Gaussian Partitioning for Reliable Learning of Multi-Index Models

Ilias Diakonikolas Giannis Iakovidis Daniel Kane Mingchen Ma

Submitted to STOC 2027

Multi-Index Models

$x^W := \mathrm{Proj}_W x$

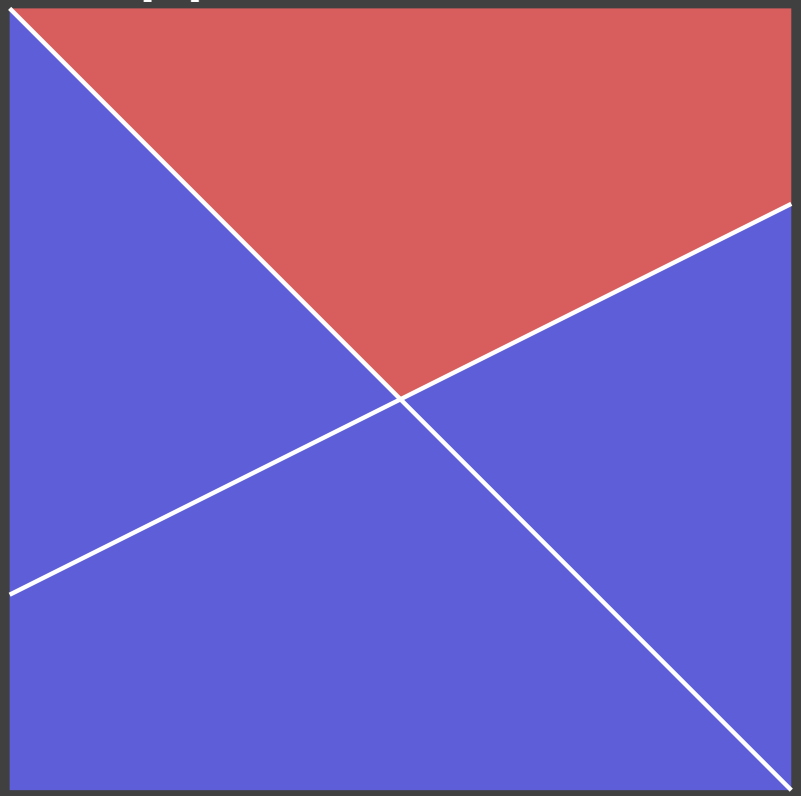
Definition (Multi-Index Models (MIMs))

A function $f : \mathbb{R}^d \to \mathcal{Y}$ is called a K-MIM if there exists a subspace $W \subseteq \mathbb{R}^d$ of dimension at most K such that $f(x) = f(x^W)$.

- Essentially each function depends on the projection onto a low dimensional subspace.
- We assume that $K \ll d$.
- We consider the case of binary classification, $\mathcal{Y} = \{\pm 1\}$.

Example MIMs

$$\prod_{i \in [K]} \mathbb{1}\left(w^{(i)} \cdot x + t_i \geq 0\right)$$



Learning Goal

- Let D be a distribution over $\mathbb{R}^d \times \{\pm 1\}$ with D_x well-behaved.
- Let $\mathcal{F}$ be a MIM class, e.g., intersections of K-halfspaces.

Learning Goal

- Let D be a distribution over $\mathbb{R}^d \times \{\pm 1\}$ with D_x well-behaved.
- Let $\mathcal{F}$ be a MIM class, e.g., intersections of K-halfspaces.

Goal:

- Given N samples, $(x^{(i)}, y_i) \sim D$.
- Find an algorithm that runs in $\mathrm{poly}(N)$ time and returns a hypothesis h comparable with the best-in-class

$$\mathop{\mathbf{Pr}}_{(x,y)\sim D}[h(x) \neq y] \leq c(\mathrm{OPT}) + \epsilon \ , \ \mathrm{OPT} := \inf_{f \in \mathcal{F}} \mathop{\mathbf{Pr}}_{D_x}[f(x) \neq y]$$

where c small function of OPT.

Learning Goal

- Let D be a distribution over $\mathbb{R}^d \times \{\pm 1\}$ with D_x well-behaved.
- Let $\mathcal{F}$ be a MIM class, e.g., intersections of K-halfspaces.

Goal:

- Given N samples, $(x^{(i)}, y_i) \sim D$.
- Find an algorithm that runs in $\mathrm{poly}(N)$ time and returns a hypothesis h comparable with the best-in-class

$$\Pr_{(x,y)\sim D}[h(x) \neq y] \leq c(\mathrm{OPT}) + \epsilon \ , \ \mathrm{OPT} := \inf_{f \in \mathcal{F}} \Pr_{D_x}[f(x) \neq y]$$

where c small function of OPT.

Semi-Agnostic Learning: Learn to error $(1+c)\mathrm{OPT} + \epsilon$ with $N = \mathrm{poly}(d/\epsilon)$ for any fixed c and class parameters.

Prior Work

[Diakonikolas–Kane–Ma, FOCS 2026] established the following guarantee:

Theorem (Agnostic Learning via Logconcave Sandwiching)

Let $\mathcal{F}$ be a Boolean K-MIM class, let $D_x \le P/\sigma$ for a known logconcave distr. P. There exists an algorithm that uses $n = d^{O(\ell(c))}\mathrm{poly}_{K,c}(1/(\sigma\epsilon))$ samples from D, runs in $d^{O(\ell(c))}\exp\big(\mathrm{poly}_{K,c}(\log(1/(\sigma\epsilon)))\big)$ time, and returns $\widehat{h}$ such that w.h.p.

$$\mathrm{err}_D(\widehat{h}) \le (1+c)\inf_{f\in\mathcal{F}}\mathrm{err}_D(f) + \epsilon.$$

Definition (Logconcave Sandwiching Degree)

For every $c \in (0,1)$, every $f \in \mathcal{F}$, and every logconcave distribution μ, there exist polynomials $p^{\pm}$ of degree at most $\ell(c)$ satisfying

$$\mathbb{E}_{z\sim\mu}\big[|p^+(z) - p^-(z)|\big] \le c, \quad p^-(z) \le f(z) \le p^+(z)$$

Central Question

Known Gaussian sandwiching bounds can be exponentially smaller than those for general logconcave distributions.

For example, for Boolean functions of K halfspaces:

- General logconcave: $\exp\big(\mathrm{poly}(K/c)\big)$.
- Standard Gaussian: $\mathrm{poly}(K/c)$.

Even for Gaussian inputs, the logconcave algorthim has exponent greater than the Gaussian sandwiching degree.

Central Question

Known Gaussian sandwiching bounds can be exponentially smaller than those for general logconcave distributions.

For example, for Boolean functions of K halfspaces:

- General logconcave: $\exp\big(\mathrm{poly}(K/c)\big)$.
- Standard Gaussian: $\mathrm{poly}(K/c)$.

Even for Gaussian inputs, the logconcave algorthim has exponent greater than the Gaussian sandwiching degree.

Question: Can the exponent of d depend on the Gaussian sandwiching degree when D_x is close to a Gaussian?

Central Question

Known Gaussian sandwiching bounds can be exponentially smaller than those for general logconcave distributions.

For example, for Boolean functions of K halfspaces:

- General logconcave: $\exp(\mathrm{poly}(K/c))$.
- Standard Gaussian: $\mathrm{poly}(K/c)$.

Even for Gaussian inputs, the logconcave algorthim has exponent greater than the Gaussian sandwiching degree.

Question: Can the exponent of d depend on the Gaussian sandwiching degree when D_x is close to a Gaussian?

We obtain such an result by introducing **Gaussian partitioning**, a new algorithmic primitive with applications beyond agnostic learning.

Roadmap

1. **Gaussian Partitioning**
2. **Agnostic and Testable learning**
3. **PUI and Reliable learning**
4. **Constructing a Gaussian Partition**

Gaussian Partitioning

Given samples from $Q \leq \mathcal{N}(0, I)/\sigma$.

