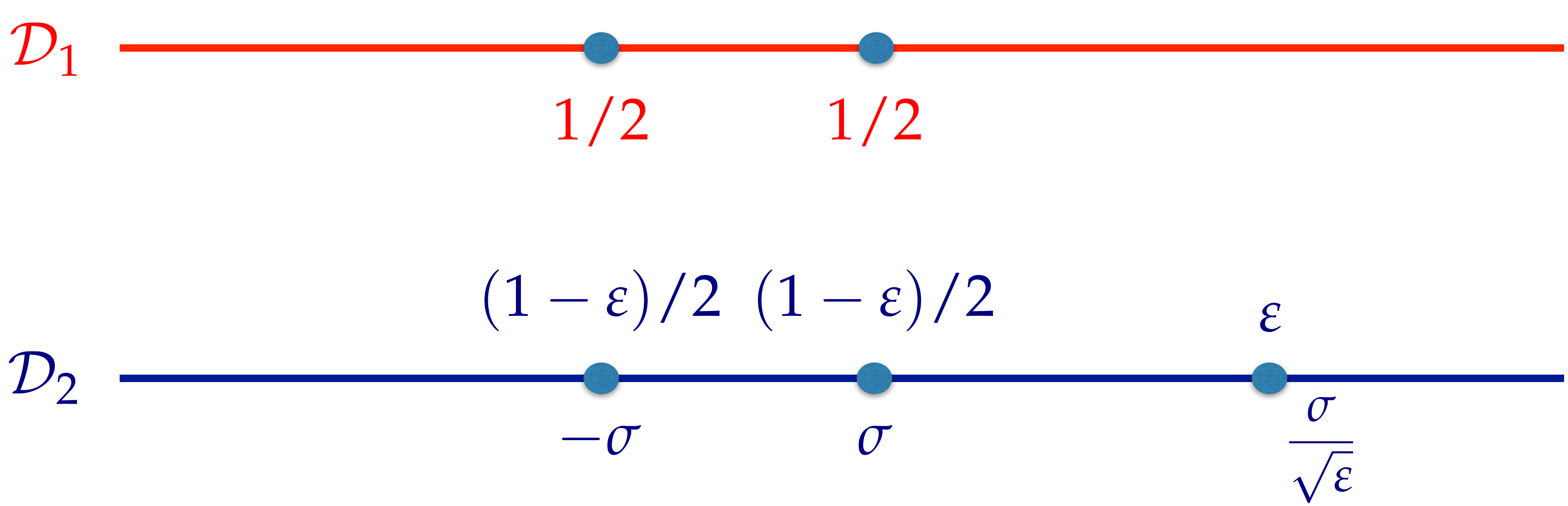
$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Bounded Variance means are $\sim \sigma\sqrt{\epsilon}$ apart.



Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

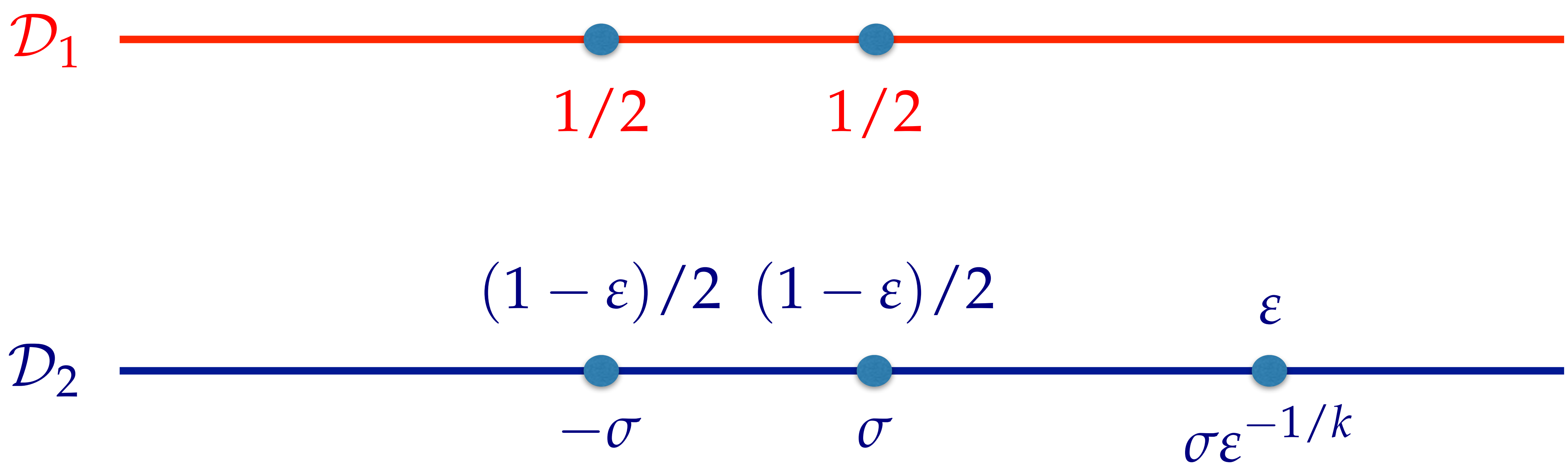
Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Bounded 2k-moments means are $\sim \sigma\epsilon^{1-1/k}$ apart.

$\mathbb{E}(x-\mu)^{2k} \le (Ck)^k(\mathbb{E}(x-\mu)^2)^k$



Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1 - \varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

high dimensional setting

Bounded Moment Distributions

$\mathcal{D}$ has **C**-bounded **2k**-moments, if for every $u \in \mathbb{R}^d$

$$\mathbb{E}_{\mathcal{D}}\langle x - \mu, u\rangle^{2k} \leq (C \cdot k \cdot \mathbb{E}_{\mathcal{D}}\langle x - \mu, u\rangle^2)^k$$

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1-\varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

high dimensional setting

Bounded Moment Distributions

$\mathcal{D}$ has **C**-bounded **2k**-moments, if for every $u \in \mathbb{R}^d$

$$\mathbb{E}_{\mathcal{D}} \langle x - \mu, u \rangle^{2k} \leq (C \cdot k \cdot \mathbb{E}_{\mathcal{D}} \langle x - \mu, u \rangle^2)^k$$

Natural families are bounded for all k.

