Polytope Sparsification

Shivam Nadimpalli
(MIT)

Joint work with

Anindya De
(Penn)

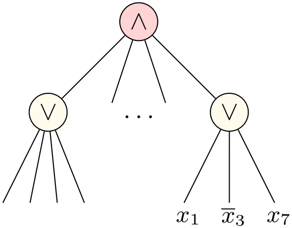
Ryan O'Donnell
(CMU)

Rocco Servedio
(Columbia)

Motivation from the Boolean Setting

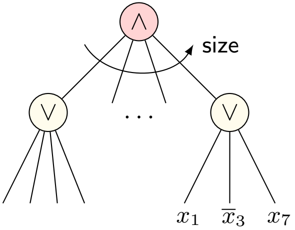
Motivation from the Boolean Setting: CNFs

Motivation from the Boolean Setting: CNFs



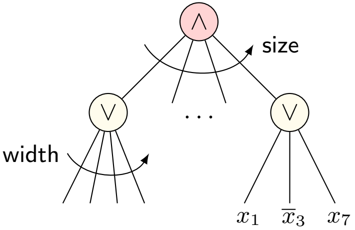
$f : \{0,1\}^n \to \{0,1\} \equiv \mathrm{AND}$-of-$\mathrm{OR}$ circuit representation

Motivation from the Boolean Setting: CNFs



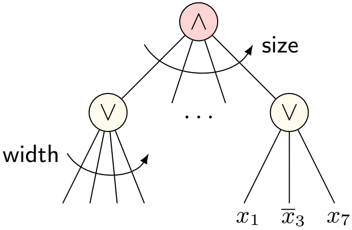
$f : \{0,1\}^n \to \{0,1\} \equiv \mathrm{AND}$-of-$\mathrm{OR}$ circuit representation

Motivation from the Boolean Setting: CNFs



$f : \{0,1\}^n \to \{0,1\} \equiv \mathrm{AND}$-of-$\mathrm{OR}$ circuit representation

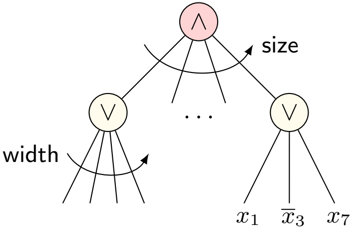
Motivation from the Boolean Setting: CNFs



$f : \{0,1\}^n \to \{0,1\} \equiv \mathrm{AND}$-of-$\mathrm{OR}$ circuit representation

Our Focus: Approximation of CNF by "simpler" CNF

Motivation from the Boolean Setting: CNFs



$f : \{0,1\}^n \to \{0,1\} \equiv \mathrm{AND\text{-}of\text{-}OR}$ circuit representation

Our Focus: Approximation of CNF by "simpler" CNF

The Hamming Distance

Definition: Given $f, g : \{0,1\}^n \to \{0,1\}$, we define

$$\mathrm{dist}(f, g) := \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \{0,1\}^n} \big[f(\boldsymbol{x}) \neq g(\boldsymbol{x}) \big].$$

The Hamming Distance

Definition: Given $f, g : \{0,1\}^n \to \{0,1\}$, we define

$$\mathrm{dist}(f,g) := \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \{0,1\}^n} \big[f(\boldsymbol{x}) \neq g(\boldsymbol{x}) \big].$$

The Hamming Distance

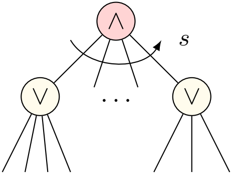
Definition: Given $f, g : \{0,1\}^n \to \{0,1\}$, we define

$$\mathrm{dist}(f,g) := \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \{0,1\}^n} \big[f(\boldsymbol{x}) \neq g(\boldsymbol{x}) \big].$$

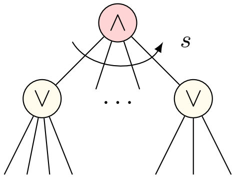
We will write $f \approx_\varepsilon g$ if $\mathrm{dist}(f,g) \leq \varepsilon$.

Warmup: Small CNFs are (Essentially) Narrow

Warmup: Small CNFs are (Essentially) Narrow

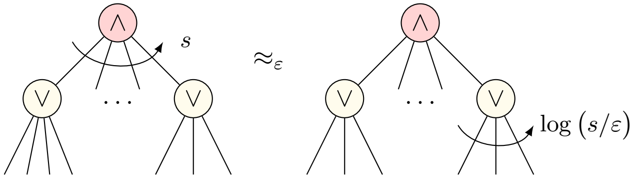


Warmup: Small CNFs are (Essentially) Narrow

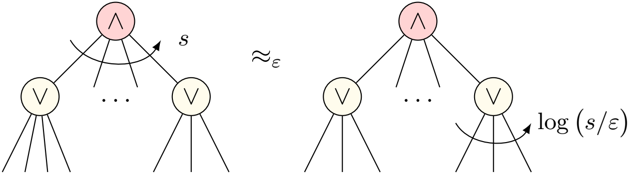


$\approx_\varepsilon$

Warmup: Small CNFs are (Essentially) Narrow

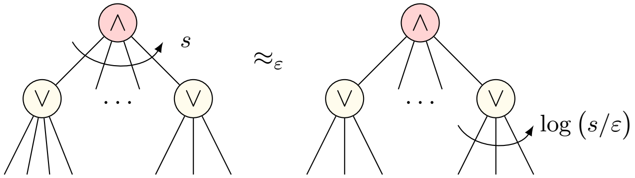


Warmup: Small CNFs are (Essentially) Narrow



(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

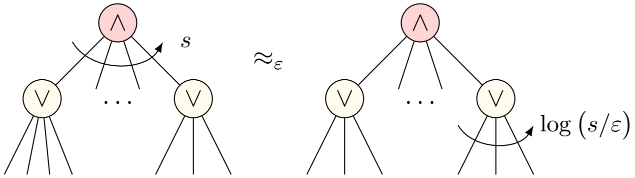
Warmup: Small CNFs are (Essentially) Narrow



(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

Idea: Drop wide clauses, take union bound ☺

Warmup: Small CNFs are (Essentially) Narrow

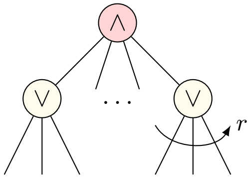


(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

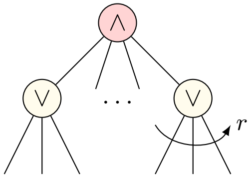
What about a converse?

Are Narrow CNFs (Essentially) Small?

Are Narrow CNFs (Essentially) Small?

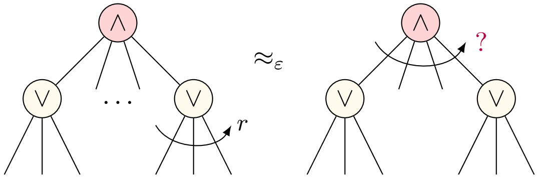


Are Narrow CNFs (Essentially) Small?

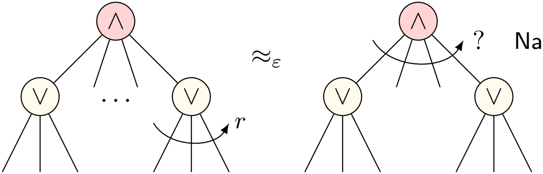


$\approx_\varepsilon$

Are Narrow CNFs (Essentially) Small?

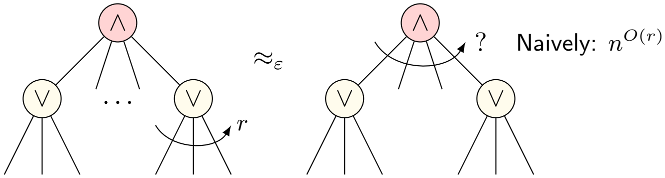


Are Narrow CNFs (Essentially) Small?



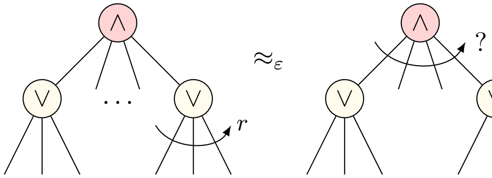
Naively: $n^{O(r)}$

Are Narrow CNFs (Essentially) Small?