Goal: Construct a small polyhedral partition with number of cells independent of d s.t. cells covering almost all of the Q-mass satisfy:

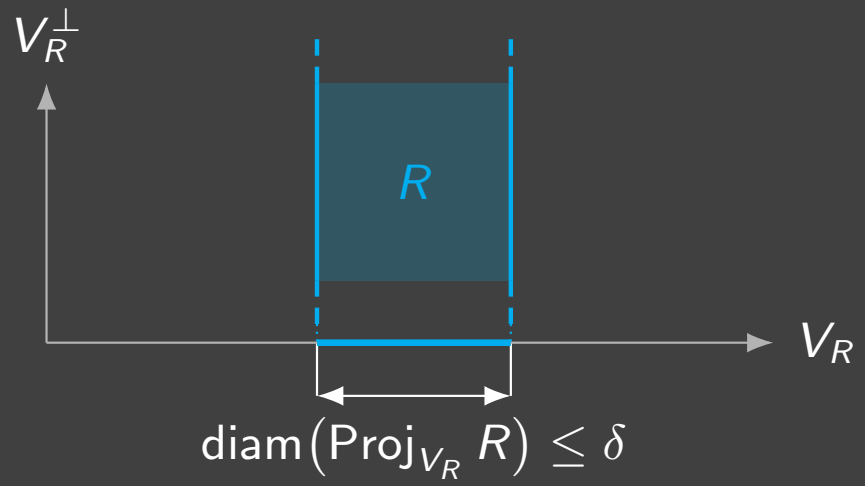
Gaussian Partitioning

$V_R := \text{span}\{\text{constraint normals of } R\}$

Given samples from $Q \leq \mathcal{N}(0, I)/\sigma$.

Goal: Construct a small polyhedral partition with number of cells independent of d s.t. cells covering almost all of the Q-mass satisfy:

- Thinness in constrained directions: $\text{diam}\big(\text{Proj}_{V_R} R\big) \leq \delta$.



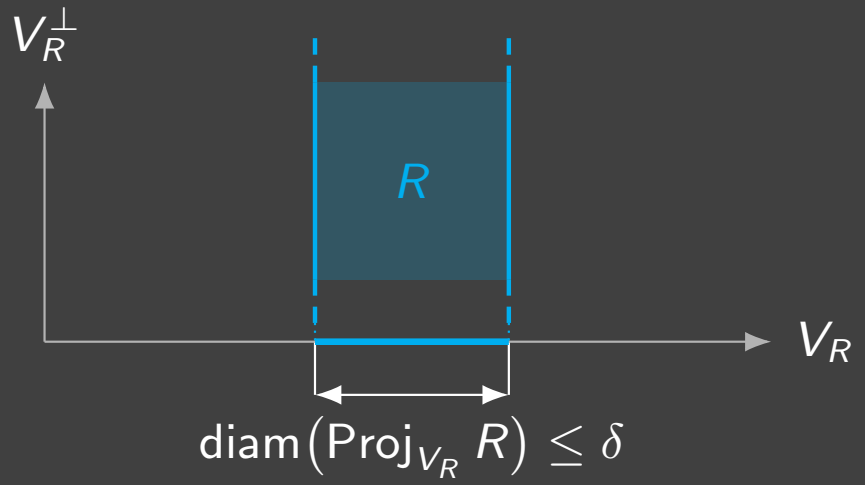
Gaussian Partitioning

$V_R := \text{span}\{\text{constraint normals of } R\}$

Given samples from $Q \leq \mathcal{N}(0, I)/\sigma$.

Goal: Construct a small polyhedral partition with number of cells independent of d s.t. cells covering almost all of the Q-mass satisfy:

- Thinness in constrained directions: $\text{diam}\left(\text{Proj}_{V_R} R\right) \leq \delta$.
- Gaussian moments in the remaining directions: For every unit $v \in V_R^{\perp}$ and $j \leq m$, $\left|\mathbb{E}_{x \sim Q|R}[(v \cdot x)^j] - \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[Z^j]\right| \leq \Delta$.



Gaussian Partitioning

$V_R := \text{span}\{\text{constraint normals of } R\}$

Given samples from $Q \leq \mathcal{N}(0, I)/\sigma$.

Goal: Construct a small polyhedral partition with number of cells independent of d s.t. cells covering almost all of the Q-mass satisfy:

- Thinness in constrained directions: $\text{diam}\left(\text{Proj}_{V_R} R\right) \leq \delta$.
- Gaussian moments in the remaining directions: For every unit $v \in V_R^{\perp}$ and $j \leq m$, $\left|\mathbb{E}_{x \sim Q|R}[(v \cdot x)^j] - \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[Z^j]\right| \leq \Delta$.

Plan:

- Thinness allows us to fix the V_R coordinates.
- Moment matching allows us to run algorithms with gaussian performance in that each cell.

Gaussian Partitioning

$V_R := \text{span}\{\text{constraint normals of } R\}$

Given samples from $Q \leq \mathcal{N}(0, I)/\sigma$.

Goal: Construct a small polyhedral partition with number of cells independent of d s.t. cells covering almost all of the Q-mass satisfy:

- Thinness in constrained directions: $\operatorname{diam}\left(\operatorname{Proj}_{V_R} R\right) \leq \delta$.
- Gaussian moments in the remaining directions: For every unit $v \in V_R^{\perp}$ and $j \leq m$, $\left|\mathbb{E}_{x \sim Q|R}[(v \cdot x)^j] - \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[Z^j]\right| \leq \Delta$.

Plan:

- Thinness allows us to fix the V_R coordinates.
- Moment matching allows us to run algorithms with gaussian performance in that each cell.

Similar partition exist for logconcave distributions [Diakonikolas–Kane–Ma, FOCS 2026], but it lacks the thinness guarantee needed to preserve gaussianity within each cell.

Gaussian Partitioning Guarantee

Theorem (Gaussian Partitioning)

Let $Q \le \mathcal{N}(0, I)/\sigma$, $m \ge 1$ *and* $\eta, \delta, \Delta \in (0, 1/2)$.
There is an algorithm using at most $n = d^m B$ *samples from* Q *and* $\mathrm{poly}(n)$ *time, where* $B = \exp\left(\mathrm{polylog}_{m,\Delta}\left(\frac{1}{\sigma\eta\delta}\right)\right)$.
With probability at least $2/3$, *it returns a polyhedral partition* $\mathcal{P}$ *and good cells* $\mathcal{G} \subseteq \mathcal{P}$ *satisfying* $|\mathcal{P}| \le B, Q\left(\bigcup_{R\in\mathcal{G}} R\right) \ge 1 - \eta$.
For every $R \in \mathcal{G}$:

1. $Q(R) \ge 1/B$.
2. $\mathrm{diam}\left(\mathrm{Proj}_{V_R} R\right) \le \delta$.
3. $Q \mid R$ *matches* m *with the Gaussian on* $V_R^\perp$, *with slack* Δ.

Gaussian Partitioning Guarantee

Theorem (Gaussian Partitioning)

Let $Q \le \mathcal{N}(0, I)/\sigma$, $m \ge 1$ *and* $\eta, \delta, \Delta \in (0, 1/2)$.
There is an algorithm using at most $n = d^m B$ *samples from* Q *and* $\mathrm{poly}(n)$ *time, where* $B = \exp\left(\mathrm{polylog}_{m,\Delta}\left(\frac{1}{\sigma\eta\delta}\right)\right)$.
With probability at least $2/3$, *it returns a polyhedral partition* $\mathcal{P}$ *and good cells* $\mathcal{G} \subseteq \mathcal{P}$ *satisfying* $|\mathcal{P}| \le B, Q\left(\bigcup_{R\in\mathcal{G}} R\right) \ge 1 - \eta$.
For every $R \in \mathcal{G}$:

1. $Q(R) \ge 1/B$.
2. $\mathrm{diam}\left(\mathrm{Proj}_{V_R} R\right) \le \delta$.
3. $Q \mid R$ *matches* m *with the Gaussian on* $V_R^\perp$, *with slack* Δ.

