SoS Certifiability of Subgaussian Distributions and its Algorithmic Applications

Ankit Pensia

Simons Institute (UC Berkeley) $\longrightarrow$ CMU Statistics

2025 TTIC Summer Workshop Program on
Information-Computation Tradeoffs for Statistical Problems

Joint work with

Ilias Diakonikolas

Sam Hopkins

Stefan Tiegel

Outline

Motivation: Distributional assumptions and accuracy

Generic Estimation Problem. Let $\mathcal{P}$ be a family of distributions over $\mathbb{R}^d$ and $\theta^* : \mathcal{P} \to \Theta$ be the target parameter.

Input: samples from (unknown) $Q \in \mathcal{P}$

Output: $\widehat{\theta}$ such that distance$\left(\widehat{\theta}, \theta^*(Q)\right)$ is small w.h.p.

Motivation: Distributional assumptions and accuracy

Generic Estimation Problem. Let $\mathcal{P}$ be a family of distributions over $\mathbb{R}^d$ and $\theta^* : \mathcal{P} \to \Theta$ be the target parameter.

Input:	samples from (unknown) $Q \in \mathcal{P}$
Output:	$\widehat{\theta}$ such that distance$\left(\widehat{\theta}, \theta^*(Q)\right)$ is small w.h.p.

- Naturally, as $\mathcal{P}$ gets **nicer**, optimal error gets **smaller**
  - Want to impose minimal assumptions on $\mathcal{P}$

Motivation: Distributional assumptions and accuracy

Generic Estimation Problem. Let $\mathcal{P}$ be a family of distributions over $\mathbb{R}^d$ and $\theta^* : \mathcal{P} \to \Theta$ be the target parameter.

Input: samples from (unknown) $Q \in \mathcal{P}$

Output: $\widehat{\theta}$ such that distance$\left(\widehat{\theta}, \theta^*(Q)\right)$ is small w.h.p.

- Naturally, as $\mathcal{P}$ gets **nicer**, optimal error gets **smaller**
  - Want to impose minimal assumptions on $\mathcal{P}$
- Unfortunately, current algorithmic guarantees often **do not improve**

Motivation: Distributional assumptions and accuracy

Generic Estimation Problem. Let $\mathcal{P}$ be a family of distributions over $\mathbb{R}^d$ and $\theta^* : \mathcal{P} \to \Theta$ be the target parameter.

Input: samples from (unknown) $Q \in \mathcal{P}$

Output: $\widehat{\theta}$ such that distance$\left(\widehat{\theta}, \theta^*(Q)\right)$ is small w.h.p.

- Naturally, as $\mathcal{P}$ gets **nicer**, optimal error gets **smaller**
  - Want to impose minimal assumptions on $\mathcal{P}$
- Unfortunately, current algorithmic guarantees often **do not improve**
  - Unless, the **niceness** assumption on $\mathcal{P}$ is **algorithmically-tractable**

Motivation: Distributional assumptions and accuracy

Generic Estimation Problem. Let $\mathcal{P}$ be a family of distributions over $\mathbb{R}^d$ and $\theta^* : \mathcal{P} \to \Theta$ be the target parameter.

Input: samples from (unknown) $Q \in \mathcal{P}$

Output: $\widehat{\theta}$ such that distance$\left(\widehat{\theta}, \theta^*(Q)\right)$ is small w.h.p.

- Naturally, as $\mathcal{P}$ gets **nicer**, optimal error gets **smaller**
  - Want to impose minimal assumptions on $\mathcal{P}$
- Unfortunately, current algorithmic guarantees often do not improve
  - Unless, the **niceness** assumption on $\mathcal{P}$ is **algorithmically-tractable**

Understanding "niceness" versus "tractable niceness"

Motivation: Tail decay and moment bounds

- Often, the optimal error is governed by **tail bounds** of **linear forms**:

  For all vectors $v \in \mathbb{R}^d$: $\quad \mathbb{P}\left(\langle v, X - \mu\rangle > y\|v\|_2\right) \ll \text{small}(y)$

Motivation: Tail decay and moment bounds

- Often, the optimal error is governed by **tail bounds** of **linear forms**:

  For all vectors $v \in \mathbb{R}^d$: $\quad \mathbb{P}\left(\langle v, X - \mu\rangle > y\|v\|_2\right) \ll \text{small}(y)$

- Equivalently, on the **moments** of **linear forms**

  For all vectors $v \in \mathbb{R}^d$: $\quad \mathbb{E}\left[|\langle v, X - \mu\rangle|^t\right] \leqslant \text{Bound}(t)\|v\|^t$

Motivation: Tail decay and moment bounds

- Often, the optimal error is governed by tail bounds of linear forms:

  For all vectors $v \in \mathbb{R}^d$: $\quad \mathbb{P}\left(\langle v, X-\mu\rangle > y\|v\|_2\right) \ll \text{small}(y)$

- Equivalently, on the moments of linear forms

  For all vectors $v \in \mathbb{R}^d$: $\quad \mathbb{E}\left[|\langle v, X-\mu\rangle|^t\right] \leqslant \text{Bound}(t)\|v\|^t$

- This motivates the definition of **subgaussian distributions**

Motivation: Subgaussianity versus gaussianity

Definition. X is **Subgaussian** if its moments grow slower than **Gaussian**.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu \rangle^t\right] \leqslant \mathbb{E}_{G \sim \mathcal{N}(0, I_d)}[\langle v, G \rangle^t]$

Motivation: Subgaussianity versus gaussianity

Definition. X is **Subgaussian** if its moments grow slower than **Gaussian**.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

Motivation: Subgaussianity versus gaussianity

Definition. X is **Subgaussian** if its moments grow slower than **Gaussian**.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu\rangle^t\right] \leqslant \mathbb{E}_{G \sim \mathcal{N}(0, I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X - \mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X-\mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X-\mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X - \mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption **while** keeping similar **information-theoretic** optimal error

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X-\mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X-\mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption while keeping similar **information-theoretic** optimal error
- However, current **computationally-efficient** algorithms seem unable to use subgaussianity

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X - \mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption while keeping similar **information-theoretic** optimal error
- However, current **computationally-efficient** algorithms seem unable to use subgaussianity
  - Even though they are able to use gaussianity

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X-\mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X-\mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption while keeping similar **information-theoretic** optimal error
- However, current **computationally-efficient** algorithms seem unable to use subgaussianity
  - Even though they are able to use gaussianity
  - In fact, their error scales as if only **the second moment** $(t=2)$ is bounded

Motivation: Subgaussianity versus gaussianity

Definition. X is Subgaussian if its moments grow slower than Gaussian.