Check: Suppose $r = 1$, i.e. f is just an AND of Boolean literals.

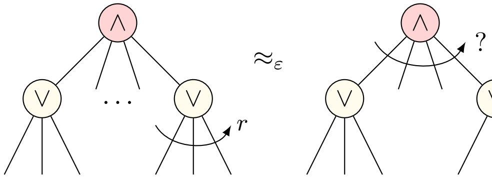
Are Narrow CNFs (Essentially) Small?



Naively: $n^{O(r)}$

Check: Suppose $r = 1$, i.e. f is just an AND of Boolean literals. If f has more than $\log(1/\varepsilon)$ literals, then $f \approx_\varepsilon 0$.

Are Narrow CNFs (Essentially) Small?



Naively: $n^{O(r)}$

Check: Suppose $r = 1$, i.e. f is just an AND of Boolean literals. If f has more than $\log(1/\varepsilon)$ literals, then $f \approx_\varepsilon 0$.

Hope: Is every width-r CNF $\approx_\varepsilon$ size-$O_{\varepsilon,r}(1)$ CNF?

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

$2^{O\left(r2^r \log\left(\frac{1}{\varepsilon}\right)\right)}$ [Trevisan '05]

Covering/packing argument

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

$2^{O\left(r 2^r \log\left(\frac{1}{\varepsilon}\right)\right)}$ [Trevisan '05]
Covering/packing argument

$\left(r \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ [Gopalan-Meka-Reingold '13]
Quasi-sunflowers

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

$2^{O\left(r2^r \log\left(\frac{1}{\varepsilon}\right)\right)}$ — [Trevisan '05]
Covering/packing argument

$\left(r \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ — [Gopalan-Meka-Reingold '13]
Quasi-sunflowers

$\left(\frac{1}{\varepsilon}\right)^{O(r)}$ — [Lovett-Zhang '18]
Switching lemma + noise stability

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

$2^{O\left(r 2^r \log\left(\frac{1}{\varepsilon}\right)\right)}$ — [Trevisan '05]
Covering/packing argument

$\left(r \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ — [Gopalan-Meka-Reingold '13]
Quasi-sunflowers

$\left(\frac{1}{\varepsilon}\right)^{O(r)}$ — [Lovett-Zhang '18]
Switching lemma + noise stability

$\left(2 + \frac{1}{r} \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ — [Lovett-Wu-Zhang '21]
"Mild" random restrictions + noise stability

SIMONS INSTITUTE for the Theory of Computing

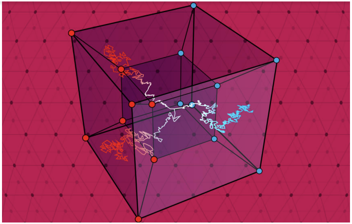
Home   Programs & Events   People   Participate   Support   News & Videos   About

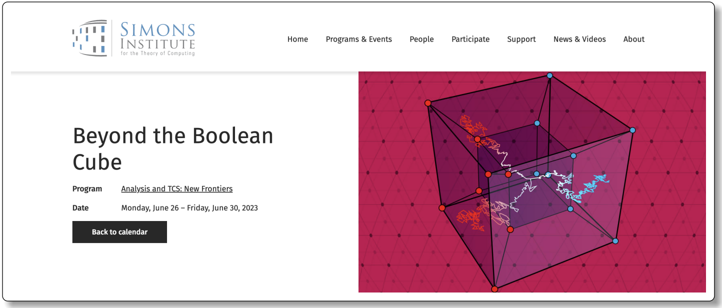
Beyond the Boolean Cube

Program Analysis and TCS: New Frontiers

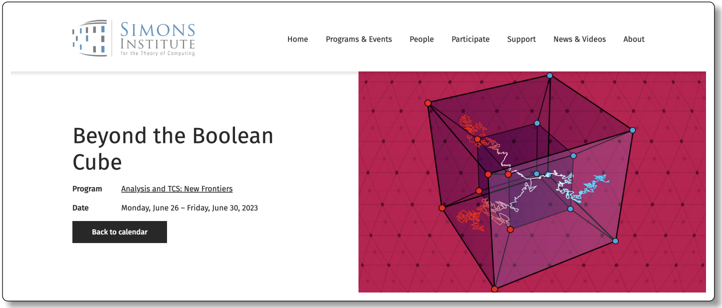
Date Monday, June 26 – Friday, June 30, 2023

Back to calendar



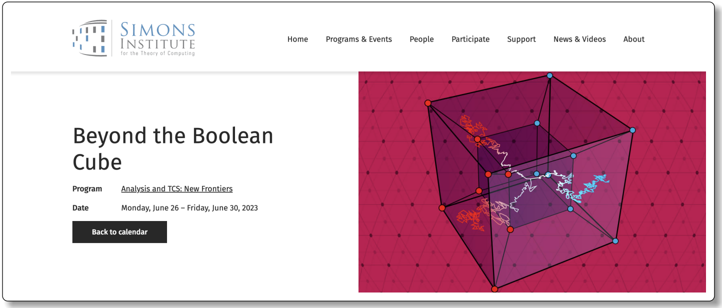


Analogue of CNF sparsification over Gaussian space?



Analogue of CNF sparsification over Gaussian space?

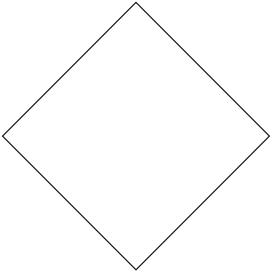
$\mathbb{R}^n$ with $N(0,1)^n$



Analogue of CNF sparsification over Gaussian space?

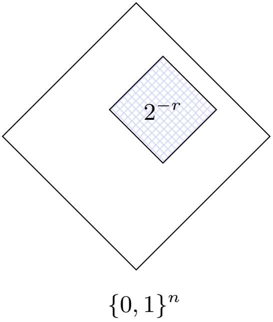
$\mathbb{R}^n$ with $N(0,1)^n$

Narrow Clauses



$\{0,1\}^n$

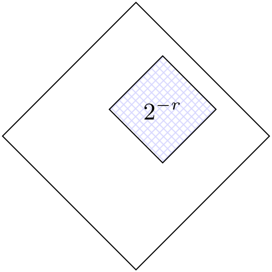
Narrow Clauses



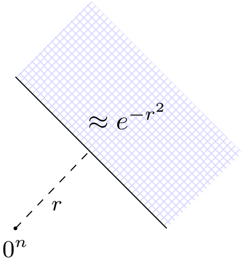
Clause of width r

e.g. $x_1 \vee \overline{x}_2 \longleftrightarrow 01 * \cdots *$

Narrow Clauses ≈ Narrow Halfspaces



$\{0,1\}^n$



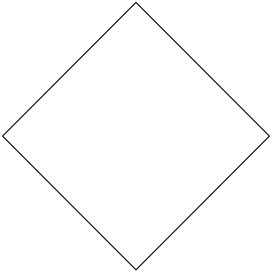
Clause of width r

e.g. $x_1 \vee \overline{x}_2 \longleftrightarrow 01 * \cdots *$

Halfspace of width r

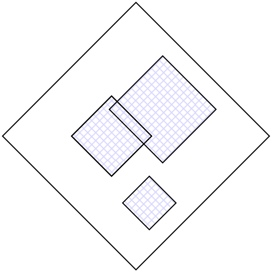
e.g. $\mathbf{1}\{x \cdot v \leq 10\}$ for $v \in \mathbb{S}^{n-1}$