First we focus on applications of Gaussian partitioning.
Then we discuss how it differs from previous logconcave partitioning result.

Gaussian Partitioning under Smoothness

Our partitioning also handles unbounded density ratios.

Definition ((q, σ)-Smoothness)

For $q \geq 1$ and $\sigma \in (0, 1]$, Q is (q, σ)-smooth with respect to P if for every measurable event E, $Q(E) \leq \frac{1}{\sigma} P(E)^{1/q}$.

For $q = 1$, this is exactly $Q \leq P/\sigma$. Events rare under P remain rare under Q.

Gaussian Partitioning under Smoothness

Our partitioning also handles unbounded density ratios.

Definition ((q, σ)-Smoothness)

For $q \geq 1$ and $\sigma \in (0, 1]$, Q is (q, σ)-smooth with respect to P if for every measurable event E, $Q(E) \leq \frac{1}{\sigma} P(E)^{1/q}$.

For $q = 1$, this is exactly $Q \leq P/\sigma$. Events rare under P remain rare under Q.

Theorem (Gaussian Partitioning under Smoothness)

Let Q be (q, σ)-smooth with respect to $\mathcal{N}(0, I)$. For $m \geq 1$ and $\eta, \delta, \Delta \in (0, 1/2)$, an algorithm uses $n = d^m B$ samples and $\mathrm{poly}(n)$ time, where $B = \exp\left(\mathrm{polylog}_{q,m,\Delta}\left(\frac{1}{\sigma\eta\delta}\right)\right)$. With probability at least $2/3$, all previous partition guarantees hold.

Gaussian Partitioning under Smoothness

Our partitioning also handles unbounded density ratios.

Definition ((q, σ)-Smoothness)

For $q \geq 1$ and $\sigma \in (0, 1]$, Q is (q, σ)-smooth with respect to P if for every measurable event E, $Q(E) \leq \frac{1}{\sigma} P(E)^{1/q}$.

For $q = 1$, this is exactly $Q \leq P/\sigma$. Events rare under P remain rare under Q.

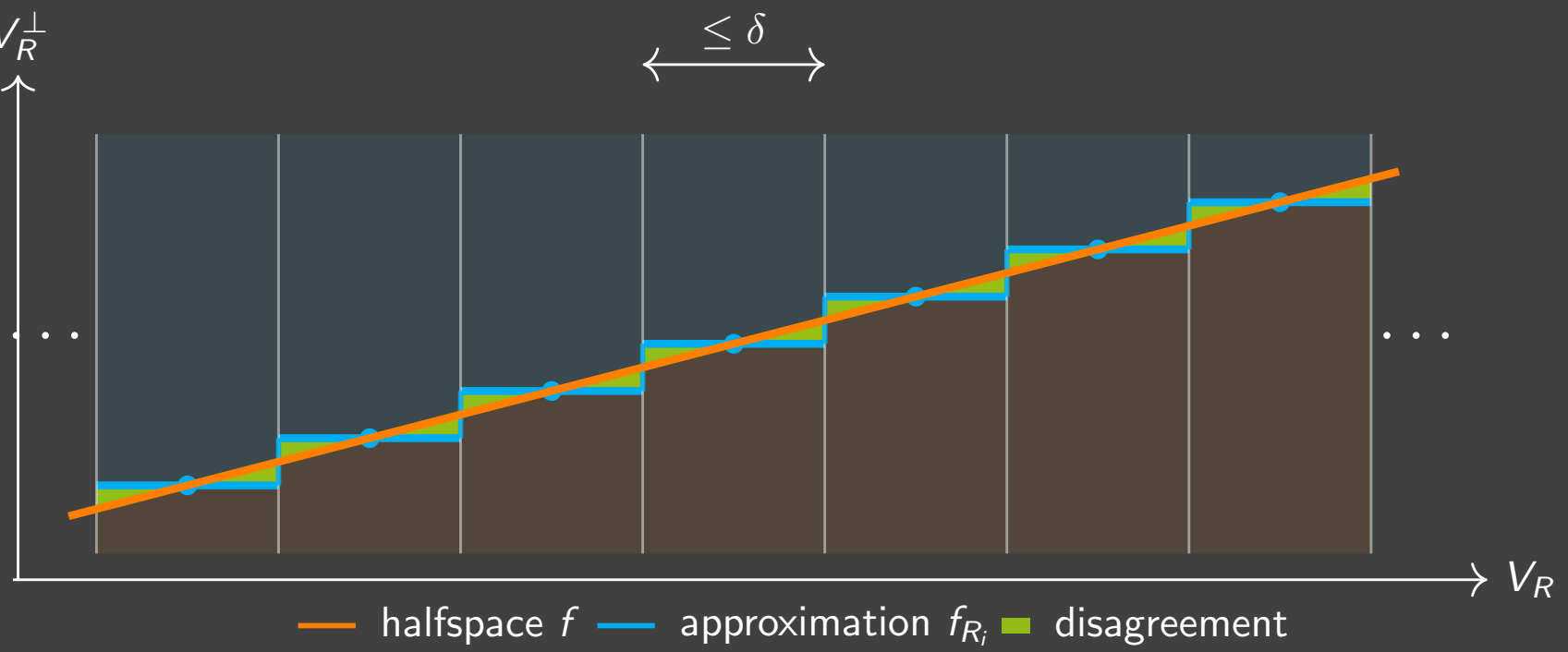
Theorem (Gaussian Partitioning under Smoothness)

Let Q be (q, σ)-smooth with respect to $\mathcal{N}(0, I)$. For $m \geq 1$ and $\eta, \delta, \Delta \in (0, 1/2)$, an algorithm uses $n = d^m B$ samples and $\mathrm{poly}(n)$ time, where $B = \exp\left(\mathrm{polylog}_{q,m,\Delta}\left(\frac{1}{\sigma\eta\delta}\right)\right)$. With probability at least $2/3$, all previous partition guarantees hold.

Smoothness w.r.t. $\mathcal{N}(0, I)$ covers many distributions, including Gaussian mixtures with bounded means and identity covariance, and distributions with bounded χ^2 or Rényi divergence from $\mathcal{N}(0, I)$.