2k-wise Product Distributions, Sub-gaussian/Sub-exp Families,…

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1 - \varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

A flurry of activity starting with the pioneering papers of

[Diakonikolas-Kane-Kamath-Li-Moitra-Stewart'16] [Lai-Rao-Vempala'16]

Robust Mean Estimation

Setting: unknown distribution $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1 - \varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

A flurry of activity starting with the pioneering papers of

[Diakonikolas-Kane-Kamath-Li-Moitra-Stewart'16] [Lai-Rao-Vempala'16]

will skip a detailed survey and instead give you punchlines.

focus on estimation error for a given dist. family.

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2})\|\Sigma\|^{1/2}$ optimal!

[Lai-Rao-Vempala'16]
[Charikar-Steinhardt-Valiant'17]
[Diakonikolas-Kane-Kamath-Li-Moitra-Stewart'17]

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2})\|\Sigma\|^{1/2}$ optimal!

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon)\sqrt{\log(1/\epsilon)}\|\Sigma\|^{1/2}$ ~optimal!

[Diakonikolas-Kane-Kamath-Li-Moitra-Stewart'16]

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2}) \|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon) \sqrt{\log(1/\epsilon)} \|\Sigma\|^{1/2}$

For covariance estimation, optimal results only for gaussians.

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2})\|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon)\sqrt{\log(1/\epsilon)}\|\Sigma\|^{1/2}$

Bounded 2k-Moments

relates to the hardness of **UG/SSE**.

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2}) \|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon) \sqrt{\log(1/\epsilon)} \|\Sigma\|^{1/2}$

***Certified* Bounded 2k-Moments**

"higher-moment information is algorithmically accessible"

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2}) \|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon) \sqrt{\log(1/\epsilon)} \|\Sigma\|^{1/2}$

***Certified* Bounded 2k-Moments**

Examples

- Gaussians, product distributions on discrete hypercube,…
- k-wise product distributions
- Distributions satisfying **Poincaré** inequality [K-Steinhardt'17]
  includes all ***strongly log-concave*** distributions

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2}) \|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon) \sqrt{\log(1/\epsilon)} \|\Sigma\|^{1/2}$

Certified **Bounded 2k-Moments**

[K-Steurer'18] $\|\hat{\mu} - \mu\| \leq O(\sqrt{Ck}) \cdot \epsilon^{1-\frac{1}{2k}} \cdot \|\Sigma\|^{1/2}$ in time $d^{O(k)}$

optimal!

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2})\|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon)\sqrt{\log(1/\epsilon)}\|\Sigma\|^{1/2}$

Certified **Bounded 2k-Moments**

[K-Steurer'18] $\|\hat{\mu} - \mu\| \leq O(\sqrt{Ck}) \cdot \epsilon^{1-\frac{1}{2k}} \cdot \|\Sigma\|^{1/2}$ in time $d^{O(k)}$

via the SoS method.

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2})\|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon)\sqrt{\log(1/\epsilon)}\|\Sigma\|^{1/2}$

***Certified* Bounded 2k-Moments**

[K-Steurer'18] $\|\hat{\mu} - \mu\| \leq O(\sqrt{Ck}) \cdot \epsilon^{1-\frac{1}{2k}} \cdot \|\Sigma\|^{1/2}$ in time $d^{O(k)}$

optimal results for **covariance** and **higher moment** estimation!

Corollary "outlier-robust *method of moments*"

[Pearson'94],...,[Kalai-Moitra-Valiant'10,Belkin-Sinha'10],...

- Robust Independent Component Analysis.
- Robust Learning of Mixture of Gaussians for *linearly indep*. means.

Robust Mean Estimation

Quick summary of what's known

Bounded Covariance $\|\hat{\mu} - \mu\| \leq O(\epsilon^{1/2}) \|\Sigma\|^{1/2}$

Gaussians $\|\hat{\mu} - \mu\| \leq O(\epsilon) \sqrt{\log(1/\epsilon)} \|\Sigma\|^{1/2}$

***Certified* Bounded 2k-Moments**

[K-Steurer'18] $\|\hat{\mu} - \mu\| \leq O(\sqrt{Ck}) \cdot \epsilon^{1-\frac{1}{2k}} \cdot \|\Sigma\|^{1/2}$ in time $d^{O(k)}$

conceptual power of SoS in robust estimation

- allows algorithmically using higher moment information in data.
- key to improved algorithms for **clustering** mixture models.

Our Goal Today

One algorithm to *robustly* estimate them all…

unified conceptual blueprint, simple proofs.

don't try this on your personal computers yet.

Robust Mean Estimation

Setting: unknown $\mathcal{D}$ on $\mathbb{R}^d$ with unknown mean $\mu \in \mathbb{R}^d$ and cov. Σ

$X = \{x_1, x_2, \ldots, x_m\}$ i.i.d. sample from $\mathcal{D}$

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X.

$y_i = x_i$ for $(1 - \varepsilon)m$ indices i

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Robust Mean Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

Step 2: ***Algorithm Design***

An efficient algorithm to find $\hat{\mu}$.

SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

= a test that only approx. **true** parameters can pass.

= a *certificate* that a purported solution is correct.

what Ilias showed you in the first part today!

SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

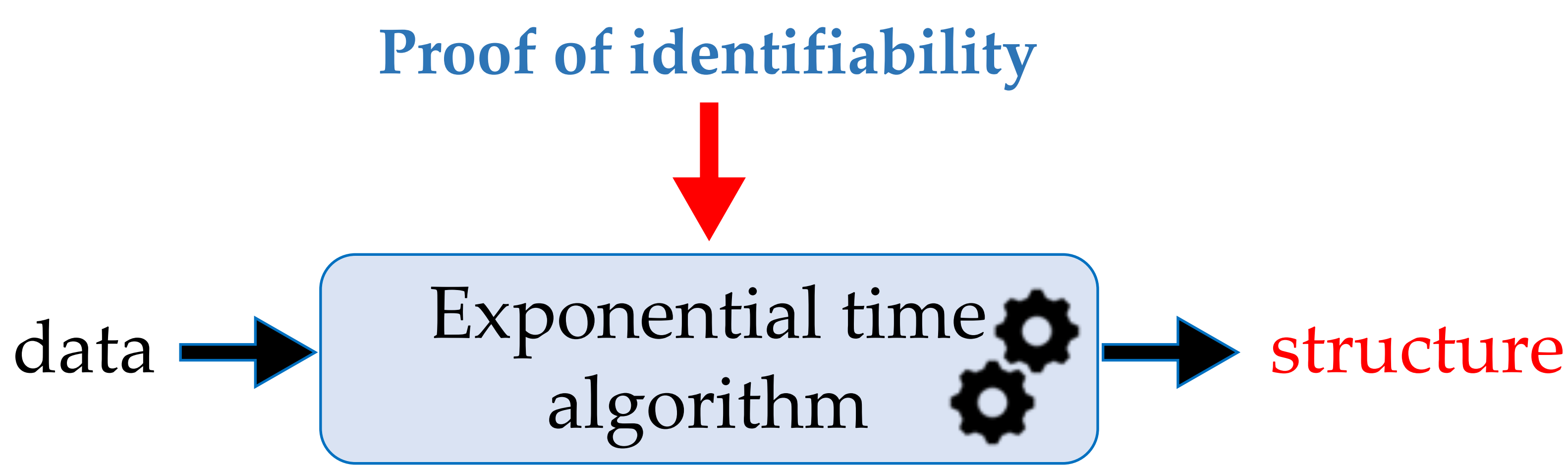
Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

= a test that only approx. **true** parameters can pass.

= a *certificate* that a purported solution is correct.

determines *sample complexity*. Implies that brute-force succeeds.



SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

I'm going to show you a magical world where "P = NP"!

SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of $X \sim \mathcal{D}^m$ with μ, Σ

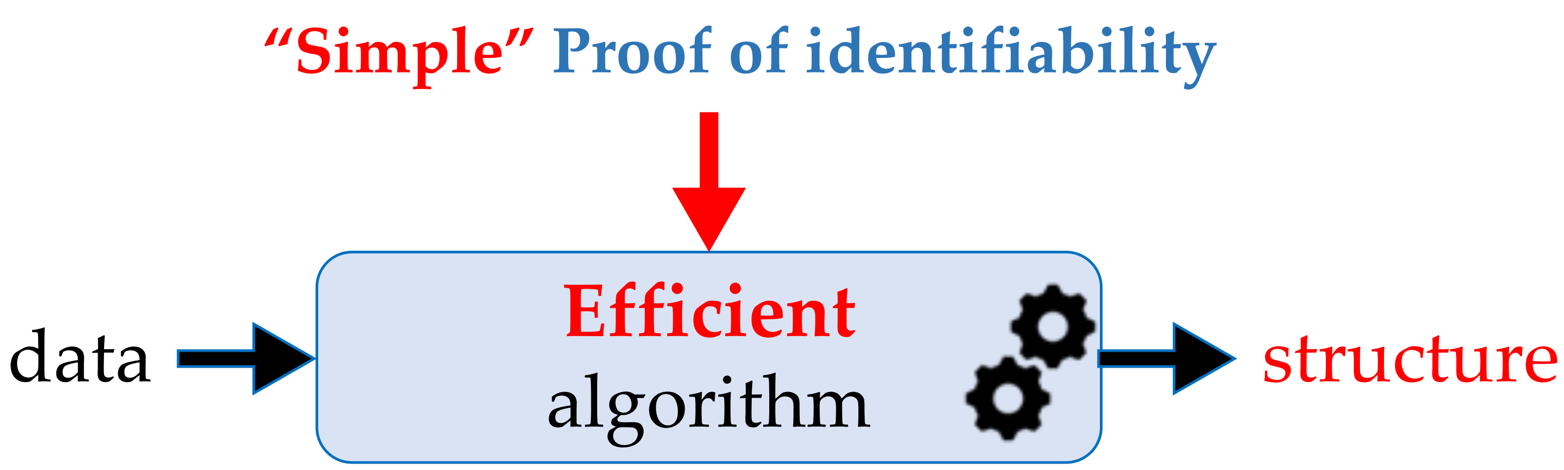
Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

simple (low degree SoS) proof of identifiability = efficient algorithm.



SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

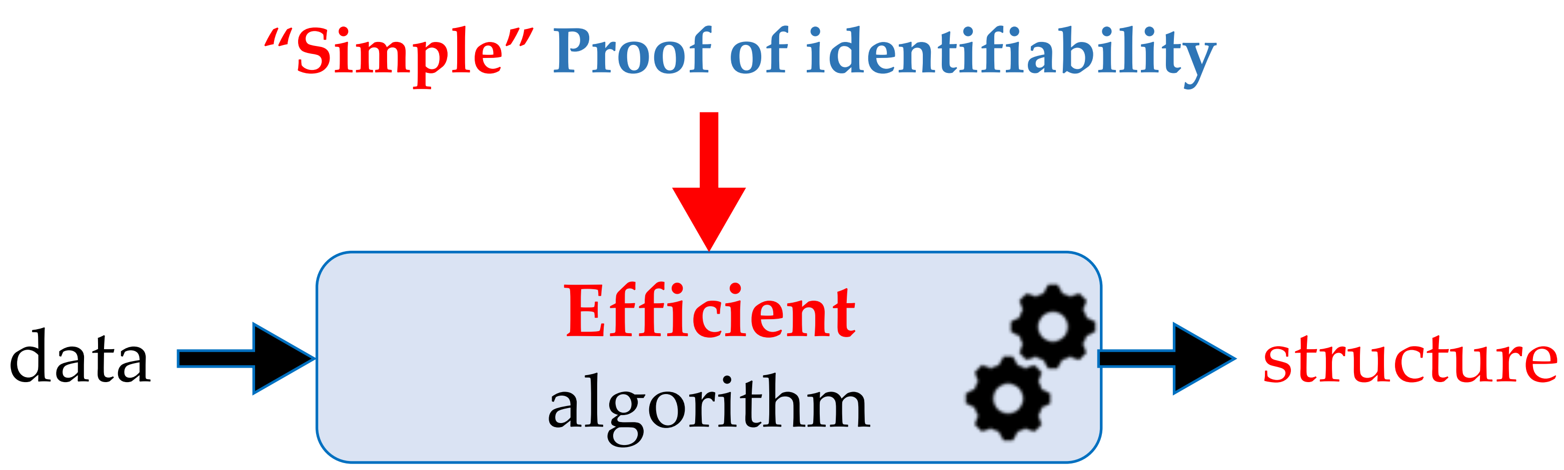
Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely* identifies*** μ up to a small error.

simple (low degree SoS) proof of identifiability = efficient algorithm.