For all $v \in \mathbb{R}^d$ and even t: $\mathbb{E}\left[\langle v, X - \mu\rangle^t\right] \leqslant \mathbb{E}_{G\sim\mathcal{N}(0,I_d)}[\langle v, G\rangle^t]$

- Equivalently, $\mathbb{P}\left(\langle v, X - \mu\rangle \gg y\|v\|\right) \leqslant \exp(-y^2)$
- Thus, we relax the *unrealistic* gaussianity assumption while keeping similar **information-theoretic** optimal error
- However, current **computationally-efficient** algorithms seem unable to use subgaussianity
  - Even though they are able to use gaussianity
  - In fact, their error scales as if only **the second moment** $(t = 2)$ is bounded

Algorithmically-tractable version of subgaussianity?

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input: outlier-corrupted samples from $\mathcal{Q}$

a small fraction of samples are arbitrary

Output: $\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}$ is small w.h.p.

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input:	outlier-corrupted samples from $\mathcal{Q}$
Output:	$\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}$ is small w.h.p.

- Information theoretic error: dimension-independent

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input:	outlier-corrupted samples from $\mathcal{Q}$
Output:	$\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}$ is small w.h.p.

- Information theoretic error: dimension-independent
- Algorithmically: not well-understood

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input: outlier-corrupted samples from $\mathcal{Q}$

Output: $\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{\text{op}}$ is small w.h.p.

- Information theoretic error: dimension-independent
- Algorithmically: not well-understood
  - $\text{Error}_{\text{Efficient-algorithms}}(\text{sub-Gaussian}) \gg \text{Error}_{\text{Efficient-algorithms}}(\text{Gaussian})$

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input: outlier-corrupted samples from $\mathcal{Q}$

Output: $\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{op}$ is small w.h.p.

- Information theoretic error: dimension-independent
- Algorithmically: not well-understood
  - $\text{Error}_{\text{Efficient-algorithms}}(\text{sub-Gaussian}) \gg \text{Error}_{\text{Efficient-algorithms}}(\text{Gaussian})$
  - Also true for (robust) regression, mean estimation, clustering, ...

Example: Robust covariance estimation of sub-Gaussians

Let $\mathcal{Q}$ be a **sub-Gaussian** distribution over $\mathbb{R}^d$ with covariance Σ

Input: outlier-corrupted samples from $\mathcal{Q}$

Output: $\widehat{\Sigma}$ such that $\|\widehat{\Sigma} - \Sigma\|_{\text{op}}$ is small w.h.p.

- Information theoretic error: dimension-independent
- Algorithmically: not well-understood
  - $\text{Error}_{\text{Efficient-algorithms}}(\text{sub-Gaussian}) \gg \text{Error}_{\text{Efficient-algorithms}}(\text{Gaussian})$
  - Also true for (robust) regression, mean estimation, clustering, ...

What are the underlying algorithmic challenges?

Challenges in robust covariance estimation

Algorithmic template: robust **mean** estimation

1. While there exists a direction v with large **variance**:
   1.1 Filter a point x if $\langle v, x \rangle$ is too large
2. **return** sample mean

Challenges in robust covariance estimation

Algorithmic template: robust **covariance** estimation

1. While there exists a direction v with large **t-th moment** (for $t \geqslant 4$):
   1.1 Filter a point x if $\langle v, x \rangle^2$ is too large
2. **return** sample covariance

Challenges in robust covariance estimation

Algorithmic template: robust **covariance** estimation

1. While there exists a direction v with large **t-th moment** (for $t \geqslant 4$):
   1.1 Filter a point x if $\langle v, x\rangle^2$ is too large
2. **return** sample covariance

- Need to find a direction $\widehat{v} \approx \operatorname{argmax}_{v \in \mathcal{S}^{d-1}} \mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$)

Challenges in robust covariance estimation

Algorithmic template: robust **covariance** estimation

1. While there exists a direction v with large **t-th moment** (for $t \geqslant 4$):
   1.1 Filter a point x if $\langle v, x \rangle^2$ is too large
2. **return** sample covariance

- Need to find a direction $\widehat{v} \approx \operatorname{argmax}_{v \in \mathcal{S}^{d-1}} \mathbb{E}_X[\langle v, X \rangle^t]$ (for $t \geqslant 4$)
- Unfortunately, intractable (in the worst-case over X)

Challenges in robust covariance estimation

Algorithmic template: robust **covariance** estimation

1. While there exists a direction v with large **t-th moment** (for $t \geqslant 4$):
   1.1 Filter a point x if $\langle v, x\rangle^2$ is too large
2. **return** sample covariance

- Need to find a direction $\hat{v} \approx \operatorname{argmax}_{v \in \mathcal{S}^{d-1}} \mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$)
- Unfortunately, intractable (in the worst-case over X)
- **Further structure?**

Challenges in robust covariance estimation

Algorithmic template: robust **covariance** estimation

1. While there exists a direction v with large **t-th moment** (for $t \geqslant 4$):
   1.1 Filter a point x if $\langle v, x\rangle^2$ is too large
2. **return** sample covariance

- Need to find a direction $\widehat{v} \approx \operatorname{argmax}_{v \in \mathcal{S}^{d-1}} \mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$)
- Unfortunately, intractable (in the worst-case over X)
- **Further structure?**

Do sub-Gaussian data (X) have algorithm-friendly structure?

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X \left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

- A central object in many applications

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

- A central object in many applications
  - Injective tensor norm of the t-tensor $\mathbb{E}_X[X^{\otimes t}]$

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

- A central object in many applications
  - Injective tensor norm of the t-tensor $\mathbb{E}_X[X^{\otimes t}]$
  - The 2-to-t norm of the $n \times d$ matrix $\mathbf{A}$,
    - where $\mathbf{A}$ has rows $a_1, \ldots, a_n$ and X is uniform on $(a_1, \ldots, a_n)$

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

- A central object in many applications
  - Injective tensor norm of the t-tensor $\mathbb{E}_X[X^{\otimes t}]$
  - The 2-to-t norm of the $n \times d$ matrix $\mathbf{A}$,
    - where $\mathbf{A}$ has rows $a_1, \ldots, a_n$ and X is uniform on $(a_1, \ldots, a_n)$

- **Challenge.** No efficient algorithm for constant approximation (for $t \geqslant 4$)
  - Under Exponential Time Hypothesis [BBHKSZ12]
  - Stronger hardness under Small Set Expansion Hypothesis [BBHKSZ12; HL19]

[BBHKSZ12] B. Barak, F. G.S.L. Brandao, A. W. Harrow, J. Kelner, D. Steurer, Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications.