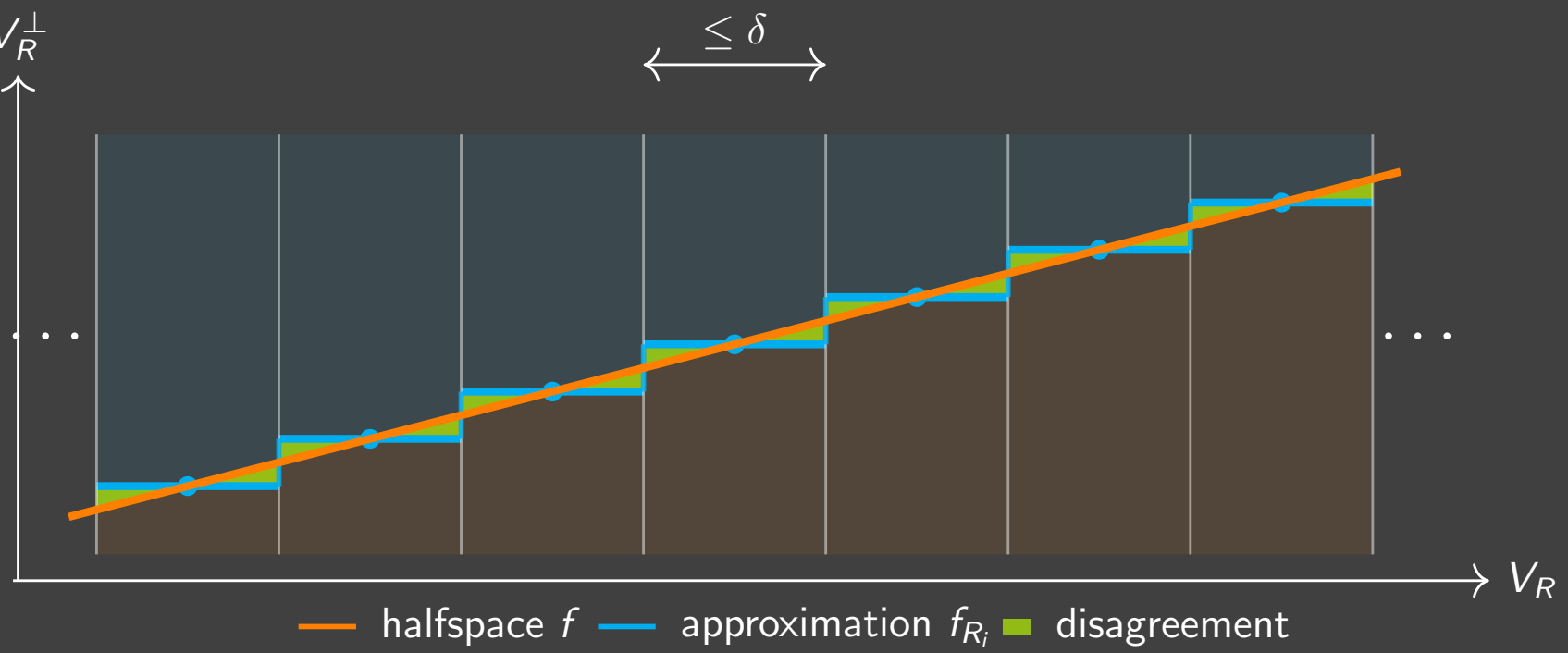
Fixing the Constrained Coordinates

Assume D_x is (q, σ)-smooth w.r.t. $\mathcal{N}(0, I)$.



Fixing the Constrained Coordinates

Assume D_x is (q, σ)-smooth w.r.t. $\mathcal{N}(0, I)$.

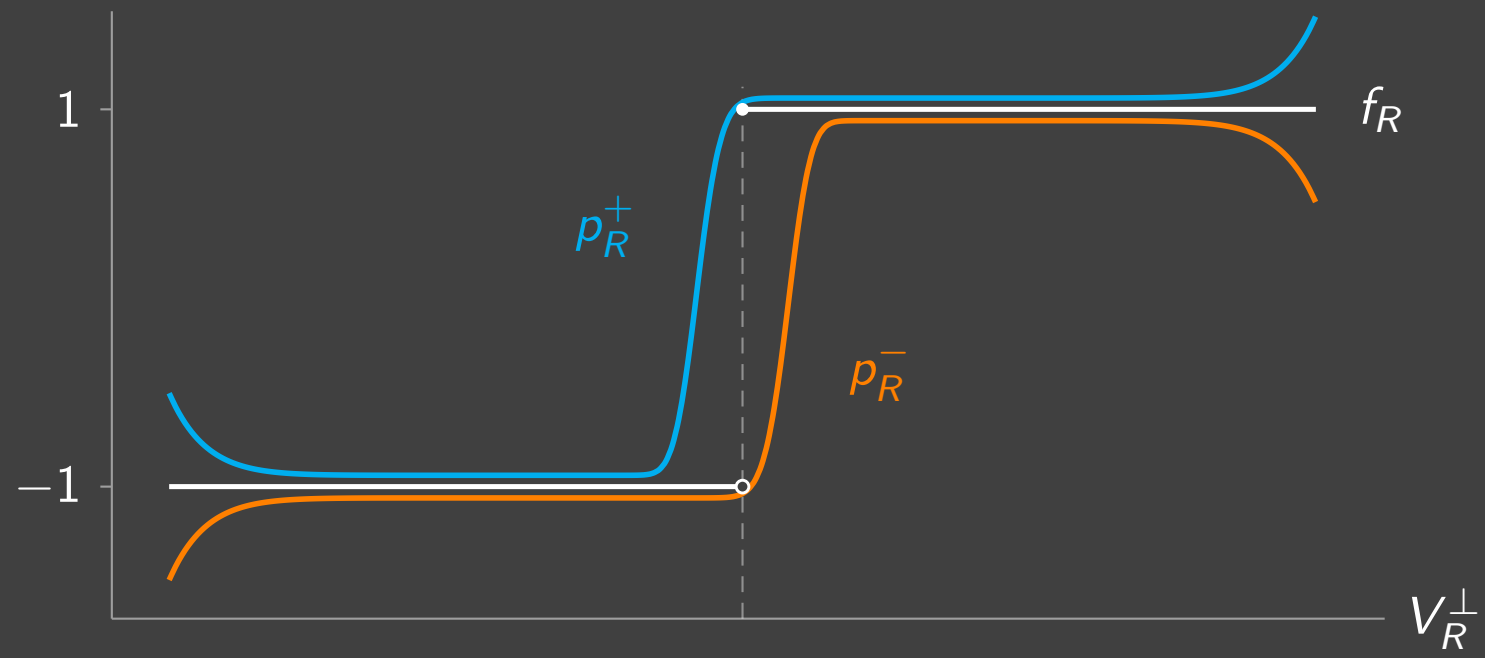


For general $\mathcal{F}$ with $\Gamma(\mathcal{F}) < \infty$ we have $\mathbf{Pr}_{D_x}[f(x) \neq g(x)] \leq \frac{(\delta\,\Gamma(\mathcal{F}))^{1/q}}{\sigma}$.
$\Gamma(\cdot)$ Gaussian surface area.
g fixes the V_R coordinates in each cell and varies only along $V_R^\perp$.

Cellwise Regression

If the function class is closed under affine restrictions, $f_R = f(x_{V_R^\perp} + z_R) \in \mathcal{F}$.
1. Gaussian sandwiching + moment matching. On each good cell R, there are degree-$\ell(\tau)$ polynomials $p_R^\pm$ satisfying

$$p_R^- \le g \le p_R^+, \qquad \mathbb{E}_{x\sim D_x|R}[p_R^+(x) - p_R^-(x)] \le O(\tau).$$



Cellwise Regression

If the function class is closed under affine restrictions, $f_R = f(x_{V_R^\perp} + z_R) \in \mathcal{F}$.

1. Gaussian sandwiching + moment matching. On each good cell R, there are degree-$\ell(\tau)$ polynomials $p_R^\pm$ satisfying

$$p_R^- \le g \le p_R^+, \qquad \mathbb{E}_{x\sim D_x|R}[p_R^+(x) - p_R^-(x)] \le O(\tau).$$

2. Regress on the observed labels. Degree-$\ell(\tau)$ L_1 regression and thresholding produce h_R with

$$\begin{aligned}\Pr_D[h_R(x) \neq y \mid R] &\le \frac{1}{2}\mathbb{E}_D[|y - p_R^+(x)| \mid R] + O(\tau) \\ &\le \Pr_D[g(x) \neq y \mid R] + O(\tau).\end{aligned}$$

3. Combine the cells. Choosing δ to be small and averaging over cells

$$\mathrm{err}_D(h) \le \mathrm{OPT} + O(\tau).$$

Cellwise Regression

If the function class is closed under affine restrictions, $f_R = f(x_{V_R^\perp} + z_R) \in \mathcal{F}$.