Luckily, our proofs are often simple without additional effort!



SoS Approach to Robust Estimation

Input: $Y = \{y_1, y_2, \ldots, y_m\}$ ε-corruption of X $\sim \mathcal{D}^m$ with μ, Σ

Goal: Compute $\hat{\mu} \in \mathbb{R}^d$ so that $\|\mu - \hat{\mu}\|_2$ is as small as possible.

Standard blueprint for problems in unsupervised learning

Step 1: ***Robust Identifiability***

A small sample Y ***uniquely*** **identifies*** μ up to a small error.

DONE! [Barak-Kelner-Steurer'15],...

Step 2 is problem independent!

~~**Step 2:** ***Algorithm Design***~~

~~An efficient algorithm to find $\hat{\mu}$.~~

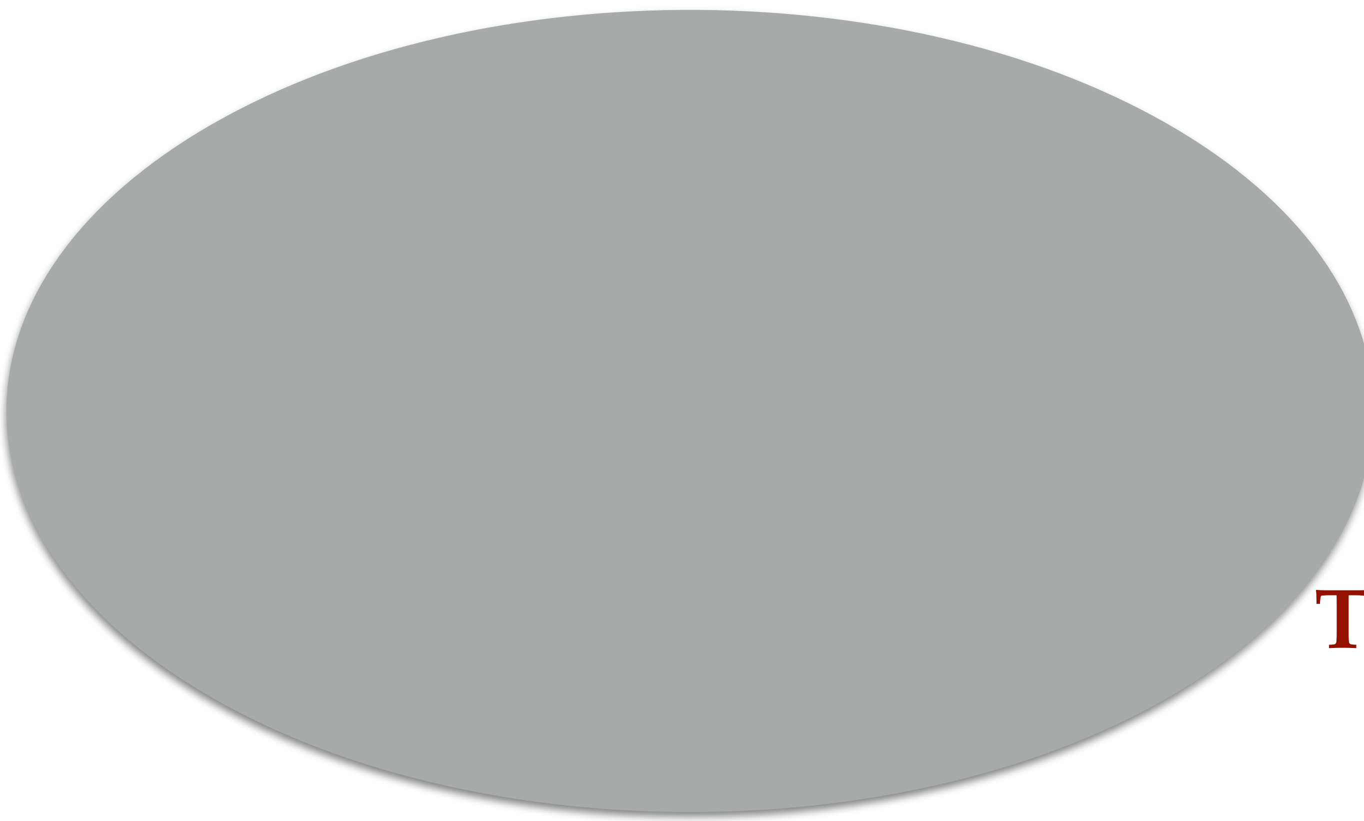
SoS Approach to Robust Estimation

Why does a corrupted sample uniquely* determine the mean?

*up to a small error

SoS Approach to Robust Estimation

Why does a corrupted sample uniquely* determine the mean?