[HL19] S. B. Hopkins, J. Li. How Hard Is Robust Mean Estimation? *COLT*. 2019

Challenge: Checking the moments efficiently

- The t-th moment bound of the distribution X is

$$\max_{v} \frac{\mathbb{E}_X\left[\langle v, X\rangle^t\right]}{\|v\|^t}$$

  - Subgaussianity $\iff$ bounded appropriately for all even t

- A central object in many applications
  - Injective tensor norm of the t-tensor $\mathbb{E}_X[X^{\otimes t}]$
  - The 2-to-t norm of the $n \times d$ matrix $\mathbf{A}$,
    - where $\mathbf{A}$ has rows $a_1, \ldots, a_n$ and X is uniform on $(a_1, \ldots, a_n)$

- **Challenge.** No efficient algorithm for constant approximation (for $t \geqslant 4$)
  - Under Exponential Time Hypothesis [BBHKSZ12]
  - Stronger hardness under Small Set Expansion Hypothesis [BBHKSZ12; HL19]

Algorithmically-tractable version of subgaussianity?

[BBHKSZ12] B. Barak, F. G.S.L. Brandao, A. W. Harrow, J. Kelner, D. Steurer, Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications.

[HL19] S. B. Hopkins, J. Li. How Hard Is Robust Mean Estimation? *COLT*. 2019

Algorithm-friendly subsets of sub-Gaussians

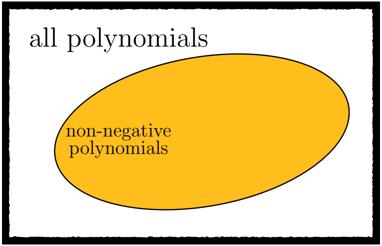
Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Given an arbitrary degree-t (for $t \geqslant 4$) polynomial $h(\cdot)$ on $\mathbb{R}^d$,

- It is **hard** to check if it is always $\geqslant 0$

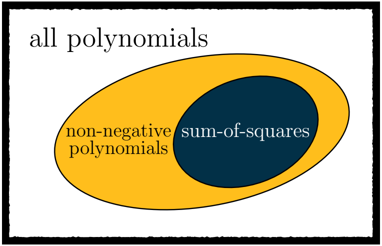


Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Given an arbitrary degree-t (for $t \geqslant 4$) polynomial $h(\cdot)$ on $\mathbb{R}^d$,

- It is **hard** to check if it is always $\geqslant 0$
- But **easy** to check if $h(v) = \sum_i q_i^2(v)$, i.e., sum-of-squares, & thus $\geqslant 0$



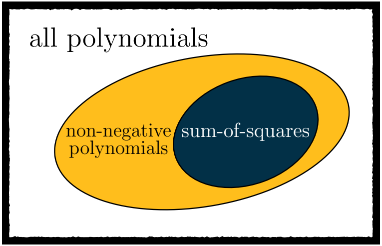
Algorithm-friendly subsets of sub-Gaussians

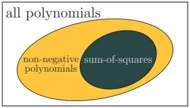
Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Given an arbitrary degree-t (for $t \geqslant 4$) polynomial $h(\cdot)$ on $\mathbb{R}^d$,

- It is **hard** to check if it is always $\geqslant 0$
- But **easy** to check if $h(v) = \sum_i q_i^2(v)$, i.e., sum-of-squares, & thus $\geqslant 0$

a semidefinite program





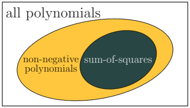
Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Definition. (Algorithm-friendly sub-Gaussians) X is **certifiably-sub-Gaussian** if

not only are the moments bounded

$$\mathbb{E}_{G\sim\mathcal{N}(0,\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X\rangle^t\right] \geqslant 0$$



Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Definition. (Algorithm-friendly sub-Gaussians) X is **certifiably-sub-Gaussian** if

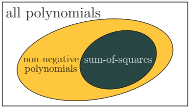
the moments are bounded in a sum-of-squares way: for all even t,

$$\mathbb{E}_{G\sim\mathcal{N}(0,\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X\rangle^t\right] = \sum_i q_i^2(v)$$

not only $\geqslant 0$,
but also a SoS polynomials

for some polynomials $q_1(\cdot), \ldots,$

sub-Gaussian: moments less than Gaussians



Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Definition. (Algorithm-friendly sub-Gaussians) X is **certifiably-sub-Gaussian** if

the moments are bounded in a sum-of-squares way: for all even t,

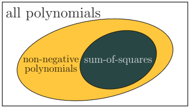
$$\mathbb{E}_{G\sim\mathcal{N}(0,\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X\rangle^t\right] = \sum_i q_i^2(v)$$

for some polynomials $q_1(\cdot), \ldots,$

Proposed in [KSS18; HL18] and hugely influential since then

[KSS18] P. K. Kothari, J. Steinhardt, D. Steurer. Robust Moment Estimation and Improved Clustering via Sum of Squares. *STOC*. 2018

[HL18] S. B. Hopkins, J. Li. Mixture Models, Robustness, and Sum of Squares Proofs. *STOC*. 2018



Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Definition. (Algorithm-friendly sub-Gaussians) X is **certifiably-sub-Gaussian** if

the moments are bounded in a sum-of-squares way: for all even t,

$$\mathbb{E}_{G\sim\mathcal{N}(0,\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X\rangle^t\right] \quad = \sum_i q_i^2(v)$$

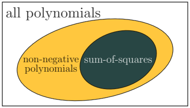
for some polynomials $q_1(\cdot), \ldots,$

Proposed in [KSS18; HL18] and hugely influential since then

- For this class of distributions, we get polynomial-time algorithms!

[KSS18] P. K. Kothari, J. Steinhardt, D. Steurer. Robust Moment Estimation and Improved Clustering via Sum of Squares. *STOC*. 2018

[HL18] S. B. Hopkins, J. Li. Mixture Models, Robustness, and Sum of Squares Proofs. *STOC*. 2018



Algorithm-friendly subsets of sub-Gaussians

Challenge: checking if the moment polynomial $\mathbb{E}_X[\langle v, X\rangle^t]$ (for $t \geqslant 4$) is bounded

Definition. (Algorithm-friendly sub-Gaussians) X is **certifiably-sub-Gaussian** if

the moments are bounded in a sum-of-squares way: for all even t,

$$\mathbb{E}_{G\sim\mathcal{N}(0,\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X\rangle^t\right] = \sum_i q_i^2(v)$$

for some polynomials $q_1(\cdot), \ldots,$

Proposed in [KSS18; HL18] and hugely influential since then

- For this class of distributions, we get polynomial-time algorithms!

Which distributions are certifiably sub-Gaussian anyway?