Narrow CNFs



$\{0,1\}^n$

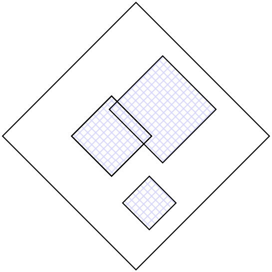
Narrow CNFs



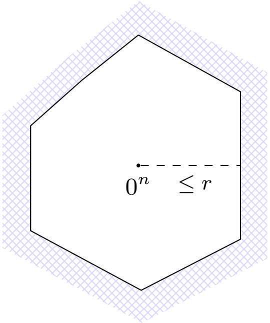
$\{0,1\}^n$

AND of width-r ORs

Narrow CNFs ≈ Bounded-Width Polytopes



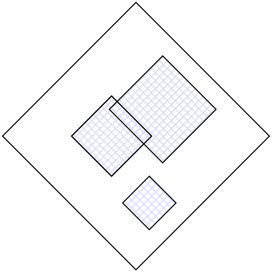
$\{0,1\}^n$



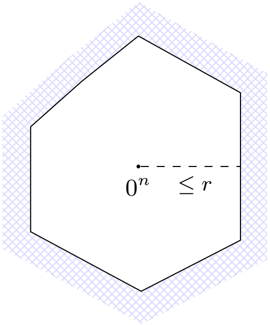
AND of width-r ORs

Intersection of width-r halfspaces

Narrow CNFs $\approx$ Bounded-Width Polytopes



$\{0,1\}^n$

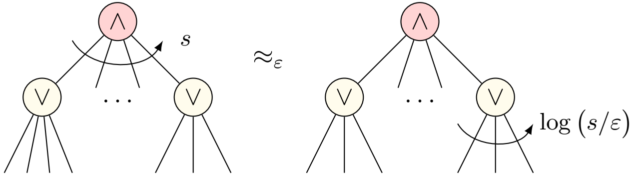


AND of width-r ORs

Intersection of width-r halfspaces

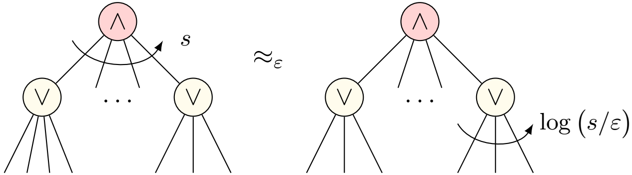
Width-r polytope

Recall: Small CNFs are (Essentially) Narrow



(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

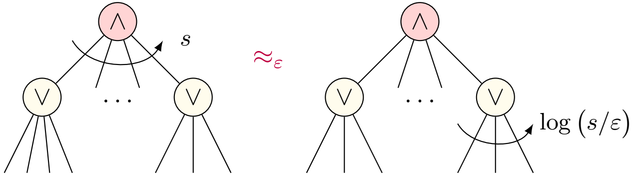
Recall: Small CNFs are (Essentially) Narrow



(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

Hope: A Gaussian analogue?

Recall: Small CNFs are (Essentially) Narrow



(Folklore) Fact: If $f : \{0,1\}^n \to \{0,1\}$ is computed by a size-s CNF, then $f \approx_\varepsilon g$ where g is a CNF of width $\log(s/\varepsilon)$.

Hope: A Gaussian analogue?

The Gaussian Distance

Definition: Given $K, L \subseteq \mathbb{R}^n$, we define

$$\mathrm{dist}(K, L) := \mathop{\mathbf{Pr}}_{\boldsymbol{g} \sim N(0,1)^n} \big[K(\boldsymbol{g}) \neq L(\boldsymbol{g}) \big].$$

The Gaussian Distance

Definition: Given $K, L \subseteq \mathbb{R}^n$, we define

$$\mathrm{dist}(K, L) := \mathop{\mathbf{Pr}}_{\boldsymbol{g} \sim N(0,1)^n} \big[K(\boldsymbol{g}) \neq L(\boldsymbol{g})\big].$$

The Gaussian Distance

Definition: Given $K, L \subseteq \mathbb{R}^n$, we define

$$\mathrm{dist}(K, L) := \mathop{\mathbf{Pr}}_{\boldsymbol{g} \sim N(0,1)^n} \big[K(\boldsymbol{g}) \neq L(\boldsymbol{g}) \big].$$

We will write $K \approx_\varepsilon L$ if $\mathrm{dist}(K, L) \leq \varepsilon$.

The Gaussian Distance

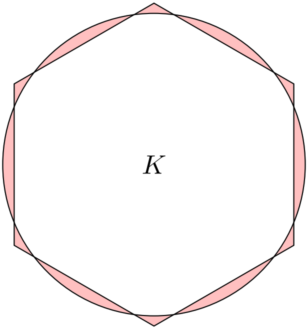
Definition: Given $K, L \subseteq \mathbb{R}^n$, we define

$$\mathrm{dist}(K, L) := \mathop{\mathbf{Pr}}_{\boldsymbol{g} \sim N(0,1)^n} \big[K(\boldsymbol{g}) \neq L(\boldsymbol{g})\big].$$

We will write $K \approx_\varepsilon L$ if $\mathrm{dist}(K, L) \leq \varepsilon$.

Note: Not translation invariant!

The Gaussian Distance

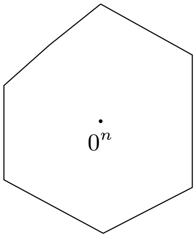


$L = H_1 \cap \ldots \cap H_6$

$$\mathrm{dist}(K, L) = \mathop{\mathbf{Pr}}_{\boldsymbol{g} \sim N(0,1)^n} \left[\boldsymbol{g} \in \square \right]$$

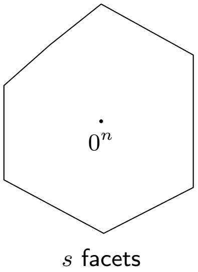
Intersections of Few Halfspaces are (Essentially) Narrow

Intersections of Few Halfspaces are (Essentially) Narrow



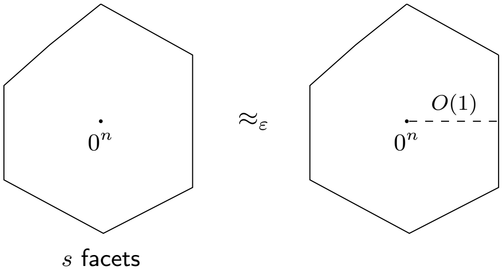
s facets

Intersections of Few Halfspaces are (Essentially) Narrow

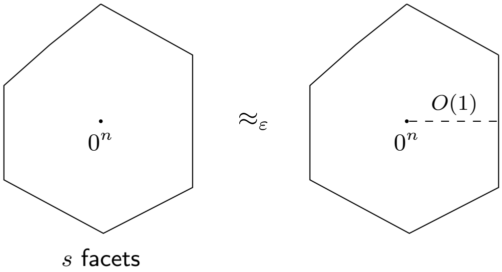


$\approx_\varepsilon$

Intersections of Few Halfspaces are (Essentially) Narrow

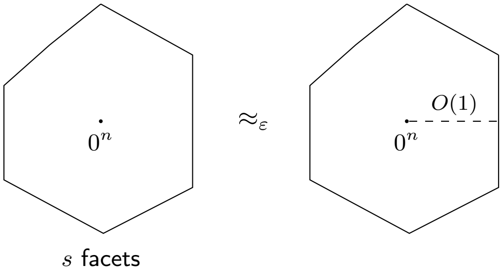


Intersections of Few Halfspaces are (Essentially) Narrow



(Folklore) Fact: If $K \subseteq \mathbb{R}^n$ is an intersection of s halfspaces, then $K \approx_\varepsilon L$ where L is a polytope of width $\sqrt{2\log(s/\varepsilon)}$.

Intersections of Few Halfspaces are (Essentially) Narrow

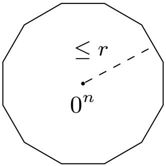


(Folklore) Fact: If $K \subseteq \mathbb{R}^n$ is an intersection of s halfspaces, then $K \approx_\varepsilon L$ where L is a polytope of width $\sqrt{2\log(s/\varepsilon)}$.

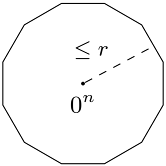
What about a converse?