1. Gaussian sandwiching + moment matching. On each good cell R, there are degree-$\ell(\tau)$ polynomials $p_R^\pm$ satisfying

$$p_R^- \le g \le p_R^+, \qquad \mathbb{E}_{x\sim D_x|R}[p_R^+(x) - p_R^-(x)] \le O(\tau).$$

2. Regress on the observed labels. Degree-$\ell(\tau)$ L_1 regression and thresholding produce h_R with

$$\begin{aligned}\Pr_D[h_R(x) \neq y \mid R] &\le \frac{1}{2}\mathbb{E}_D[|y - p_R^+(x)| \mid R] + O(\tau) \\ &\le \Pr_D[g(x) \neq y \mid R] + O(\tau).\end{aligned}$$

3. Combine the cells. Choosing δ to be small and averaging over cells

$$\mathrm{err}_D(h) \le \mathrm{OPT} + O(\tau).$$

If $\mathrm{OPT} < 1/2 - c$, choosing $\tau = O(c)$ gives a weak learner.

Agnostic Learning Result

A weak learner for every $D' \leq D/\epsilon$ satisfying $\mathrm{OPT}_{D'} \leq 1/2 - c$ can be boosted to $(1+c)\mathrm{OPT}_D + \epsilon$ error [Diakonikolas–Kane–Ma, FOCS 2026].

Agnostic Learning Result

A weak learner for every $D' \le D/\epsilon$ satisfying $\mathrm{OPT}_{D'} \le 1/2 - c$ can be boosted to $(1+c)\mathrm{OPT}_D + \epsilon$ error [Diakonikolas–Kane–Ma, FOCS 2026].

Theorem (Agnostic Learning via Gaussian Partitioning)

Let $\mathcal{F}$ be a Boolean K-MIM class:

- *Closed under affine restrictions*
- *Bounded Gaussian surface area*
- *Gaussian sandwiching degree $\ell(\cdot)$*

Let D_x (q,σ)-smooth w.r.t. $\mathcal{N}(0,I)$. Let $c, \epsilon \in (0, 1/2)$.
There is an algorithm using $n = d^{O(\ell(c))} \exp\big(\mathrm{polylog}_{\ell(c),q,K}(1/(\sigma\epsilon))\big)$ samples, running in $\mathrm{poly}(n)$ time, and returning $\widehat{h}$ such that, w.h.p.

$$\mathrm{err}_D(\widehat{h}) \le (1+c) \inf_{f \in \mathcal{F}} \mathrm{err}_D(f) + \epsilon.$$

Agnostic Learning Result

A weak learner for every $D' \le D/\epsilon$ satisfying $\mathrm{OPT}_{D'} \le 1/2 - c$ can be boosted to $(1+c)\mathrm{OPT}_D + \epsilon$ error [Diakonikolas–Kane–Ma, FOCS 2026].

Theorem (Agnostic Learning via Gaussian Partitioning)

Let $\mathcal{F}$ be a Boolean K-MIM class:

- *Closed under affine restrictions*
- *Bounded Gaussian surface area*
- *Gaussian sandwiching degree $\ell(\cdot)$*

Let D_x (q,σ)-smooth w.r.t. $\mathcal{N}(0,I)$. Let $c, \epsilon \in (0, 1/2)$.
There is an algorithm using $n = d^{O(\ell(c))} \exp\big(\mathrm{polylog}_{\ell(c),q,K}(1/(\sigma\epsilon))\big)$ samples, running in $\mathrm{poly}(n)$ time, and returning $\widehat{h}$ such that, w.h.p.

$$\mathrm{err}_D(\widehat{h}) \le (1+c) \inf_{f \in \mathcal{F}} \mathrm{err}_D(f) + \epsilon.$$

The exponent of d is the Gaussian sandwiching degree.
An exponential improvement over the logconcave bound.

Agnostic Learning Result

A weak learner for every $D' \le D/\epsilon$ satisfying $\mathrm{OPT}_{D'} \le 1/2 - c$ can be boosted to $(1+c)\mathrm{OPT}_D + \epsilon$ error [Diakonikolas–Kane–Ma, FOCS 2026].

Theorem (Agnostic Learning via Gaussian Partitioning)

Let $\mathcal{F}$ be a Boolean K-MIM class:

- *Closed under affine restrictions*
- *Bounded Gaussian surface area*
- *Gaussian sandwiching degree $\ell(\cdot)$*

Let D_x (q,σ)-smooth w.r.t. $\mathcal{N}(0,I)$. Let $c, \epsilon \in (0, 1/2)$.
There is an algorithm using $n = d^{O(\ell(c))} \exp\big(\mathrm{polylog}_{\ell(c),q,K}(1/(\sigma\epsilon))\big)$ samples, running in $\mathrm{poly}(n)$ time, and returning $\widehat{h}$ such that, w.h.p.

$$\mathrm{err}_D(\widehat{h}) \le (1+c) \inf_{f \in \mathcal{F}} \mathrm{err}_D(f) + \epsilon.$$

We prove a $d^{\Omega(\ell_1(c))}$ SQ lower bound, where $\ell_1(\cdot)$ is the Gaussian L_1 approx. degree.

Agnostic Learning Results

Function class	**Dimension dependence**
Intersections of K halfspaces	$d^{\widetilde{O}(K^3/c^4)}$
Boolean functions of K halfspaces	$d^{\widetilde{O}(K^5/c^4)}$
Convex concepts of intrinsic dimension K	$d^{\widetilde{O}(K^5/c^4)}$
Degree-r PTFs of intrinsic dimension K	$d^{\widetilde{O}(r^6K^5/c^4)}$

For the two halfspace classes, the same dimension bounds hold for testable agnostic learning.

PUI Learning Model

Unknown: A target $f \in \mathcal{F}$ and an input distribution D^* that is (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$.

Given:

- Positive examples from $D^*_+ := D^* \mid (f(x) = 1)$.
- Unlabeled examples from the reference $\mathcal{N}(0, I_d)$.
- The parameters q, σ and $p_+ > 0$ s.t. $D^*(f = 1) \geq p_+ > 0$.

PUI Learning Model

Unknown: A target $f \in \mathcal{F}$ and an input distribution D^* that is (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$.

Given:

- Positive examples from $D^*_+ := D^* \mid (f(x) = 1)$.
- Unlabeled examples from the reference $\mathcal{N}(0, I_d)$.
- The parameters q, σ and $p_+ > 0$ s.t. $D^*(f = 1) \geq p_+ > 0$.

Goal: Output a hypothesis h such that, $\Pr_{x \sim D^*}[h(x) \neq f(x)] \leq \epsilon$.

PUI Learning Model

Unknown: A target $f \in \mathcal{F}$ and an input distribution D^* that is (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$.

Given:

- Positive examples from $D^*_+ := D^* \mid (f(x) = 1)$.
- Unlabeled examples from the reference $\mathcal{N}(0, I_d)$.
- The parameters q, σ and $p_+ > 0$ s.t. $D^*(f = 1) \geq p_+ > 0$.

Goal: Output a hypothesis h such that, $\Pr_{x \sim D^*}[h(x) \neq f(x)] \leq \epsilon$.

Introduced by [Lee–Mehrotra–Zampetakis, STOC 2026] as a generalization of the PU learning model.

$d^{O(\ell_1((\sigma p_+ \epsilon)^{\mathrm{poly}(q)}))}$, $\ell_1(\cdot)$ the Gaussian L_1 approximation degree.

Example: $d^{O\left(\log(K)(1/(\sigma p_+ \epsilon))^{\mathrm{poly}(q)}\right)}$ for intersections of K halfspaces.

Partitioning the Positive Examples

Apply Gaussian partitioning to $Q := D_+^*$, which is $(q, \sigma p_+)$-smooth w.r.t. $\mathcal{N}(0, I_d)$. Let $\mathcal{G}$ be the good cells and $S := \bigcup_{R \in \mathcal{G}} R$. Then $Q(S) \geq 1 - \eta$, and for every $R \in \mathcal{G}$:

- Thinness: $\mathrm{diam}(\mathrm{Proj}_{V_R} R) \leq \delta$.
- Moment matching: $Q \mid R$ matches m moments with $\mathcal{N}(0, I_{V_R^\perp})$, to slack Δ.