Outline

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)

[KSS18] P. K. Kothari, J. Steinhardt, D. Steurer. Robust Moment Estimation and Improved Clustering via Sum of Squares. *STOC*. 2018

Prior work: **Certifiable sub-Gaussianity vs. sub-Gaussianity**

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top \mathbf{J} x) \lesssim \|\mathbf{J}\|_{\mathrm{Fr}}^2$

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top \mathbf{J} x) \lesssim \|\mathbf{J}\|_{\mathrm{Fr}}^2$ (SoS in variable $\mathbf{J}$)

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top \mathbf{J} x) \lesssim \|\mathbf{J}\|_{\mathrm{Fr}}^2$ (SoS in variable $\mathbf{J}$)
  - Which is not true for many subgaussians

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger*[HL19]

[HL19] S. B. Hopkins, J. Li. How Hard Is Robust Mean Estimation? *COLT*. 2019

Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger*[HL19]
  - There exists X with bounded moments $\mathbb{E}_X[\langle v, X\rangle^{100}] \leqslant O(1)\|v\|_2^{100}$ for all v

[HL19] S. B. Hopkins, J. Li. How Hard Is Robust Mean Estimation? *COLT*. 2019

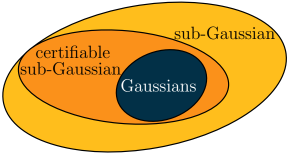
Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger*[HL19]
  - There exists X with bounded moments $\mathbb{E}_X[\langle v, X\rangle^{100}] \leqslant O(1)\|v\|_2^{100}$ for all v
  - but X does NOT have certifiably bounded moments $C\|v\|_2^4 - \mathbb{E}_X[\langle v, X\rangle^4]$ is not sum-of-squares for any constant C

[HL19] S. B. Hopkins, J. Li. How Hard Is Robust Mean Estimation? *COLT*. 2019

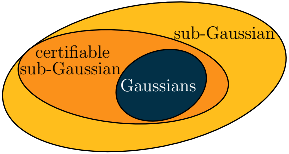
Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

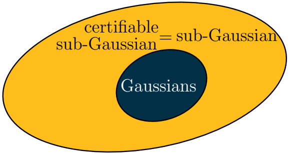
- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger* [HL19]
  - There exists X with bounded moments
  - but X does NOT have certifiably bounded moments
- Led to the belief that certifiable sub-Gaussianity is also strictly stronger than sub-Gaussianity



Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

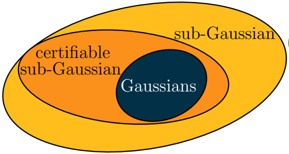
- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger*[HL19]
  - There exists X with bounded moments
  - but X does NOT have certifiably bounded moments
- Led to the belief that certifiable sub-Gaussianity is also strictly stronger than sub-Gaussianity



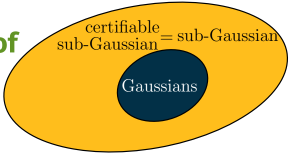


Prior work: Certifiable sub-Gaussianity vs. sub-Gaussianity

- **Sufficient** conditions for certifiability are rather strong
  - rotational invariance, independent coordinates
  - log-Sobolev distributions (more generally, Poincare) [KSS18]
  - and simple transformations thereof (simple within SoS)
- Many sub-Gaussians **do not** satisfy these sufficient conditions
  - Existing proofs crucially need $\mathrm{Var}(x^\top J x) \lesssim \|J\|_{Fr}^2$ (SoS in variable J)
  - Which is not true for many subgaussians
- In general, certifiability of (constantly-many) moments is strictly stronger*[HL19]
  - There exists X with bounded moments
  - but X does NOT have certifiably bounded moments
- Led to the belief that certifiable sub-Gaussianity is also strictly stronger than sub-Gaussianity



Can we characterize the class of certifiably sub-Gaussians?



Outline

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-**P**-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

,

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-**P**-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G\sim\mathcal{N}(0,C\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X-\mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1,\dots,.$$

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-P-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G \sim \mathcal{N}(0, C\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X - \mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1, \dots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-P-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G \sim \mathcal{N}(0, C\mathbf{I})}\left[\langle v, G \rangle^t\right] - \mathbb{E}_X\left[\langle v, X - \mu \rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1, \ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-**P**-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G\sim\mathcal{N}(0,C\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X-\mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1,\ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$
- We show sub-Gaussians also have other algorithm-friendly structures

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: **Certifiability of sub-Gaussian distributions**

Theorem: [Diakonikolas-Hopkins-P-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G\sim\mathcal{N}(0,C\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X-\mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1,\ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$
- We show sub-Gaussians also have other algorithm-friendly structures
- **Conceptual contribution**: connecting SDP relaxations & empirical processes

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-P-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G\sim\mathcal{N}(0,C\mathbf{I})}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X-\mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1, \ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$
- We show sub-Gaussians also have other algorithm-friendly structures
- Conceptual contribution: connecting SDP relaxations & empirical processes
- Corollary: **new polynomial-time algorithms** for sub-Gaussian data

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Our result: **Certifiability of sub-Gaussian distributions**

Theorem: [Diakonikolas-Hopkins-**P**-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G\sim\mathcal{N}(0,CI)}\left[\langle v, G\rangle^t\right] - \mathbb{E}_X\left[\langle v, X-\mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1, \ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$
- We show sub-Gaussians also have other algorithm-friendly structures
- Conceptual contribution: connecting SDP relaxations & empirical processes
- Corollary: new polynomial-time algorithms for sub-Gaussian data
- Corollary: **new computational lower bounds** for Gaussian data (certification)

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

New algorithmic guarantees for sub-Gaussian data

New algorithmic guarantees for sub-Gaussian data

- We give the **first polynomial-time algorithms** for
  - robust covariance estimation
  - robust linear regression
  - robust covariance-aware mean estimation

New algorithmic guarantees for sub-Gaussian data

- We give the first polynomial-time algorithms for
  - robust covariance estimation
  - robust linear regression
  - robust covariance-aware mean estimation

In contrast, prior algorithmic guarantees were vacuous

New algorithmic guarantees for sub-Gaussian data

- We give the first polynomial-time algorithms for
  - robust covariance estimation
  - robust linear regression
  - robust covariance-aware mean estimation

  In contrast, prior algorithmic guarantees were vacuous

- We also **vastly improve** the error of polynomial-time algorithms for
  - robust mean estimation
  - clustering and mixture models

New algorithmic guarantees for sub-Gaussian data

- We give the first polynomial-time algorithms for
  - robust covariance estimation
  - robust linear regression
  - robust covariance-aware mean estimation

  In contrast, prior algorithmic guarantees were vacuous
- We also vastly improve the error of polynomial-time algorithms for
  - robust mean estimation
  - clustering and mixture models