Our Main Result: Polytope Sparsification

Our Main Result: Polytope Sparsification

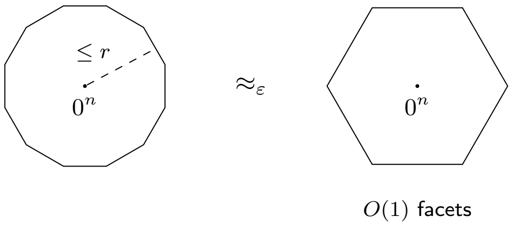


Our Main Result: Polytope Sparsification

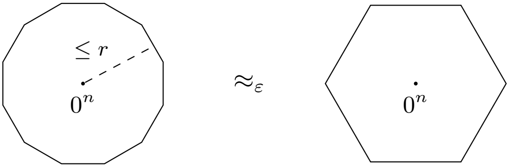


$\approx_\varepsilon$

Our Main Result: Polytope Sparsification



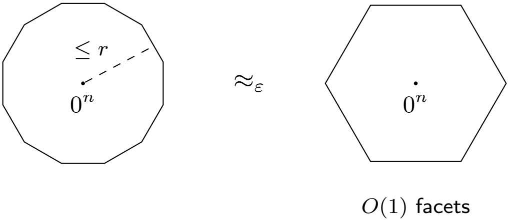
Our Main Result: Polytope Sparsification



$O(1)$ facets

[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$ halfspaces.

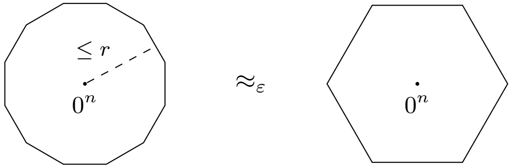
Our Main Result: Polytope Sparsification



[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$ halfspaces.

Yields efficient learning & tolerant testing algorithms

Our Main Result: Polytope Sparsification



$O(1)$ facets

[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$ halfspaces.

Yields efficient learning & tolerant testing algorithms

CNF Sparsification: A Brief History

Width-r CNFs are ε-approximated by CNFs of size:

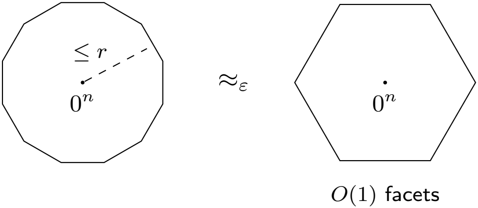
$2^{O\left(r2^r \log\left(\frac{1}{\varepsilon}\right)\right)}$ [Trevisan '05]
Covering/packing argument

$\left(r \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ [Gopalan-Meka-Reingold '13]
Quasi-sunflowers

$\left(\frac{1}{\varepsilon}\right)^{O(r)}$ [Lovett-Zhang '18]
Switching lemma + noise stability

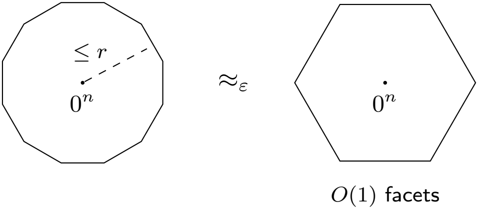
$\left(2 + \frac{1}{r} \log\left(\frac{1}{\varepsilon}\right)\right)^{O(r)}$ [Lovett-Wu-Zhang '21]
"Mild" random restrictions + noise stability

Our Main Result



[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$.

Our Main Result

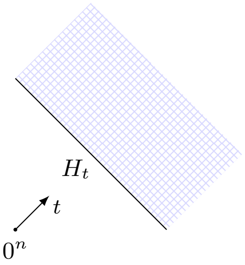


[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$.

Lower Bound: $\exp(1/\varepsilon)$ halfspaces necessary

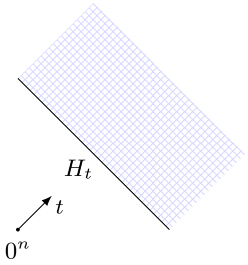
From Polytopes to Suprema

From Polytopes to Suprema



$$H_t := \{x \in \mathbb{R}^n : \langle x, t\rangle \leq 1\}$$

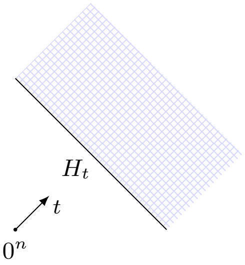
From Polytopes to Suprema

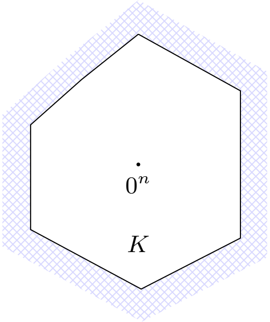


$$H_t := \{x \in \mathbb{R}^n : \langle x, t\rangle \leq 1\}$$

$$x \in H_t \iff \langle x, t\rangle \leq 1$$

From Polytopes to Suprema



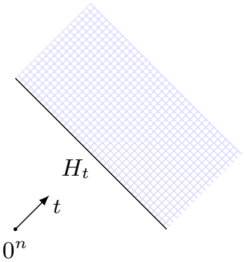


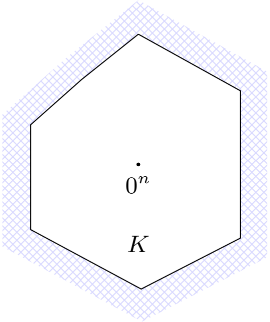
$$H_t := \{x \in \mathbb{R}^n : \langle x, t\rangle \leq 1\}$$

$$x \in H_t \iff \langle x, t\rangle \leq 1$$

$$K = \bigcap_{t \in T} H_t$$

From Polytopes to Suprema

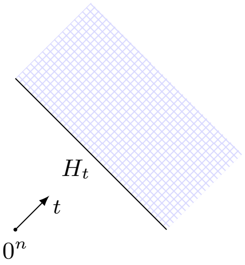


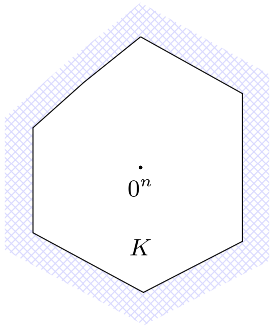


$$H_t := \{x \in \mathbb{R}^n : \langle x, t \rangle \le 1\} \qquad K = \bigcap_{t \in T} H_t$$

$$x \in H_t \iff \langle x, t \rangle \le 1 \qquad x \in K \iff \sup_{t \in T} \langle x, t \rangle \le 1$$

From Polytopes to Suprema





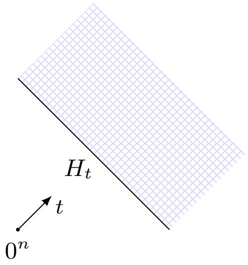
$$H_t := \{x \in \mathbb{R}^n : \langle x, t \rangle \le 1\}$$

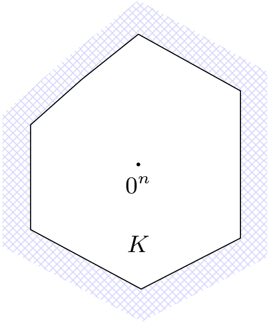
$$K = \bigcap_{t \in T} H_t$$

$$x \in H_t \iff \langle x, t \rangle \le 1$$

$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \le 1$$

From Polytopes to Suprema





$$H_t := \{x \in \mathbb{R}^n : \langle x, t\rangle \le 1\}$$

$$K = \bigcap_{t \in T} H_t$$

$$x \in H_t \iff \langle x, t\rangle \le 1$$

$$x \in K \iff \sup_{t \in T} \langle x, t\rangle \le 1$$

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

for the (canonical) Gaussian process on T.

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

for the (canonical) Gaussian process on T.

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

for the (canonical) Gaussian process on T.

Our Focus: $\sup\limits_{t \in T} \boldsymbol{X}_t$

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

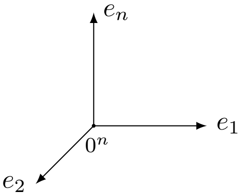
for the (canonical) Gaussian process on T.

Our Focus: $\sup_{t \in T} \boldsymbol{X}_t$

Compressed sensing, metric embeddings, random matrix theory, ...