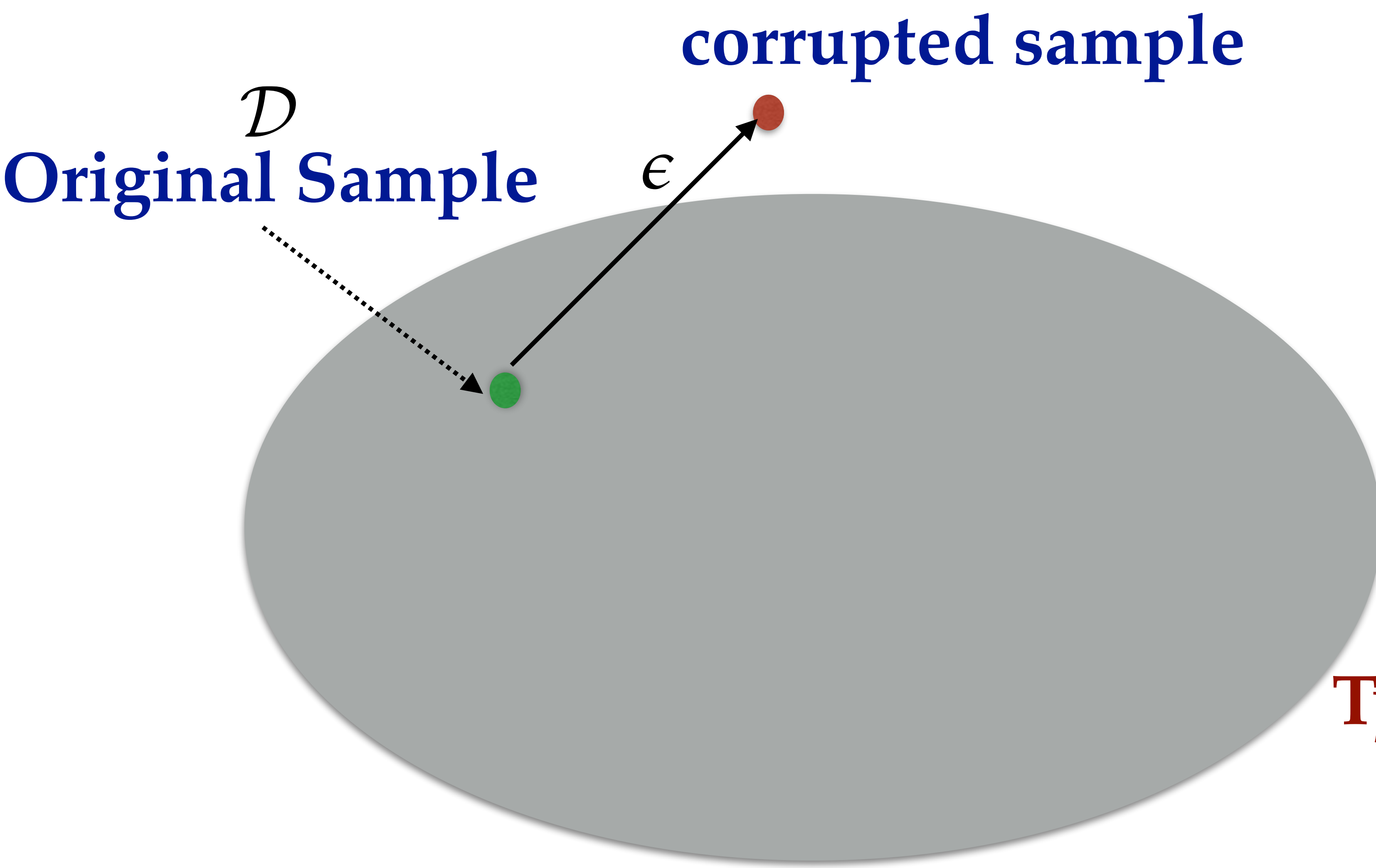


Typical m-samples from bounded covariance distributions

If $m \approx d/\epsilon^2$, the uniform distribution on the **sample** satisfies the bounded variance property whp.

SoS Approach to Robust Estimation

Why does a corrupted sample uniquely* determine the mean?

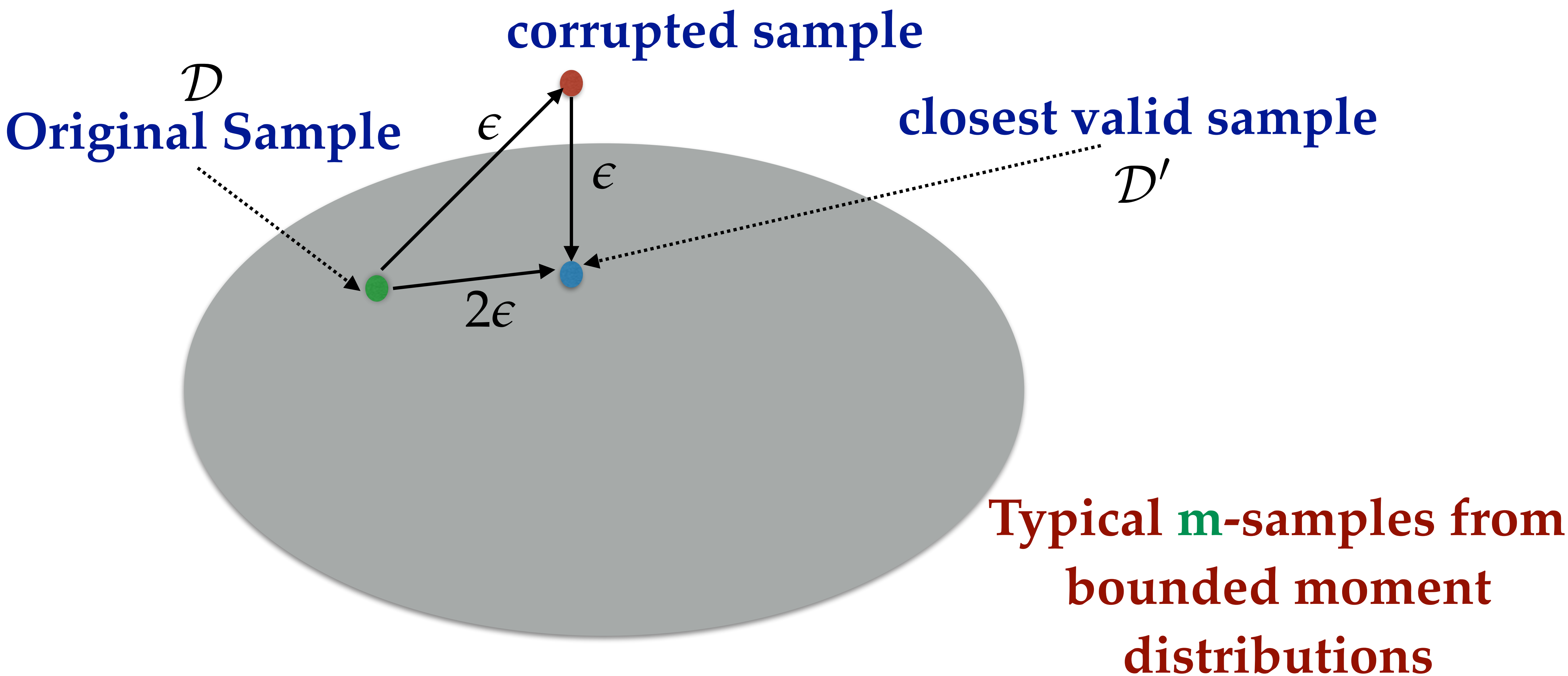


Typical m-samples from bounded moment distributions

If $m \approx d/\epsilon^2$, the uniform distribution on the **sample** satisfies the bounded variance property whp.