  In contrast, prior algorithmic guarantees were worse (sub-Gaussian vs. second moment)

New algorithmic guarantees for sub-Gaussian data

- We give the first polynomial-time algorithms for
  - robust covariance estimation
  - robust linear regression
  - robust covariance-aware mean estimation

  In contrast, prior algorithmic guarantees were vacuous

- We also vastly improve the error of polynomial-time algorithms for
  - robust mean estimation
  - clustering and mixture models

  In contrast, prior algorithmic guarantees were worse (sub-Gaussian vs. second moment)

- Our algorithmic guarantees are qualitatively-**optimal*** (within low-degree polynomial tests, statistical query algorithms, sum-of-squares hierarchy)

New algorithmic guarantees for sub-Gaussian data

Robust Estimation Task	Inlier Distribution	Information-theoretic Error	Previous Best Guarantee in Polynomial Time	New Guarantees
Covariance estimation: Relative spectral norm	hypercontractive sub-Gaussian	$\widetilde{\Theta}(\epsilon)$	No general algorithm	$\epsilon^{1-2/m}$
Mean estimation: Mahalanobis norm	hypercontractive sub-Gaussian	$\widetilde{\Theta}(\epsilon)$	No general algorithm	$\epsilon^{1-1/m}$
Linear regression: Arbitrary noise	hypercontractive sub-Gaussian	$\widetilde{\Theta}(\epsilon)$	No general algorithm	$\epsilon^{1-\frac{2}{m}}$
Mean estimation: Euclidean norm	sub-Gaussian	$\widetilde{\Theta}(\epsilon)$	$\sqrt{\epsilon}$	$\epsilon^{1-1/m}$
List-decodable mean estimation	sub-Gaussian	$\widetilde{\Theta}(\epsilon)$	$\sqrt{\frac{1}{1-\epsilon}}$	$\left(\frac{1}{1-\epsilon}\right)^{-\Omega\left(\frac{1}{m}\right)}$
Mixture models: Mixture of k Δ-separated components	Each component is sub-Gaussian	$\Delta \gtrsim \sqrt{\log k}$	$\Delta \gtrsim k^{\Omega(1)}$	$\Delta \gtrsim k^{O\left(\frac{1}{m}\right)}$

Outline

- Motivation
- Prior Work
- Our Result
- **Proof Sketch**
- Conclusion

Proof sketch

A series of equivalences using

1. duality
2. empirical processes
3. linearization

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$
$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

Certifiable formulation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

Certifiable formulation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

Certifiable formulation

X is sub-Gaussian

$$\Updownarrow$$

$$\forall v: \quad q(v) \leqslant B\|v\|^t$$

$$q(v) := \mathbb{E}_X[\langle v, X\rangle^t]$$

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$$\Updownarrow$$

$$\forall v: \quad q(v) \leqslant B\|v\|^t$$

$$\Updownarrow$$

$$\max_{\mathbb{E}\text{:over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$$\Updownarrow$$

$$\forall v:\ \ q(v) \leqslant B\|v\|^t$$

$$\Updownarrow$$

$$\max_{\mathbb{E}\text{ :over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Depends only on the degree-t moments of $\mathbb{E}$

Certifiable formulation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v: \;\; q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E} \text{ :over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v: \quad q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E}\text{:over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

X is certifiably sub-Gaussian

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v: \; q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E} \text{ :over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

X is certifiably sub-Gaussian

$\Updownarrow$

$q(v) \leqslant_{\text{sos}} B\|v\|^t$

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v: \quad q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E}\text{:over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

X is certifiably sub-Gaussian

$\Updownarrow$

$q(v) \leqslant_{\text{sos}} B\|v\|^t$

$\Updownarrow$

$$\max_{\widetilde{\mathbb{E}}\text{:degree-t}} \frac{\widetilde{\mathbb{E}}_v[q(v)]}{\widetilde{\mathbb{E}}_v[\|v\|^t]} \leqslant B$$

A "pseudo"-expectation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v: \;\; q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E} \text{ :over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Set of degree-t moments over $\mathbb{R}^p$

Certifiable formulation

X is certifiably sub-Gaussian

$\Updownarrow$

$q(v) \leqslant_{\text{sos}} B\|v\|^t$

$\Updownarrow$

$$\max_{\widetilde{\mathbb{E}} \text{ :degree-}t} \frac{\widetilde{\mathbb{E}}_v[q(v)]}{\widetilde{\mathbb{E}}_v[\|v\|^t]} \leqslant B$$

A "pseudo"-expectation

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$

$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

Original formulation

X is sub-Gaussian

$\Updownarrow$

$\forall v:\ q(v) \leqslant B\|v\|^t$

$\Updownarrow$

$$\max_{\mathbb{E}\text{ :over } \mathbb{R}^d} \frac{\mathbb{E}_v[q(v)]}{\mathbb{E}_v[\|v\|^t]} \leqslant B$$

Certifiable formulation

X is certifiably sub-Gaussian

$\Updownarrow$

$q(v) \leqslant_{\text{sos}} B\|v\|^t$

$\Updownarrow$

$$\max_{\widetilde{\mathbb{E}}\text{ :degree-t}} \frac{\widetilde{\mathbb{E}}_v[q(v)]}{\widetilde{\mathbb{E}}_v[\|v\|^t]} \leqslant B$$

A "pseudo"-expectation

Set of degree-t $\widetilde{\mathbb{E}}$

$q(v) := \mathbb{E}[\langle v, X\rangle^t]$
$B := (\Theta(\sqrt{t}))^t$

Proof sketch: (1/3) Duality

$$\max_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}_v[q(v)]}{\widetilde{\mathbb{E}}_v[\|v\|^t]} \leqslant B$$

A "pseudo"-expectation

Equivalent to showing that the dual is bounded

certifiabile sub-Gaussianity $\iff$ dual is bounded

Proof sketch: (2/3) Empirical process

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}\left[\mathbb{E}[\langle v, X\rangle^t]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]}$$

certifiabile sub-Gaussianity $\iff$ dual is bounded

Proof sketch: (2/3) Empirical process

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}\left[\mathbb{E}[\langle v, X\rangle^t]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]} = \sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\langle v, X_i\rangle^t]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]}$$

certifiabile sub-Gaussianity $\Longleftrightarrow$ dual is bounded

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

Proof sketch: (2/3) Empirical process

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}\left[\mathbb{E}\left[\langle v, X\rangle^t\right]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]} = \sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \widetilde{\mathbb{E}}\left[\langle v, X_i\rangle^t\right]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]}$$