Partitioning the Positive Examples

Apply Gaussian partitioning to $Q := D_+^*$, which is $(q, \sigma p_+)$-smooth w.r.t. $\mathcal{N}(0, I_d)$. Let $\mathcal{G}$ be the good cells and $S := \bigcup_{R \in \mathcal{G}} R$. Then $Q(S) \geq 1 - \eta$, and for every $R \in \mathcal{G}$:

- Thinness: $\mathrm{diam}(\mathrm{Proj}_{V_R} R) \leq \delta$.
- Moment matching: $Q \mid R$ matches m moments with $\mathcal{N}(0, I_{V_R^\perp})$, to slack Δ.

Label $x \notin S$ by -1. The false negative rate satisfies:

$$\Pr_{x \sim D^*}[f(x) = 1,\ x \notin S] = D^*(f = 1)\, Q(S^c) \leq \eta.$$

Partitioning the Positive Examples

Apply Gaussian partitioning to $Q := D_+^*$, which is $(q, \sigma p_+)$-smooth w.r.t. $\mathcal{N}(0, I_d)$. Let $\mathcal{G}$ be the good cells and $S := \bigcup_{R \in \mathcal{G}} R$. Then $Q(S) \geq 1 - \eta$, and for every $R \in \mathcal{G}$:

- Thinness: $\mathrm{diam}(\mathrm{Proj}_{V_R} R) \leq \delta$.
- Moment matching: $Q \mid R$ matches m moments with $\mathcal{N}(0, I_{V_R^\perp})$, to slack Δ.

Label $x \notin S$ by -1. The false negative rate satisfies:

$$\Pr_{x \sim D^*}[f(x) = 1,\ x \notin S] = D^*(f = 1)\, Q(S^c) \leq \eta.$$

On each $R \in \mathcal{G}$, positive examples match Gaussian moments along $V_R^\perp$.
Can we safely label the whole cell $+1$?

Moment Matching for Intersections

Lemma (Structural Result Intersections of Halfspaces)

Let $\epsilon \in (0, 1/2)$, f be an intersection of K halfspaces with $\Pr_{\mathcal{N}(0,I_d)}[f = -1] \geq \epsilon$. Then any distribution Q supported on $\{x : f(x) = 1\}$ cannot match the first $O(\log(K/\epsilon))$ moments of $\mathcal{N}(0, I_d)$ to slack less than $1/4$.

However, we have moment matching for $\mathrm{Proj}_{V_R^{\perp}}(Q|x \in R, f(x))$.

Moment Matching for Intersections

Lemma (Structural Result Intersections of Halfspaces)

Let $\epsilon \in (0, 1/2)$, f be an intersection of K halfspaces with $\Pr_{\mathcal{N}(0,I_d)}[f = -1] \geq \epsilon$. Then any distribution Q supported on $\{x : f(x) = 1\}$ cannot match the first $O(\log(K/\epsilon))$ moments of $\mathcal{N}(0, I_d)$ to slack less than $1/4$.

However, we have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f(x))$.

Sketch:

- Thinness allows us to fix the V_R coordinates. Have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f_R(x)) = \mathrm{Proj}_{V_R^\perp}(Q|x \in R)|f_R(x)$.

Moment Matching for Intersections

Lemma (Structural Result Intersections of Halfspaces)

Let $\epsilon \in (0, 1/2)$, f be an intersection of K halfspaces with $\Pr_{\mathcal{N}(0,I_d)}[f = -1] \geq \epsilon$. Then any distribution Q supported on $\{x : f(x) = 1\}$ cannot match the first $O(\log(K/\epsilon))$ moments of $\mathcal{N}(0, I_d)$ to slack less than $1/4$.

However, we have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f(x))$.

Sketch:

- Thinness allows us to fix the V_R coordinates. Have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f_R(x)) = \mathrm{Proj}_{V_R^\perp}(Q|x \in R)|f_R(x)$.
- Apply the structural result on $\mathrm{Proj}_{V_R^\perp}(Q|x \in R)$ and f_R.

Moment Matching for Intersections

Lemma (Structural Result Intersections of Halfspaces)

Let $\epsilon \in (0, 1/2)$, f be an intersection of K halfspaces with $\Pr_{\mathcal{N}(0,I_d)}[f = -1] \geq \epsilon$. Then any distribution Q supported on $\{x : f(x) = 1\}$ cannot match the first $O(\log(K/\epsilon))$ moments of $\mathcal{N}(0, I_d)$ to slack less than $1/4$.

However, we have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f(x))$.

Sketch:

- Thinness allows us to fix the V_R coordinates. Have moment matching for $\mathrm{Proj}_{V_R^\perp}(Q|x \in R, f_R(x)) = \mathrm{Proj}_{V_R^\perp}(Q|x \in R)|f_R(x)$.
- Apply the structural result on $\mathrm{Proj}_{V_R^\perp}(Q|x \in R)$ and f_R.

Averaging over good cells bounds the false positive rate by ϵ.

Gaussian PUI Exponent

Definition

The Gaussian PUI exponent $m^*(\epsilon,\Delta,\sigma,q)$ is the smallest $m \in \mathbb{Z}_+$ s.t., for every $f \in \mathcal{F}$ and every D^* that is (q,σ)-smooth w.r.t. $\mathcal{N}(0,I)$ with $D^*(f=1)>0$:
If $D^*_+ := D^*|(f=1)$ matches m mom. to err. Δ with $\mathcal{N}(0,I)$, then $D^*(f=-1) < \epsilon$.

Gaussian PUI Exponent

Definition

The Gaussian PUI exponent $m^*(\epsilon,\Delta,\sigma,q)$ is the smallest $m \in \mathbb{Z}_+$ s.t., for every $f \in \mathcal{F}$ and every D^* that is (q,σ)-smooth w.r.t. $\mathcal{N}(0,I)$ with $D^*(f=1)>0$:
If $D^*_+ := D^*|(f=1)$ matches m mom. to err. Δ with $\mathcal{N}(0,I)$, then $D^*(f=-1) < \epsilon$.

For intersections of K halfspaces, for any $\Delta < 1/4$, $m^*(\epsilon,\Delta,\sigma,q) = O\left(\log \frac{2K}{(\sigma\epsilon)^q}\right)$.

General PUI Learning Result

Let $\mathcal{F}$ closed under affine restrictions with bounded Gaussian surface area.

Theorem (Gaussian PUI Learning)

Let $f \in \mathcal{F}$, let D^* (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$, and let $p_+ > 0$ be a known lower bound on $D^*(f = 1)$. For $\epsilon, \Delta \in (0, 1/4)$, set $m = m^*(\epsilon, \Delta, \sigma\epsilon p_+, q)$. There is an algorithm using $n = d^m \exp\left(\mathrm{polylog}_{m,q,\Delta} \frac{1}{\sigma\epsilon p_+}\right)$ samples from $D^*_+ = D^* \mid (f = 1)$ and $\mathrm{poly}(n)$ time that returns h satisfying

$$\Pr_{x \sim D^*}[h(x) \neq f(x)] \leq \epsilon.$$

General PUI Learning Result

Let $\mathcal{F}$ closed under affine restrictions with bounded Gaussian surface area.