SoS Approach to Robust Estimation

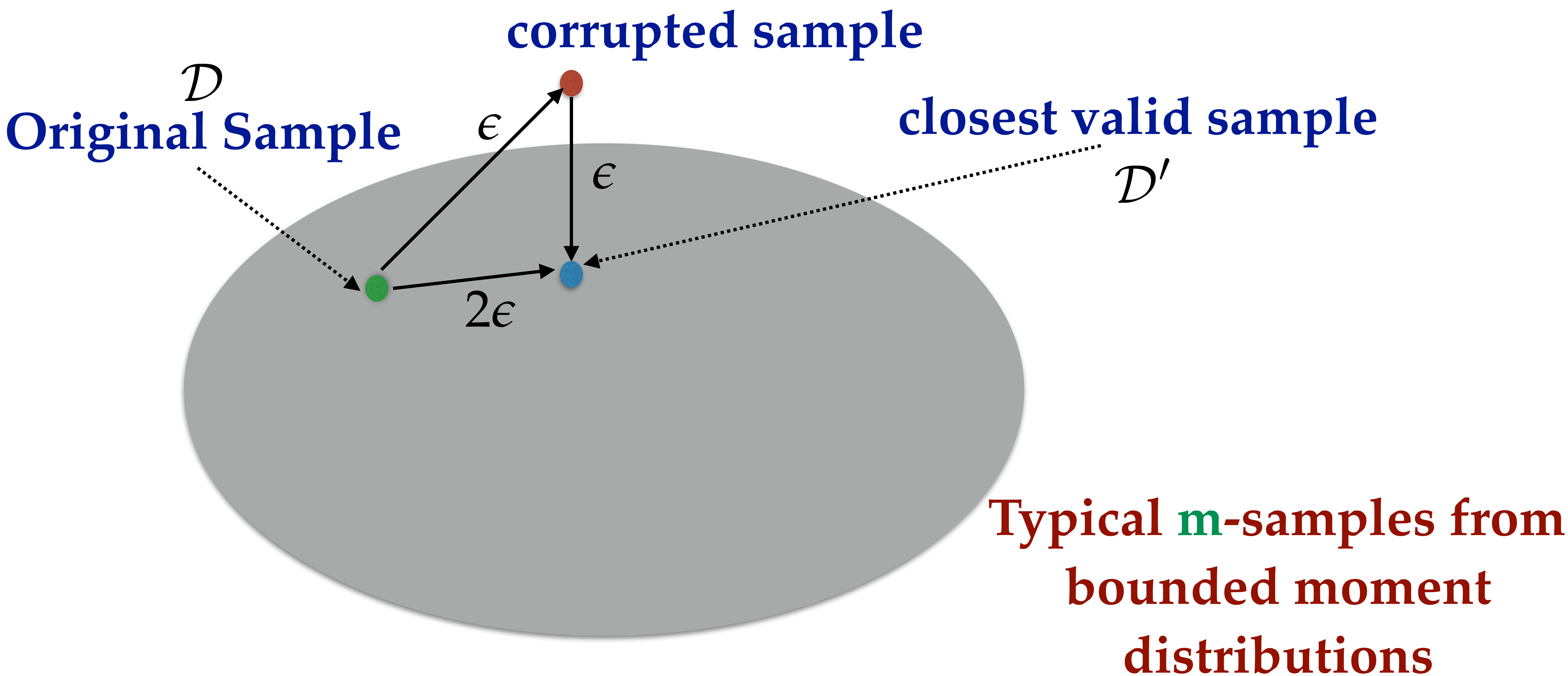
Why does a corrupted sample uniquely* determine the mean?



If $m \approx d/\epsilon^2$, the uniform distribution on the **sample** satisfies the bounded variance property whp.