certifiabile sub-Gaussianity $\iff$ dual is bounded

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

Proof sketch: (2/3) Empirical process

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-t}} \frac{\widetilde{\mathbb{E}}\left[\mathbb{E}[\langle v, X\rangle^t]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]} = \sup_{\widetilde{\mathbb{E}}\text{:degree-t}} \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]}$$

$$\leqslant \mathbb{E}_{X_1,\ldots,X_n}\left[\sup_{\widetilde{\mathbb{E}}\text{:degree-t}} \frac{1}{n}\sum_{i=1}^n \frac{\widetilde{\mathbb{E}}\left[\langle v, X_i\rangle^t\right]}{\widetilde{\mathbb{E}}\left[\|v\|^t\right]}\right]$$

certifiabile sub-Gaussianity $\iff$ dual is bounded

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

Proof sketch: (2/3) Empirical process

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}\left[\mathbb{E}[\langle v, X\rangle^t]\right]}{\widetilde{\mathbb{E}}\,[\|v\|^t]} = \sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]\right]}{\widetilde{\mathbb{E}}\,[\|v\|^t]}$$

$$\leqslant \mathbb{E}_{X_1,\dots,X_n}\left[\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{1}{n}\sum_{i=1}^n \frac{\widetilde{\mathbb{E}}\,[\langle v, X_i\rangle^t]}{\widetilde{\mathbb{E}}\,[\|v\|^t]}\right]$$

an empirical process

$$= \mathbb{E}_{X_1,\dots,X_n}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n f(X_i)\right]$$

$$\mathcal{F} := \left\{ x \mapsto \frac{\widetilde{\mathbb{E}}\,[\langle v, x\rangle^t]}{\widetilde{\mathbb{E}}\,[\|v\|^t]} : \widetilde{\mathbb{E}} \text{ degree-}t \right\}$$

certifiabile sub-Gaussianity $\iff$ dual is bounded

$\widetilde{\mathbb{E}}$: a "pseudo"-expectation

Proof sketch: (2/3) Empirical process

$$\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\widetilde{\mathbb{E}}[\mathbb{E}[\langle v, X\rangle^t]]}{\widetilde{\mathbb{E}}[\|v\|^t]} = \sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{\mathbb{E}[\frac{1}{n}\sum_{i=1}^n \widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

$$\leqslant \mathbb{E}_{X_1,\dots,X_n}\left[\sup_{\widetilde{\mathbb{E}}\text{:degree-}t} \frac{1}{n}\sum_{i=1}^n \frac{\widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}\right]$$

an empirical process

$$= \mathbb{E}_{X_1,\dots,X_n}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n f(X_i)\right]$$

$$\mathcal{F} := \left\{x \mapsto \frac{\widetilde{\mathbb{E}}[\langle v, x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]} : \widetilde{\mathbb{E}} \text{ degree-}t\right\}$$

Is this empirical process bounded for **all** sub-Gaussians?

Proof sketch: (3/3) Chaining

$$\mathop{\mathbb{E}}_{X_i' s}\left[\sup_{f\text{ :from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian **linear** process $\leqslant$ Gaussian **linear** process

Proof sketch: (3/3) Chaining

$$\mathbb{E}_{X_i' s}\left[\sup_{f:\text{from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian linear process $\leqslant$ Gaussian linear process
  - but our process is non-linear

$$\mathop{\mathbb{E}}_{X_i's}\left[\sup_{f:\text{from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Proof sketch: (3/3) Chaining

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian linear process $\leqslant$ Gaussian linear process
  - but our process is non-linear. Can we still use it?

Proof sketch: (3/3) Chaining

$$\mathbb{E}_{X_i' s}\left[\sup_{f:\text{from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian linear process $\leqslant$ Gaussian linear process
  - but our process is non-linear. Can we still use it?
- Beyond linear processes, chaining also applies if f is **"linear-ish"**
  - $\sup_{g \in \mathcal{G}} \langle g, x\rangle$ or $\left(\sup_{g \in \mathcal{G}} \langle g, x\rangle\right)^t$

$$\mathbb{E}_{X_i's}\left[\sup_{f\text{ : from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Proof sketch: (3/3) Chaining

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian linear process $\leqslant$ Gaussian linear process
  - but our process is non-linear. Can we still use it?

- Beyond linear processes, chaining also applies if f is **"linear-ish"**
  - $\sup_{g\in\mathcal{G}}\langle g, x\rangle$ or $\left(\sup_{g\in\mathcal{G}}\langle g, x\rangle\right)^t$

Lemma. $\sum_i f(x_i)$ is "linear-ish" and generic chaining is applicable

$$\mathop{\mathbb{E}}_{X_i' s}\left[\sup_{f:\text{from degree-}t\ \widetilde{\mathbb{E}}} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v,x\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Proof sketch: (3/3) Chaining

- Foundational result in probability: Talagrands's generic chaining
  - all sub-Gaussian linear process $\leqslant$ Gaussian linear process
  - but our process is non-linear. Can we still use it?

- Beyond linear processes, chaining also applies if f is **"linear-ish"**
  - $\sup_{g\in\mathcal{G}}\langle g,x\rangle$ or $\left(\sup_{g\in\mathcal{G}}\langle g,x\rangle\right)^t$

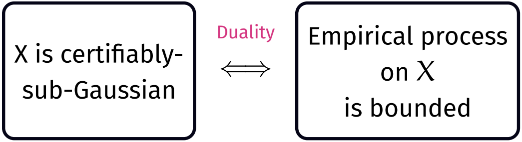
Lemma. $\sum_i f(x_i)$ is "linear-ish" and generic chaining is applicable

For all sub-Gaussians, our process is bounded by Gaussians!

Proof sketch: Putting the pieces together

X is certifiably-sub-Gaussian

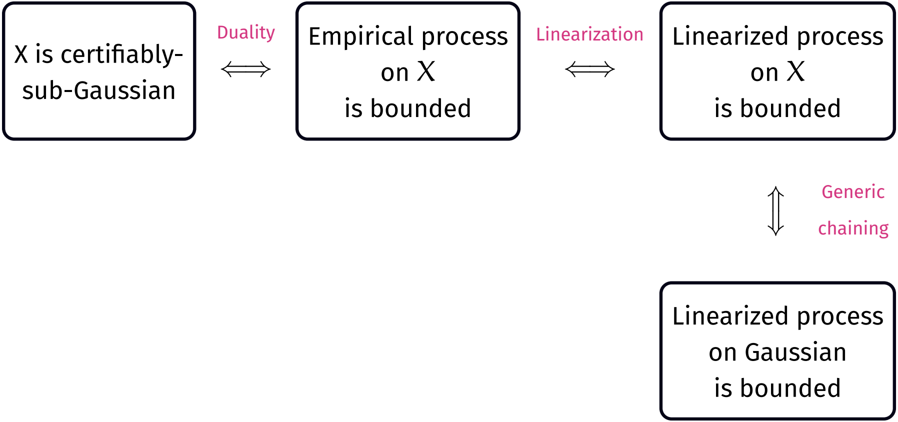
Proof sketch: Putting the pieces together