Theorem (Gaussian PUI Learning)

Let $f \in \mathcal{F}$, let D^* (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$, and let $p_+ > 0$ be a known lower bound on $D^*(f = 1)$. For $\epsilon, \Delta \in (0, 1/4)$, set $m = m^*(\epsilon, \Delta, \sigma\epsilon p_+, q)$. There is an algorithm using $n = d^m \exp\left(\mathrm{polylog}_{m,q,\Delta} \frac{1}{\sigma\epsilon p_+}\right)$ samples from $D^*_+ = D^* \mid (f = 1)$ and $\mathrm{poly}(n)$ time that returns h satisfying

$$\Pr_{x \sim D^*}[h(x) \neq f(x)] \leq \epsilon.$$

Concept Class	**Prior Work**	**This Work**
Intersections of k halfspaces	$d^{O(\log(2k) \cdot (1/(\sigma p_+ \epsilon))^{\mathrm{poly}(q)})}$	$d^{O(q \log(kq/(\sigma\epsilon p_+)))}$
Convex sets (intrinsic dim. K)	$d^{O(\sqrt{K} \cdot (1/(\sigma p_+ \epsilon))^{\mathrm{poly}(q)})}$	$d^{O(K + q \log(q/(\sigma\epsilon p_+)))}$

Dimension dependence.

General PUI Learning Result

Let $\mathcal{F}$ closed under affine restrictions with bounded Gaussian surface area.

Theorem (Gaussian PUI Learning)

Let $f \in \mathcal{F}$, let D^* (q, σ)-smooth w.r.t. $\mathcal{N}(0, I_d)$, and let $p_+ > 0$ be a known lower bound on $D^*(f = 1)$. For $\epsilon, \Delta \in (0, 1/4)$, set $m = m^*(\epsilon, \Delta, \sigma\epsilon p_+, q)$. There is an algorithm using $n = d^m \exp\left(\mathrm{polylog}_{m,q,\Delta} \frac{1}{\sigma\epsilon p_+}\right)$ samples from $D^*_+ = D^* \mid (f = 1)$ and $\mathrm{poly}(n)$ time that returns h satisfying

$$\Pr_{x \sim D^*}[h(x) \neq f(x)] \leq \epsilon.$$

Concept Class	Prior Work	This Work
Intersections of k halfspaces	$d^{O(\log(2k) \cdot (1/(\sigma p_+ \epsilon))^{\mathrm{poly}(q)})}$	$d^{O(q \log(kq/(\sigma\epsilon p_+)))}$
Convex sets (intrinsic dim. K)	$d^{O(\sqrt{K} \cdot (1/(\sigma p_+ \epsilon))^{\mathrm{poly}(q)})}$	$d^{O(K + q \log(q/(\sigma\epsilon p_+)))}$

Dimension dependence.

SQ algorithms require $d^{\Omega(m^*)}$ in general and $d^{\Omega(\log(K/\epsilon))}$ for intersections of K halfspaces.

Reliable Learning via PUI

Reliable learning was introduced by [Kalai–Kanade–Mansour, COLT 2009].

Theorem (Black-box reduction)

If $\mathcal{F}$ contains the constant function $+1$, a PUI learner for $\mathcal{F}$ yields a negative reliable learner for marginals (q, σ)-smooth with respect to $\mathcal{N}(0, I_d)$.

Intersections of K halfspaces	$d^{O(\log(K/\epsilon))}$
Convex concepts of intrinsic dimension K	$d^{O(K+\log(1/\epsilon))}$

Logconcave Partitioning

We have a known logconcave reference distribution P and samples from Q.

Moment discrepancy tester. In cell R, compare the moments of $Q \mid R$ and $P \mid R$ through degree m.

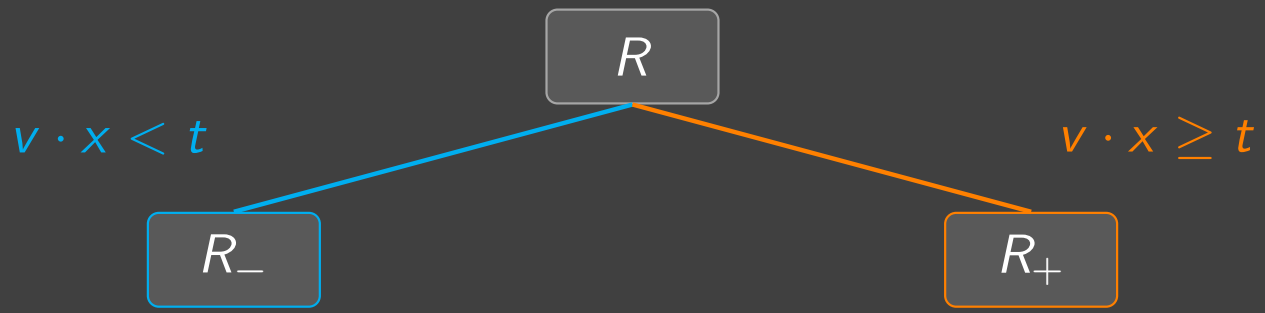
- If they agree within tolerance, certify R and stop splitting it.
- Otherwise, return a unit direction v with a moment gap.

Logconcave Partitioning

We have a known logconcave reference distribution P and samples from Q.

Moment discrepancy tester. In cell R, compare the moments of $Q \mid R$ and $P \mid R$ through degree m.

- If they agree within tolerance, certify R and stop splitting it.
- Otherwise, return a unit direction v with a moment gap.



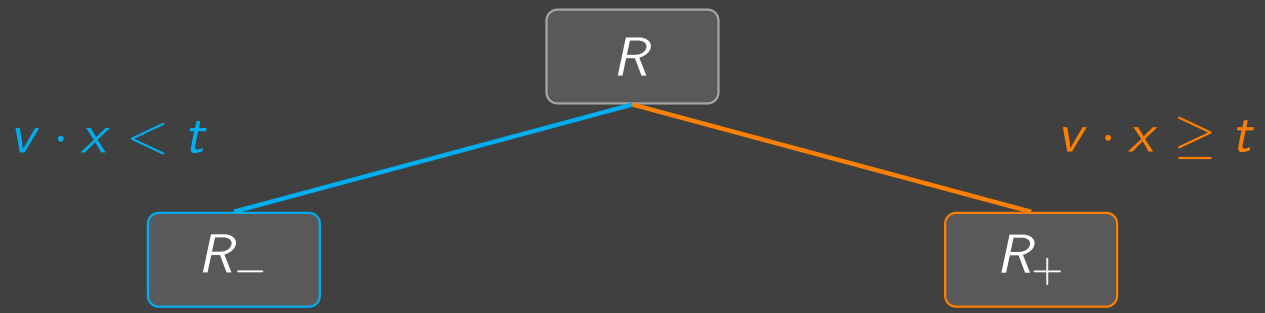
Sparsifying split. Use the gap along v to choose t with $P(R_- \mid R) \neq Q(R_- \mid R)$.

Logconcave Partitioning

We have a known logconcave reference distribution P and samples from Q.

Moment discrepancy tester. In cell R, compare the moments of $Q \mid R$ and $P \mid R$ through degree m.

- If they agree within tolerance, certify R and stop splitting it.
- Otherwise, return a unit direction v with a moment gap.



Sparsifying split. Use the gap along v to choose t with $P(R_- \mid R) \neq Q(R_- \mid R)$.

Controlling the partition. A lower bound on the mass of each child bounds the number of cells. A decreasing potential bounds the total Q-mass of the small cells.

From Logconcave to Gaussian Partitioning

Logconcave partitioning: Match the moments of $Q \mid R$ to those of the conditional reference $P \mid R$ in all directions.

From Logconcave to Gaussian Partitioning

Logconcave partitioning: Match the moments of $Q \mid R$ to those of the conditional reference $P \mid R$ in all directions.

For target is the standard Gaussian. A cut can make moment matching impossible. In a direction v along which R has width $\delta < 2$, $\mathrm{Var}_{x \sim Q|R}(v \cdot x) \leq \frac{\delta^2}{4} < 1$.

From Logconcave to Gaussian Partitioning

Logconcave partitioning: Match the moments of $Q \mid R$ to those of the conditional reference $P \mid R$ in all directions.

For target is the standard Gaussian. A cut can make moment matching impossible. In a direction v along which R has width $\delta < 2$, $\mathrm{Var}_{x \sim Q|R}(v \cdot x) \leq \frac{\delta^2}{4} < 1$.
We aim to match standard Gaussian moments on $V_R^{\perp}$.