SoS Approach to Robust Estimation

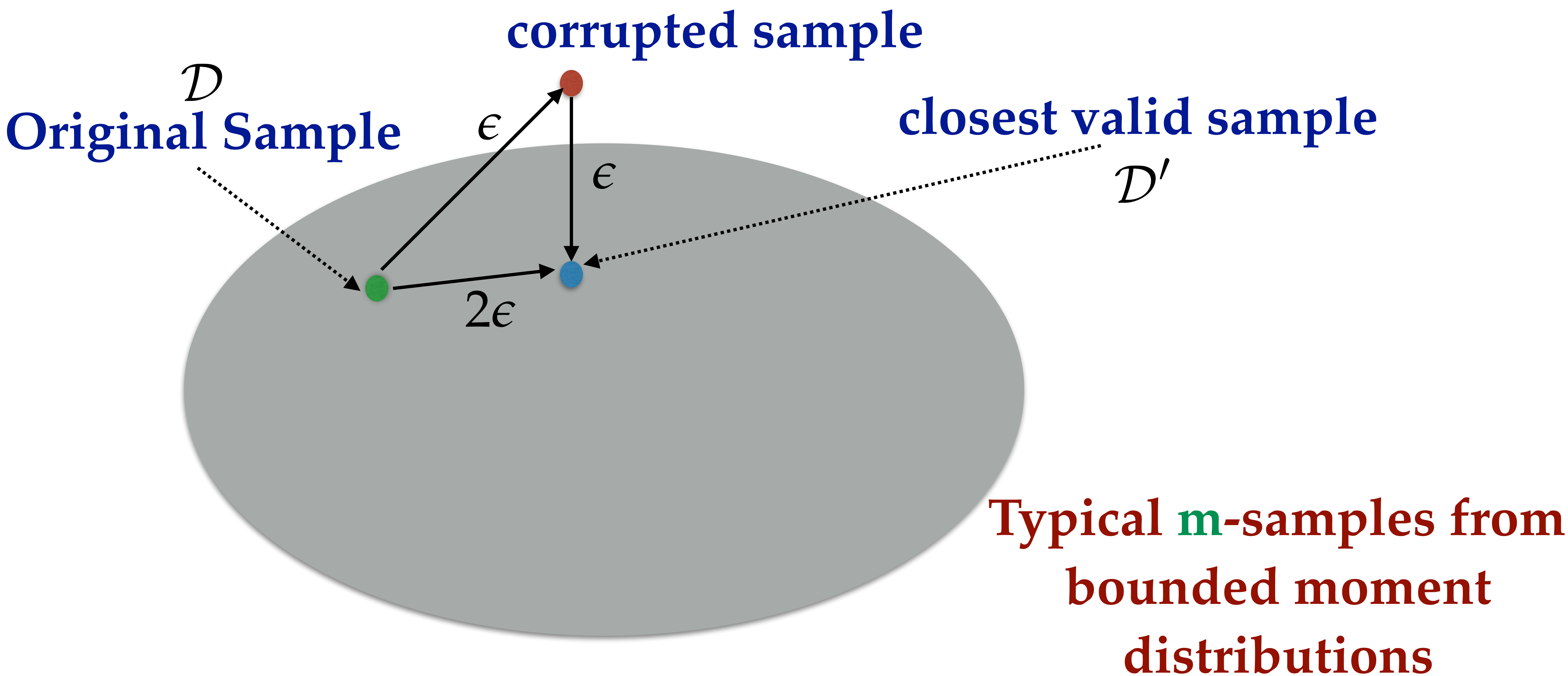
Why does a corrupted sample uniquely* determine the mean?



"Unique Decodability"

SoS Approach to Robust Estimation

Why does a corrupted sample uniquely* determine the mean?



Why do nearby samples have close parameters?

Identifiability for Mean Estimation

Why does a corrupted sample uniquely* determine the mean?

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

Soon we will obtain better guarantees under bounded moment assumptions.

Identifiability for Mean Estimation

Why does a corrupted sample uniquely* determine the mean?

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

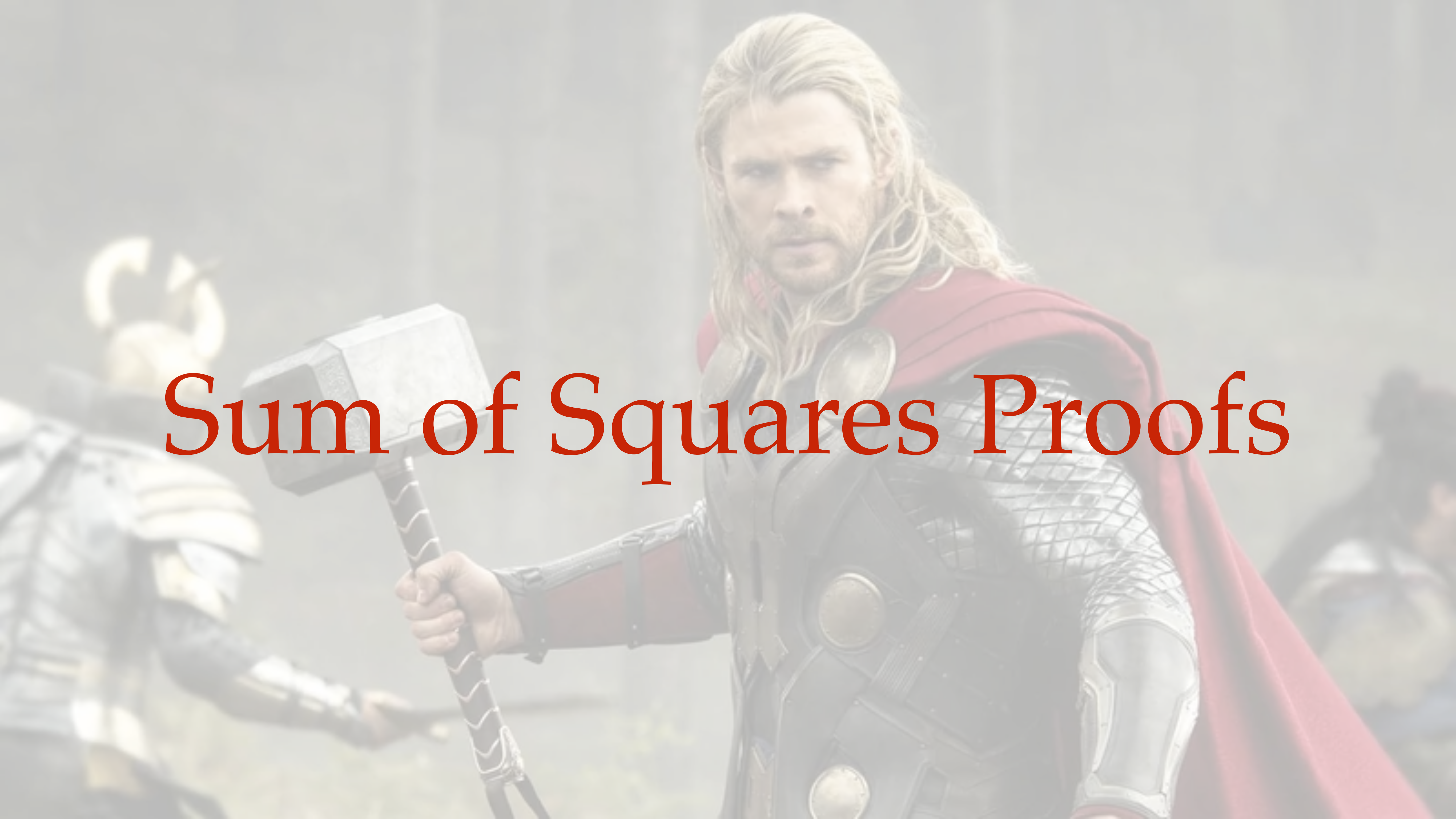
Inefficient Algorithm

1. Find an ϵ-close sample that has the smallest covariance
2. Return its mean.

In 1-D, corresponds to modifying the largest/smallest points.