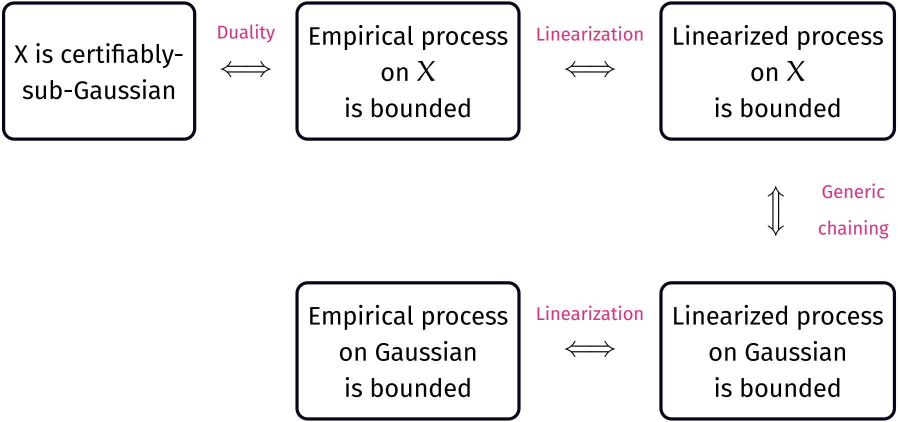
Proof sketch: Putting the pieces together

X is certifiably-sub-Gaussian $\overset{\text{Duality}}{\iff}$ Empirical process on X is bounded $\overset{\text{Linearization}}{\iff}$ Linearized process on X is bounded

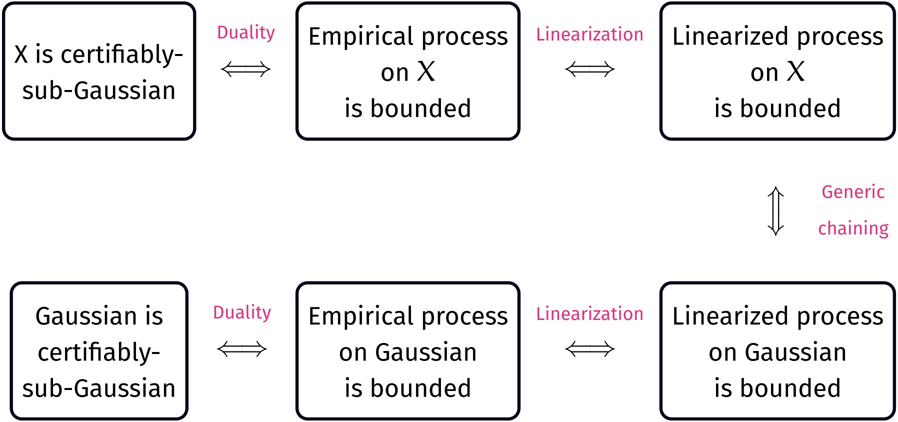
Proof sketch: Putting the pieces together



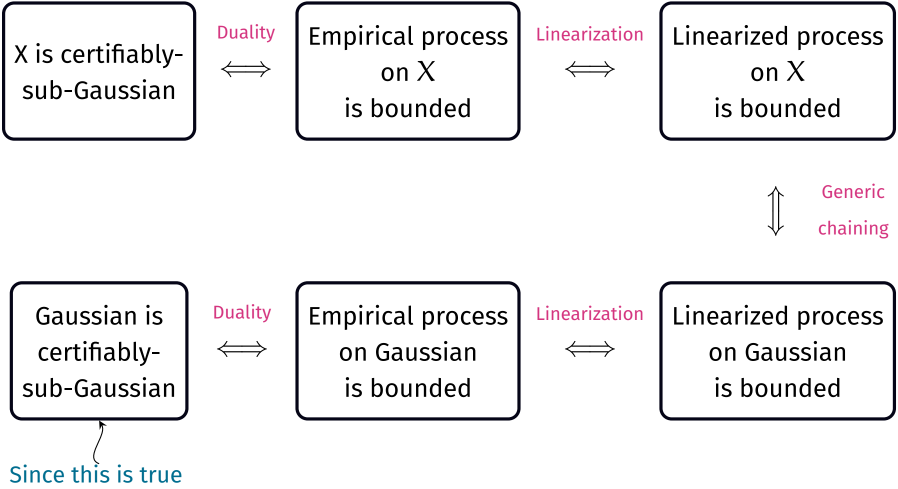
Proof sketch: Putting the pieces together



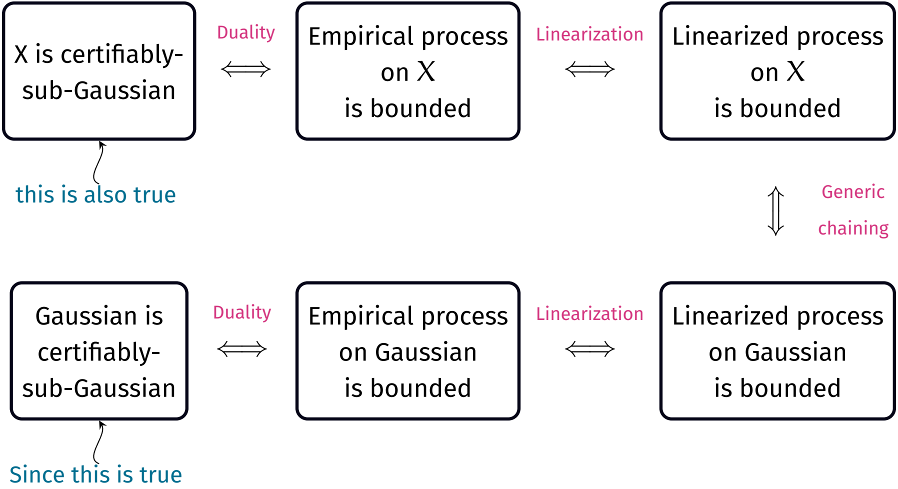
Proof sketch: Putting the pieces together



Proof sketch: Putting the pieces together



Proof sketch: Putting the pieces together



Our result: Certifiability of sub-Gaussian distributions

Theorem: [Diakonikolas-Hopkins-**P**-Tiegel]

All sub-Gaussian distributions are **certifiably** so! (upto an absolute constant C)