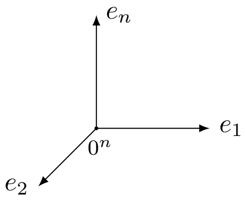
Example 1

Example 1



$$T = \{e_1, \ldots, e_n\}$$

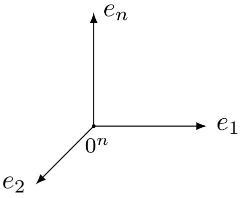
Example 1



$$T = \{e_1, \ldots, e_n\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right]$$

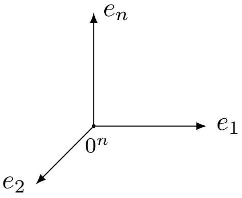
Example 1



$$T = \{e_1, \ldots, e_n\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right] = \mathop{\mathbf{E}}_{\boldsymbol{x}_i \sim N(0,1)}\left[\max_{1 \le i \le n} \boldsymbol{x}_i\right]$$

Example 1

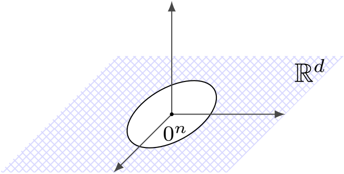


$$T = \{e_1, \dots, e_n\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right] = \mathop{\mathbf{E}}_{\boldsymbol{x}_i \sim N(0,1)}\left[\max_{1\le i\le n} \boldsymbol{x}_i\right] \asymp \sqrt{\log n}$$

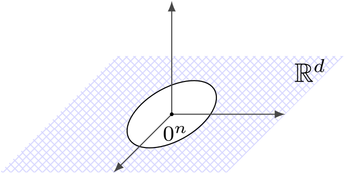
Example 2

Example 2



$$T = \left\{ x : \sum_{i=1}^{d} x_1^2 + \dots x_d^2 \leq 1 \text{ and } x_{d+1} = \cdots = x_n = 0 \right\}$$

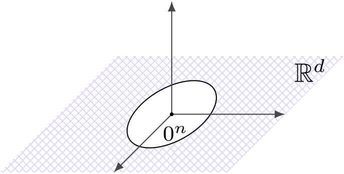
Example 2



$$T = \left\{ x : \sum_{i=1}^{d} x_1^2 + \dots x_d^2 \leq 1 \text{ and } x_{d+1} = \dots = x_n = 0 \right\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right]$$

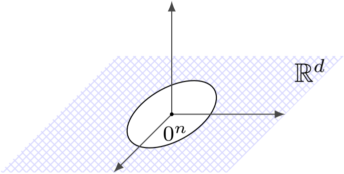
Example 2



$$T = \left\{ x : \sum_{i=1}^{d} x_1^2 + \dots x_d^2 \leq 1 \text{ and } x_{d+1} = \dots = x_n = 0 \right\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right] = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim N(0,1)^d} \left[\|\boldsymbol{x}\|_2\right]$$

Example 2



$$T = \left\{ x : \sum_{i=1}^{d} x_1^2 + \dots x_d^2 \le 1 \text{ and } x_{d+1} = \cdots = x_n = 0 \right\}$$

$$\mathbf{E}\left[\sup_t \boldsymbol{X}_t\right] = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim N(0,1)^d}\left[\|\boldsymbol{x}\|_2\right] \asymp \sqrt{d}$$

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

for the (canonical) Gaussian process on T.

Gaussian Processes

Definition: Let $T \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \sim N(0,1)^n$. We will write $\{\boldsymbol{X}_t\}_{t \in T}$ where

$$\boldsymbol{X}_t := \langle \boldsymbol{x}, t \rangle$$

for the (canonical) Gaussian process on T.

Informally: Complexity of T determines $\mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right]$

Sparsifying Gaussian Suprema

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t \in T}$ be the associated Gaussian process.

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t \in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$,

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t \in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t\in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s\in S}$

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t\in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s\in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t\in T} \boldsymbol{X}_t - \sup_{s\in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\, \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right],$$

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t\in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s\in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t\in T} \boldsymbol{X}_t - \sup_{s\in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\,\mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right],$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Sparsifying Gaussian Suprema

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t \in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

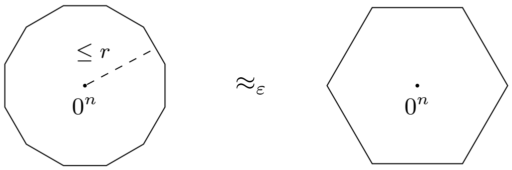
$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\, \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right],$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Note: Shifts c_s are necessary 😕

Application 1: Polytope Sparsification

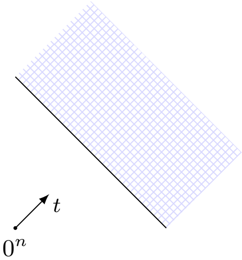
Application 1: Polytope Sparsification

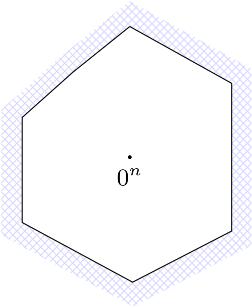


$O(1)$ facets

[DNOS'25]: If $K \subseteq \mathbb{R}^n$ is a polytope of width r, then $K \approx_\varepsilon L$ where L is an intersection of $\exp\exp(r^4/\varepsilon^2)$.

Application 1: Polytope Sparsification





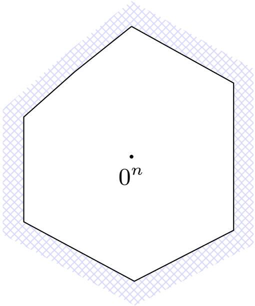
$$H_t := \{x \in \mathbb{R}^n : \langle x, t \rangle \leq 1\}$$

$$x \in H_t \iff \langle x, t \rangle \leq 1$$

$$K = \bigcap_{t \in T} H_t$$

$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \leq 1$$

Application 1: Polytope Sparsification

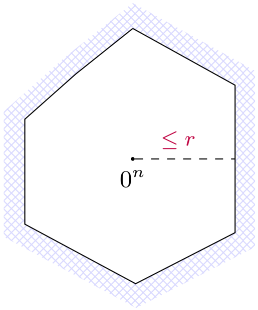


$$K = \bigcap_{t \in T} H_t$$

$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \leq 1$$

Application 1: Polytope Sparsification

Small width

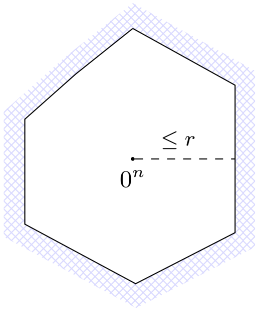


$$K = \bigcap_{t \in T} H_t$$

$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \leq 1$$

Application 1: Polytope Sparsification

Small width $\implies \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right] = O(1)$



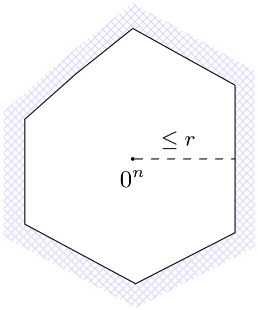
$$K = \bigcap_{t\in T} H_t$$

$$x \in K \iff \sup_{t\in T} \langle x, t\rangle \leq 1$$

Application 1: Polytope Sparsification

Small width $\implies \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality



$$K = \bigcap_{t \in T} H_t$$

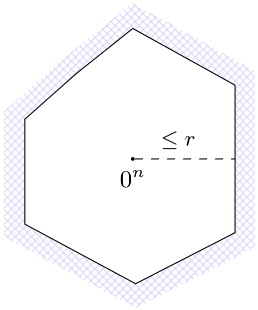
$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \leq 1$$

Application 1: Polytope Sparsification

Small width $\implies \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t\in T} \boldsymbol{X}_t = \sup_{s\in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.



$$K = \bigcap_{t\in T} H_t$$

$$x \in K \iff \sup_{t\in T} \langle x, t\rangle \leq 1$$

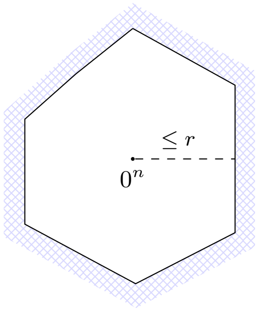
Application 1: Polytope Sparsification

Small width $\implies \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t \in T} \boldsymbol{X}_t = \sup_{s \in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.