~ median

Sum of Squares Proofs



Thank you for your attention!

Identifiability for Mean Estimation

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

1) $\mathcal{U}_X$ and $\mathcal{U}_{X'}$ have 1-bounded 4th moments, and

2) $\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{3/4})(\sigma_X + \sigma_{X'})$$

Coming up…

Automatically translate "simple" *identifiability* proofs into algorithms!

What does simple mean?

captured in the sum of squares proof system

- A proof system that reasons about polynomial inequalities
- Degree t proofs can be found in time $d^{O(t)}$
- Many natural inequalities have low-degree SoS proofs

  Holder's, Cauchy-Schwarz, Triangle Inequality, Brascamp-Lieb inequalities…

growing general toolkit for ready to use SoS facts*!

***See notes at sumofsquares.org**

Identifiability for Mean Estimation

Why does a corrupted sample uniquely* determine the mean?

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

Proof By Cauchy-Schwarz

$$\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle = \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i \rangle$$

$$\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\})\right)^{1/2} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle\right)^{1/2}$$

Identifiability for Mean Estimation

Why does a corrupted sample uniquely* determine the mean?

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

Proof By Cauchy-Schwarz

$$\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle = \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i \rangle$$

$$\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\})\right)^{1/2} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle\right)^{1/2}$$

$$\leq \epsilon^{1/2} \cdot (\mathbb{E}_i \langle u, x_i - \mu(X) \rangle) + (\langle u, x'_i - \mu(X') \rangle + \langle u, \mu(X) - \mu(X') \rangle)^{1/2}$$

Identifiability for Mean Estimation

Why does a corrupted sample uniquely* determine the mean?

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

Proof By Cauchy-Schwarz

$$\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle = \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i \rangle$$

$$\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\})\right)^{1/2} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle\right)^{1/2}$$

$$\leq O(\epsilon^{1/2})(\sigma_X + \sigma_{X'} + |\langle u, \mu(X) - \mu(X') \rangle|^{1/2})$$

Rearrange to get the lemma!

Algorithm from Identifiability

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

$\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{1/2})(\sigma_X + \sigma_{X'})$$

$\sigma_X^2 = \|\Sigma(X)\|$

$\sigma_{X'}^2 = \|\Sigma(X')\|$

SDP relaxation for the following quadratic program works!

Input $\{y_1, y_2, \ldots, y_n\}$ ϵ-***corrupted*** sample.

Variables/Constraints

$X' = \{x'_1, x'_2, \ldots, x'_n\}$ a guess for original sample. A coupling w.

$$w_i^2 = w_i \quad w_i(y_i - x'_i) = 0 \ \ \forall i \qquad \sum_i w_i = (1-\epsilon)n$$

Minimize $\|\Sigma(X')\|$

Identifiability for Mean Estimation

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

1) $\mathcal{U}_X$ and $\mathcal{U}_{X'}$ have 1-bounded 4th moments, and

2) $\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{3/4})(\sigma_X + \sigma_{X'})$$

$\sigma_X^2 = \|\Sigma(X)\|$
$\sigma_{X'}^2 = \|\Sigma(X')\|$

Proof $\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle \leq \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i \rangle$

Holder $\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}^{4/3})\right)^{3/4} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle^4\right)^{1/4}$

Identifiability for Mean Estimation

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

1) $\mathcal{U}_X$ and $\mathcal{U}_{X'}$ have 1-bounded 4th moments, and

2) $\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{3/4})(\sigma_X + \sigma_{X'})$$

Proof $$\frac{1}{n}\sum_i \langle u, x_i - x'_i\rangle \leq \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i\rangle$$

Holder $$\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}^{4/3})\right)^{3/4} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i\rangle^4\right)^{1/4}$$

$$\leq O(\epsilon^{3/4})\Bigg(\left(\mathbb{E}_i\langle u, x_i - \mu(X)\rangle^4\right)^{1/4} + \left(\mathbb{E}_i\langle u, x'_i - \mu(X')\rangle^4\right)^{1/4}$$
$$+ \left(\langle u, \mu(X) - \mu(X')\rangle^4\right)^{1/4}\Bigg)$$

Identifiability for Mean Estimation

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

1) $\mathcal{U}_X$ and $\mathcal{U}_{X'}$ have 1-bounded 4th moments, and

2) $\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{3/4})(\sigma_X + \sigma_{X'})$$

Proof $\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle \leq \frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}) \cdot \langle u, x_i - x'_i \rangle$

Holder $\leq \left(\frac{1}{n}\sum_i \mathbb{1}(\{x_i \neq x'_i\}^{4/3})\right)^{3/4} \cdot \left(\frac{1}{n}\sum_i \langle u, x_i - x'_i \rangle^4\right)^{1/4}$

certified bounded moment property $\leq O(\epsilon^{3/4})\,(\sigma_X + \sigma_{X'} + |\langle u, \mu(X) - \mu(X') \rangle|)$

Rearrange!

Identifiability for Mean Estimation

Lemma (Identifiability)

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $X' = \{x'_1, x'_2, \ldots, x'_n\}$ be such that:

1) $\mathcal{U}_X$ and $\mathcal{U}_{X'}$ have 1-bounded 4th moments, and

2) $\Pr_{i \in [n]} \{x_i \neq x'_i\} = \epsilon < 0.9$. Then,

$$\sigma_X^2 = \|\Sigma(X)\|$$
$$\sigma_{X'}^2 = \|\Sigma(X')\|$$

$$\|\mu(X) - \mu(X')\| < O(\epsilon^{3/4})(\sigma_X + \sigma_{X'})$$

Again yields a simple SDP* relaxation as before!

*some care to have a constraint for "bounded moment property"