There exists C s.t. for any $d \in \mathbb{N}$, sub-Gaussian distribution X on $\mathbb{R}^d$, and even $t \in \mathbb{N}$,

$$\mathbb{E}_{G \sim \mathcal{N}(0, CI)} \left[\langle v, G\rangle^t\right] - \mathbb{E}_X \left[\langle v, X - \mu\rangle^t\right] = \sum_i q_i^2(v) \qquad \text{for some polynomials } q_1, \ldots,.$$

- Answers open questions of Kothari-Steinhardt (2018), Hopkins (2019) & extends Barak-Brandao-Harrow-Kelner-Steurer-Zhou (2012)
  - Even for the fourth moment ($t = 4$), the prior best bound was $\sqrt{d}$
- We show sub-Gaussians also have other algorithm-friendly structures
- Conceptual contribution: connecting SDP relaxations & empirical processes
- Corollary: new polynomial-time algorithms for sub-Gaussian data
- Corollary: **new computational lower bounds** for Gaussian data (certification)

[DHPT25] I. Diakonikolas, S. Hopkins, A. Pensia, S. Tigel. SoS Certifiability of Subgaussian Distributions & Its Algorithmic Applications. *STOC*. 2025

Outline

Conclusion

- Established that all subgaussian distributions are certifiably so

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Open problems:

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Open problems:

- Going beyond subgaussianity for certifiably bounded moments.
  - Subexponential distributions?

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Open problems:

- Going beyond subgaussianity for certifiably bounded moments.
  - Subexponential distributions?
- What is the **largest** class of distributions $\mathcal{P}$ and **smallest** B that leads to (quasi)-polynomial algorithm for deciding/refuting between:

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Open problems:

- Going beyond subgaussianity for certifiably bounded moments.
  - Subexponential distributions?
- What is the **largest** class of distributions $\mathcal{P}$ and **smallest** B that leads to (quasi)-polynomial algorithm for deciding/refuting between:

Let $B \gg 1$ and $\mathcal{P} \subset \left\{P : \sup_{v \in \mathcal{S}^{d-1}} \mathbb{E}_{X \sim P}[\langle v, X \rangle^4] \leqslant 1\right\}$.

Null: $X_1, \ldots, X_n$ are iid from $P \in \mathcal{P}$

Alternate: $X_1, \ldots, X_n$ are iid from Q with $\sup_{v \in \mathcal{S}^{d-1}} \mathbb{E}_{X \sim Q}[\langle v, X \rangle^4] \geqslant B$

Conclusion

- Established that all subgaussian distributions are certifiably so
- As a corollary, we get a host of new algorithmic guarantees

Open problems:

- Going beyond subgaussianity for certifiably bounded moments.
  - Subexponential distributions?
- What is the **largest** class of distributions $\mathcal{P}$ and **smallest** B that leads to (quasi)-polynomial algorithm for deciding/refuting between:

Let $B \gg 1$ and $\mathcal{P} \subset \left\{P : \sup_{v \in \mathcal{S}^{d-1}} \mathbb{E}_{X\sim P}[\langle v, X\rangle^4] \leqslant 1\right\}$.

Null: $X_1, \ldots, X_n$ are iid from $P \in \mathcal{P}$

Alternate: $X_1, \ldots, X_n$ are iid from Q with $\sup_{v \in \mathcal{S}^{d-1}} \mathbb{E}_{X\sim Q}[\langle v, X\rangle^4] \geqslant B$

Thank you for your attention!

Outline

- **Proof of SoS Certifiability**

$$\mathop{\mathbb{E}}_{X_i's}\left[\sup_{\widetilde{\mathbb{E}}:\text{degree-}t}\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Chaining and linearization

Talagrand's generic chaining. For any linear function class $\mathcal{F}_{\text{lin}}$

$$\mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}}\tfrac{1}{n}\sum_{i=1}^{n} f_i(X_i)\Big] \lesssim \mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}}\tfrac{1}{n}\sum_{i=1}^{n} f_i(G_i)\Big]$$

Our linearization lemma

Lemma. (f is linear-ish) For every $f \in \mathcal{F}$, there exists $\mathcal{G}$ such that

$$\sum_i f(x_i) := \Big(\sup_{(g_1,\dots,g_n)\in\mathcal{G}}\sum_i \langle g_i, x_i\rangle\Big)^t$$

$$\mathbb{E}_{X_i's}\left[\sup_{\widetilde{\mathbb{E}}:\text{degree-}t} \frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Chaining and linearization

Talagrand's generic chaining. For any linear function class $\mathcal{F}_{\text{lin}}$

$$\mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}} \frac{1}{n}\sum_{i=1}^{n} f_i(X_i)\Big] \lesssim \mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}} \frac{1}{n}\sum_{i=1}^{n} f_i(G_i)\Big]$$

Our linearization lemma

Lemma. (f is linear-ish) For every $f \in \mathcal{F}$, there exists $\mathcal{G}$ such that

$$\sum_i f(x_i) := \Big(\sup_{(g_1,\dots,g_n)\in\mathcal{G}} \sum_i \langle g_i, x_i\rangle\Big)^t$$

- Intuition: $\sum_i y_i^t = \|y\|_t^t = \sup_z \left\langle \frac{z}{\|z\|_q}, y\right\rangle^t$ by Holder's inequality

$$\mathbb{E}_{X_i's}\left[\sup_{\widetilde{\mathbb{E}}:\text{degree-}t}\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right]$$

$$f(x) = \frac{\widetilde{\mathbb{E}}[\langle v, X_i\rangle^t]}{\widetilde{\mathbb{E}}[\|v\|^t]}$$

Chaining and linearization

Talagrand's generic chaining. For any linear function class $\mathcal{F}_{\text{lin}}$

$$\mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}}\tfrac{1}{n}\sum_{i=1}^{n} f_i(X_i)\Big] \lesssim \mathbb{E}\Big[\sup_{(f_1,\dots,f_n)\in\mathcal{F}_{\text{lin}}}\tfrac{1}{n}\sum_{i=1}^{n} f_i(G_i)\Big]$$

Our linearization lemma

Lemma. (f is linear-ish) For every $f \in \mathcal{F}$, there exists $\mathcal{G}$ such that

$$\sum_i f(x_i) := \Big(\sup_{(g_1,\dots,g_n)\in\mathcal{G}}\sum_i \langle g_i, x_i\rangle\Big)^t$$

- Intuition: $\sum_i y_i^t = \|y\|_t^t = \sup_z \left\langle \frac{z}{\|z\|_q}, y\right\rangle^t$ by Holder's inequality
- Extends to $\widetilde{\mathbb{E}}$: (i) $y_i = \langle v, X_i\rangle$ and (ii) Holder's inequality holds for $\widetilde{\mathbb{E}}$