Obtain a Polytope



$$K = \bigcap_{t \in T} H_t$$

$$x \in K \iff \sup_{t \in T} \langle x, t \rangle \leq 1$$

Application 1: Polytope Sparsification

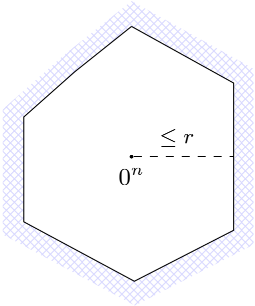
Small width $\Longrightarrow \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t\in T} \boldsymbol{X}_t = \sup_{s\in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.

Obtain a Polytope

$x \in L$



$$K = \bigcap_{t\in T} H_t$$

$$x \in K \iff \sup_{t\in T} \langle x, t\rangle \leq 1$$

Application 1: Polytope Sparsification

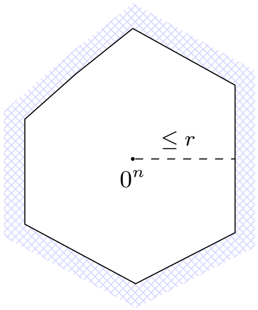
Small width $\implies \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t\in T} \boldsymbol{X}_t = \sup_{s\in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.

Obtain a Polytope

$x \in L \iff \sup_{s\in S}\{\langle x, s\rangle + c_s\} \leq 1$



$$K = \bigcap_{t\in T} H_t$$

$x \in K \iff \sup_{t\in T} \langle x, t\rangle \leq 1$

Application 1: Polytope Sparsification

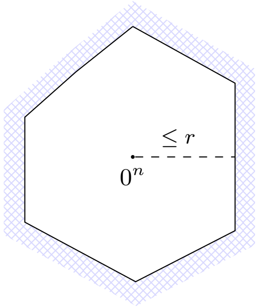
Small width $\implies \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t \in T} \boldsymbol{X}_t = \sup_{s \in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.

Obtain a Polytope

$$x \in L \iff \sup_{s \in S}\{\langle x, s\rangle + c_s\} \leq 1$$



$$K = \bigcap_{t \in T} H_t$$

$$x \in K \iff \sup_{t \in T} \langle x, t\rangle \leq 1$$

Application 1: Polytope Sparsification

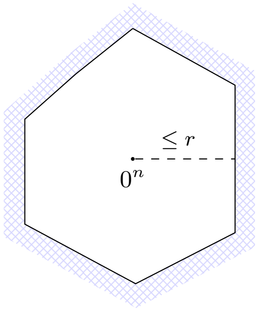
Small width $\implies \mathbf{E}\left[\sup_{t\in T} \boldsymbol{X}_t\right] = O(1)$

Sparsification + Markov's Inequality

$\sup_{t\in T} \boldsymbol{X}_t = \sup_{s\in S}\{\boldsymbol{X}_s + c_s\} \pm \varepsilon$ w.h.p.

Obtain a Polytope

$$x \in L \iff \sup_{s\in S}\{\langle x, s\rangle + c_s\} \le 1$$



$$K = \bigcap_{t\in T} H_t$$

$$x \in K \iff \sup_{t\in T} \langle x, t\rangle \le 1$$

Anti-concentration: $K \approx_\varepsilon L$

Back to Our Main Result

Back to Our Main Result

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ and let $\{\boldsymbol{X}_t\}_{t \in T}$ be the associated Gaussian process. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\, \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right],$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Application 2: Norms are Gaussian Juntas

Application 2: Norms are Gaussian Juntas

[DNOS'25]:

Application 2: Norms are Gaussian Juntas

[DNOS'25]: Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is a norm.

Application 2: Norms are Gaussian Juntas

[DNOS'25]: Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is a norm. For any $\varepsilon \in (0, 0.5)$, there exists a norm $\psi : \mathbb{R}^n \to \mathbb{R}$ such that

Application 2: Norms are Gaussian Juntas

[DNOS'25]: Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is a norm. For any $\varepsilon \in (0, 0.5)$, there exists a norm $\psi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0,1)^n}\left[1 - \varepsilon \le \frac{\phi(\boldsymbol{x})}{\psi(\boldsymbol{x})} \le 1 + \varepsilon\right] \ge 1 - \varepsilon\,,$$

Application 2: Norms are Gaussian Juntas

[DNOS'25]: Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is a norm. For any $\varepsilon \in (0, 0.5)$, there exists a norm $\psi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0,1)^n} \left[1 - \varepsilon \le \frac{\phi(\boldsymbol{x})}{\psi(\boldsymbol{x})} \le 1 + \varepsilon \right] \ge 1 - \varepsilon \, ,$$

and furthermore ψ only depends on $\exp\exp(1/\varepsilon^3)$ dimensions.

Application 2: Norms are Gaussian Juntas

[DNOS'25]: Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is a norm. For any $\varepsilon \in (0, 0.5)$, there exists a norm $\psi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim N(0,1)^n}\left[1 - \varepsilon \leq \frac{\phi(\boldsymbol{x})}{\psi(\boldsymbol{x})} \leq 1 + \varepsilon\right] \geq 1 - \varepsilon\,,$$

and furthermore ψ only depends on $\exp\exp(1/\varepsilon^3)$ dimensions.

Idea: $\phi(x) := \sup_{t \in T} \langle x, t \rangle$ for symmetric T

Recap

The Main Technical Result

Sparsifying Suprema of Gaussian Processes

Some Consequences

Polytope Sparsification, Norms are (Essentially) Juntas

Recap & Rest of Talk

The Main Technical Result

Sparsifying Suprema of Gaussian Processes

Some Consequences

Polytope Sparsification, Norms are (Essentially) Juntas

Elements of the Proof

Talagrand's Majorizing Measures Theorem

Our Main Theorem

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\, \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right],$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Our Main Theorem

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\, \mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right],$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

First: Rescale so that $\mathbf{E}[\sup_{t \in T} \boldsymbol{X}_t] = 1$

Our Main Theorem

[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ with $\mathbf{E}[\sup_{t \in T} \boldsymbol{X}_t] = 1$. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\,,$$

and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Our Main Theorem

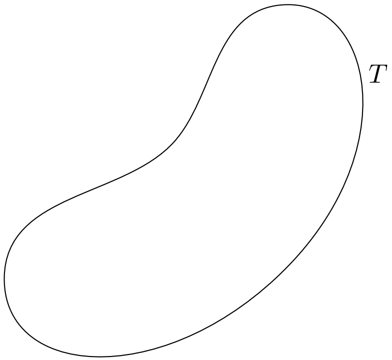
[DNOS'25]: Let $T \subseteq \mathbb{R}^n$ with $\mathbf{E}[\sup_{t \in T} \boldsymbol{X}_t] = 1$. For any $\varepsilon \in (0, 0.5)$, there is a subset $S \subseteq T$ and constants $\{c_s \geq 0\}_{s \in S}$ such that

$$\mathbf{E}\left[\left|\sup_{t \in T} \boldsymbol{X}_t - \sup_{s \in S}\{\boldsymbol{X}_s + c_s\}\right|\right] \leq \varepsilon\,,$$

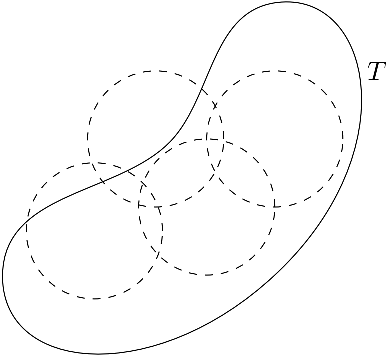
and furthermore $|S| \leq \exp\exp(1/\varepsilon)$.

Natural Idea: Cluster the points in T

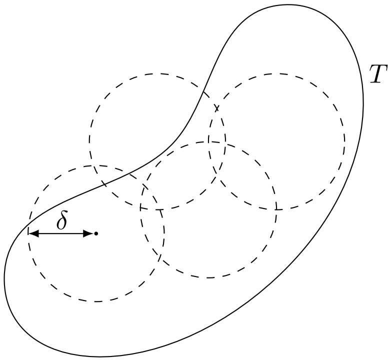
Attempt 1: A δ-Cover



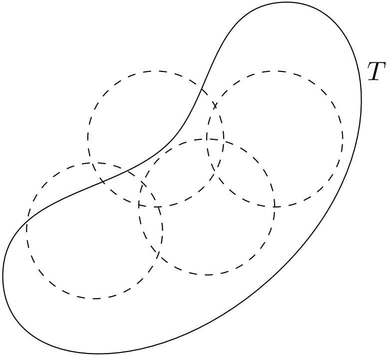
Attempt 1: A δ-Cover



Attempt 1: A δ-Cover

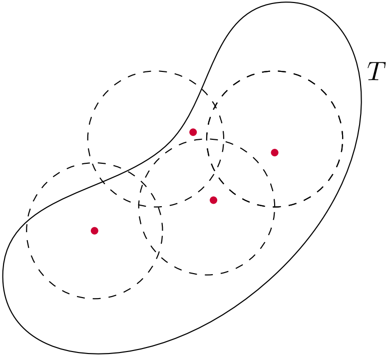


Attempt 1: A δ-Cover



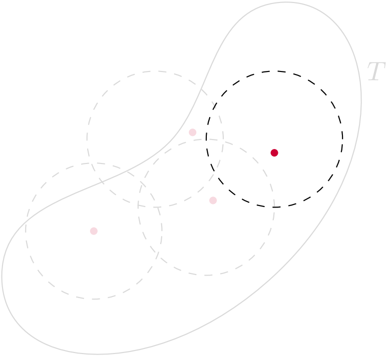
Suppose M clusters: $C_1, \ldots, C_M$.

Attempt 1: A δ-Cover



Suppose M clusters: $C_1, \ldots, C_M$. Pick rep. r_i from each C_i.

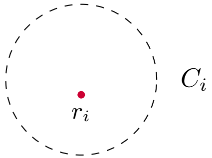
Attempt 1: A δ-Cover



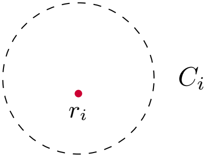
Suppose M clusters: $C_1, \ldots, C_M$. Pick rep. r_i from each C_i.

Approximation Within a Single δ-Ball

Approximation Within a Single δ-Ball

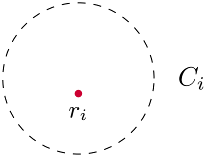


Approximation Within a Single δ-Ball



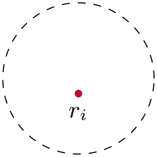
$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle$$

Approximation Within a Single δ-Ball



$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle = \langle \boldsymbol{x}, r_i \rangle + \sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle$$

Approximation Within a Single δ-Ball

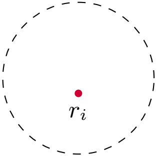


C_i

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle = \langle \boldsymbol{x}, r_i \rangle + \sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle$$

Super-concentrated!

Approximation Within a Single δ-Ball



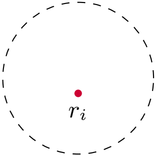
C_i

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle = \langle \boldsymbol{x}, r_i \rangle + \sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle$$

Super-concentrated!

Take $c_i := \mathbf{E}\left[\sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle\right]$,

Approximation Within a Single δ-Ball



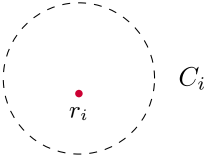
C_i

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle = \langle \boldsymbol{x}, r_i \rangle + \sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle$$

Super-concentrated!

Take $c_i := \mathbf{E}\left[\sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle\right]$, have $\mathbf{E}\left[\left|\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle - \langle \boldsymbol{x}, r_i \rangle - c_i\right|\right] \lesssim \delta$

Approximation Within a Single δ-Ball



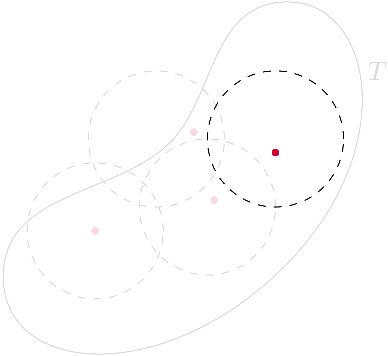
$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle = \langle \boldsymbol{x}, r_i \rangle + \sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle$$

Super-concentrated!

Take $c_i := \mathbf{E}\left[\sup_{t \in C_i} \langle \boldsymbol{x}, t - r_i \rangle\right]$, have $\mathbf{E}\left[\left|\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle - \langle \boldsymbol{x}, r_i \rangle - c_i\right|\right] \lesssim \delta$

(Consequence of Lipschitz concentration of Gaussians)

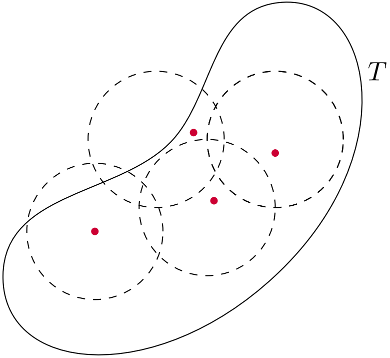
Approximation Across M Clusters



Within each cluster C_i:

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle \approx_\delta \langle \boldsymbol{x}, r_i \rangle + c_i$$

Approximation Across M Clusters

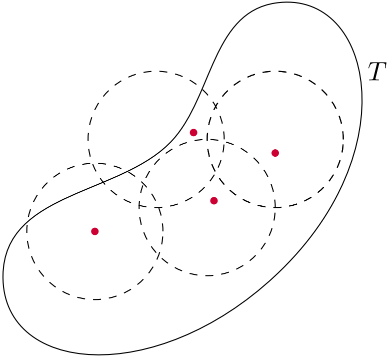


Within each cluster C_i:

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle \approx_\delta \langle \boldsymbol{x}, r_i \rangle + c_i$$

Natural Idea: Approximate $\sup_t \langle \boldsymbol{x}, t \rangle$ by $\sup_{i \in [M]} \{\langle \boldsymbol{x}, r_i \rangle + c_i\}$

Approximation Across M Clusters



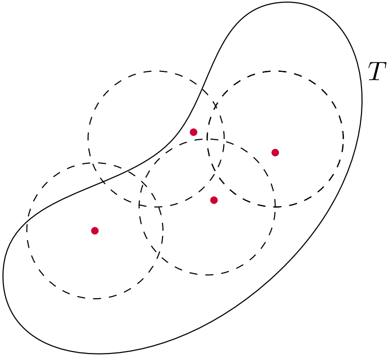
Within each cluster C_i:

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle \approx_\delta \langle \boldsymbol{x}, r_i \rangle + c_i$$

Natural Idea: Approximate $\sup_t \langle \boldsymbol{x}, t \rangle$ by $\sup_{i \in [M]} \{\langle \boldsymbol{x}, r_i \rangle + c_i\}$

Size $= M$

Approximation Across M Clusters



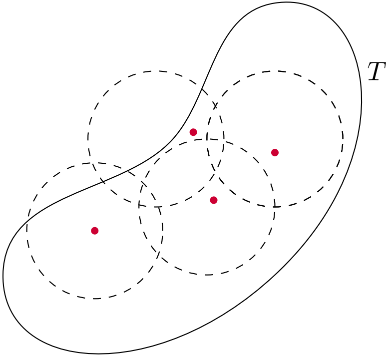
Within each cluster C_i:

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle \approx_\delta \langle \boldsymbol{x}, r_i \rangle + c_i$$

Natural Idea: Approximate $\sup_t \langle \boldsymbol{x}, t \rangle$ by $\sup_{i \in [M]} \{ \langle \boldsymbol{x}, r_i \rangle + c_i \}$

Size $= M$ Error

Approximation Across M Clusters



Within each cluster C_i:

$$\sup_{t \in C_i} \langle \boldsymbol{x}, t \rangle \approx_\delta \langle \boldsymbol{x}, r_i \rangle + c_i$$

Natural Idea: Approximate $\sup_t \langle \boldsymbol{x}, t \rangle$ by $\sup_{i \in [M]} \{\langle \boldsymbol{x}, r_i \rangle + c_i\}$

Size $= M$ Error $\asymp \delta \sqrt{\log M}$

Error Analysis: Controlling $\delta\sqrt{\log M}$

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$,

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$

Set δ as an appropriate polynomial of ε ☺

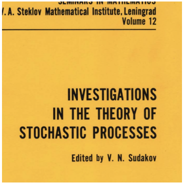
Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$

Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$,

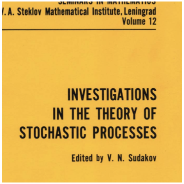
Sudakov Minoration



Theorem: If smallest δ net for $T \subseteq \mathbb{R}^n$ has size M, then

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim N(0,1)^n} \left[\sup_{t \in T} \langle \boldsymbol{x}, t \rangle \right] \gtrsim \delta \sqrt{\log M}.$$

Sudakov Minoration



Theorem: If smallest δ net for $T \subseteq \mathbb{R}^n$ has size M, then

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim N(0,1)^n} \left[\sup_{t \in T} \langle \boldsymbol{x}, t \rangle \right] \gtrsim \delta \sqrt{\log M}.$$

Recall: We rescaled to ensure LHS $= 1$

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$
Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$,

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$

Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$, then contradiction

Thanks to Sudakov minoration

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$
Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$, then contradiction
Thanks to Sudakov minoration

If $M = 2^{\Theta(1/\delta^2)} \ldots$

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$

Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$, then contradiction

Thanks to Sudakov minoration

If $M = 2^{\Theta(1/\delta^2)}$ . . . we're stuck?

Error Analysis: Controlling $\delta\sqrt{\log M}$

If $M = 2^{o(1/\delta^2)}$, then $\delta\sqrt{\log M} \to 0$ as $\delta \to 0$

Set δ as an appropriate polynomial of ε ☺

If $M = 2^{\omega(1/\delta^2)}$, then contradiction

Thanks to Sudakov minoration

If $M = 2^{\Theta(1/\delta^2)}$ ... we're stuck?

Issue: Gaussian processes s.t. the best δ-covering of the set has size $2^{\Theta(1/\delta^2)}$ at every fixed scale δ ☹

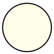

20
E3-20S
ESTWING

Talagrand's Majorizing Measures Theorem

Gives hierarchical sequence of partitions of T:

Talagrand's Majorizing Measures Theorem

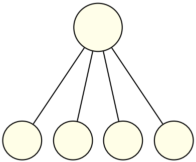
Gives hierarchical sequence of partitions of T:



$$\mathcal{A}_0 = \{T\}$$

Talagrand's Majorizing Measures Theorem

Gives hierarchical sequence of partitions of T:



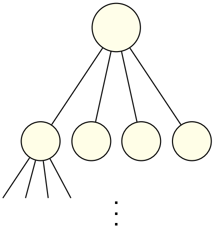
$\mathcal{A}_0 = \{T\}$

Refinement $\downarrow$

$\mathcal{A}_1$

Talagrand's Majorizing Measures Theorem

Gives hierarchical sequence of partitions of T:



$$\mathcal{A}_0 = \{T\}$$

Refinement

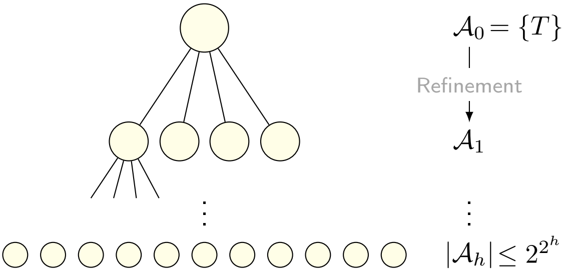
$$\downarrow$$

$$\mathcal{A}_1$$

$$\vdots$$

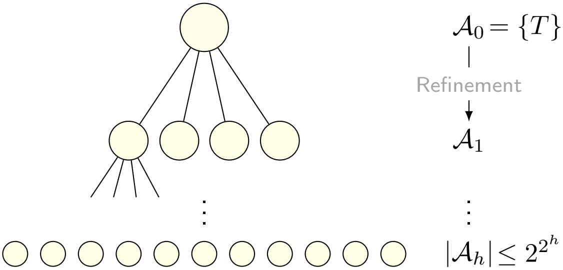
Talagrand's Majorizing Measures Theorem

Gives hierarchical sequence of partitions of T:



Talagrand's Majorizing Measures Theorem

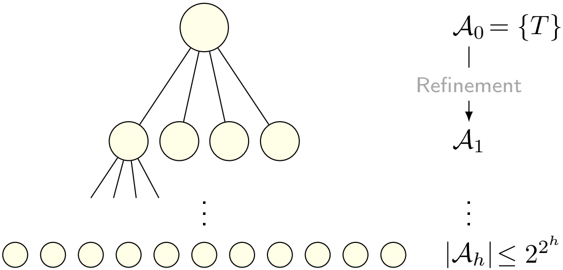
Gives hierarchical sequence of partitions of T:



$$\mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right]$$

Talagrand's Majorizing Measures Theorem

Gives hierarchical sequence of partitions of T:



$$\mathbf{E}\left[\sup_{t \in T} \boldsymbol{X}_t\right] \geq \sum_{h \geq 0} 2^{h/2} \cdot \mathrm{diam}(\mathcal{A}_h(t)) \qquad \text{for all } t \in T$$

Our Proof from 30,000 Feet

Recall we have Gaussian process $\{\boldsymbol{X}_t\}_{t \in T}$ with $\mathbf{E}[\sup \boldsymbol{X}_t] = 1$

Thanks to rescaling assumption

Our Proof from 30,000 Feet

Recall we have Gaussian process $\{\boldsymbol{X}_t\}_{t\in T}$ with $\mathbf{E}[\sup \boldsymbol{X}_t] = 1$

Thanks to rescaling assumption

Take MM Tree: $1 \geq \sum_{h\geq 0} 2^{h/2} \cdot \mathrm{diam}(\mathcal{A}_h(t))$

The h^{th} level has $\exp\exp(h)$ nodes

Our Proof from 30,000 Feet

Recall we have Gaussian process $\{\boldsymbol{X}_t\}_{t\in T}$ with $\mathbf{E}[\sup \boldsymbol{X}_t] = 1$

Thanks to rescaling assumption

Take MM Tree: $1 \geq \sum_{h\geq 0} 2^{h/2} \cdot \mathrm{diam}(\mathcal{A}_h(t))$

The h^{th} level has $\exp\exp(h)$ nodes

Simple algorithm to choose parts from Talagrand's tree

Terminates by depth $1/\varepsilon$, then proceed as before

Our Proof from 30,000 Feet

Recall we have Gaussian process $\{\boldsymbol{X}_t\}_{t \in T}$ with $\mathbf{E}[\sup \boldsymbol{X}_t] = 1$

Thanks to rescaling assumption

Take MM Tree: $1 \geq \sum_{h \geq 0} 2^{h/2} \cdot \mathrm{diam}(\mathcal{A}_h(t))$

The h^{th} level has $\exp\exp(h)$ nodes

Simple algorithm to choose parts from Talagrand's tree

Terminates by depth $1/\varepsilon$, then proceed as before

See paper for full details (and more) ☺

Recap

The Main Technical Result

Sparsifying Suprema of Gaussian Processes

Some Consequences

Polytope Sparsification, Norms are (Essentially) Juntas

Elements of the Proof

Talagrand's Majorizing Measures Theorem

Many Open Directions

Closing the gap in our bounds?

Singly- versus doubly-exponential in $1/\varepsilon$?

Many Open Directions

Closing the gap in our bounds?

Singly- versus doubly-exponential in $1/\varepsilon$?

Other applications?

Suprema are everywhere

Many Open Directions

Closing the gap in our bounds?

Singly- versus doubly-exponential in $1/\varepsilon$?

Other applications?

Suprema are everywhere

Generalizations or variants?

Sparsifying suprema of Bernoulli processes? Empirical processes?

Many Open Directions

Closing the gap in our bounds?

Singly- versus doubly-exponential in $1/\varepsilon$?

Other applications?

Suprema are everywhere

Generalizations or variants?

Sparsifying suprema of Bernoulli processes? Empirical processes?

Thanks for listening! Questions? ☺