From Logconcave to Gaussian Partitioning

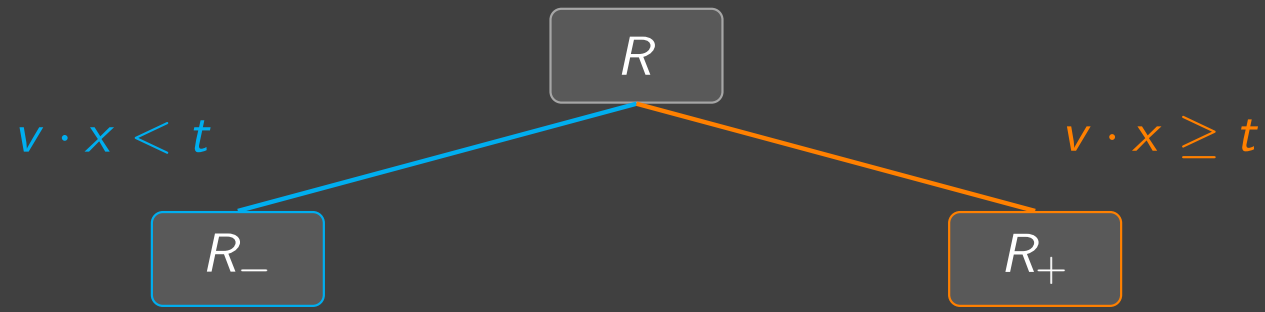
Logconcave partitioning: Match the moments of $Q \mid R$ to those of the conditional reference $P \mid R$ in all directions.

For target is the standard Gaussian. A cut can make moment matching impossible. In a direction v along which R has width $\delta < 2$, $\mathrm{Var}_{x \sim Q|R}(v \cdot x) \leq \frac{\delta^2}{4} < 1$.
We aim to match standard Gaussian moments on $V_R^{\perp}$.

Along the constrained directions in V_R, we refine to make the cells thin enough that we can fix those coordinates with small total error.

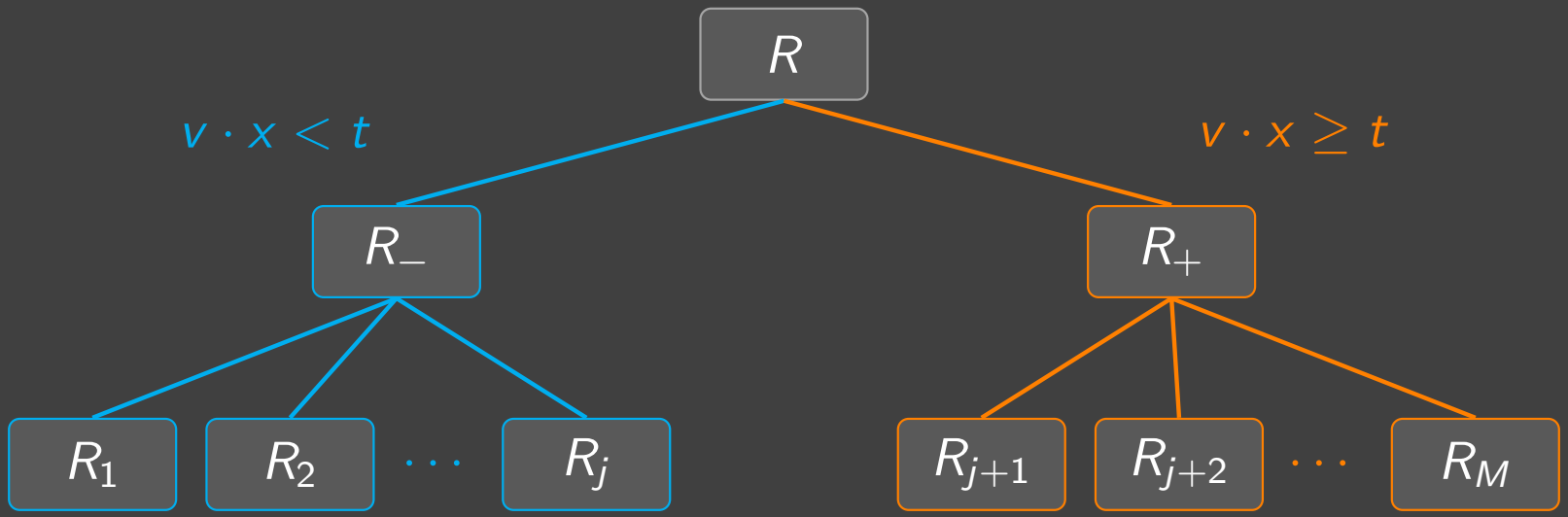
Gaussian Partitioning

For and a witness $v \in V_R^{\perp}$, make the same cut as in the logconcave case.



Gaussian Partitioning

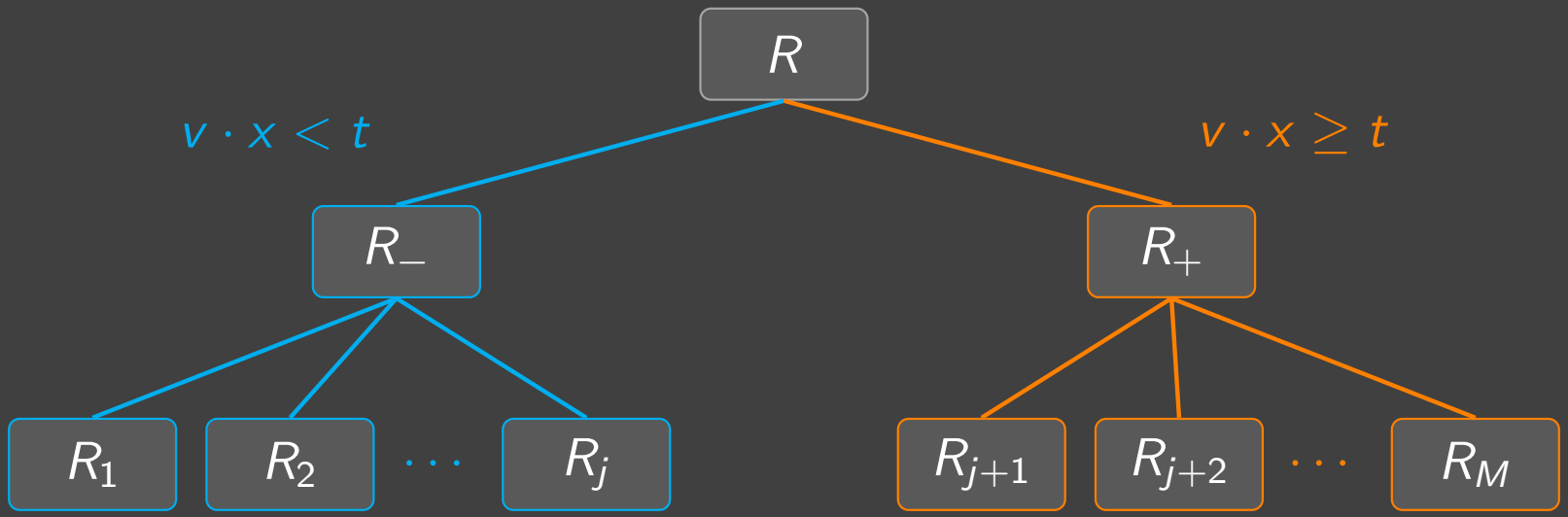
For and a witness $v \in V_R^{\perp}$, make the same cut as in the logconcave case.



Inside $|v \cdot x| \leq T$, use $M \leq 1 + \lceil 2T/\delta_0 \rceil$ slabs of width at most δ_0 along v.

Gaussian Partitioning

For and a witness $v \in V_R^{\perp}$, make the same cut as in the logconcave case.

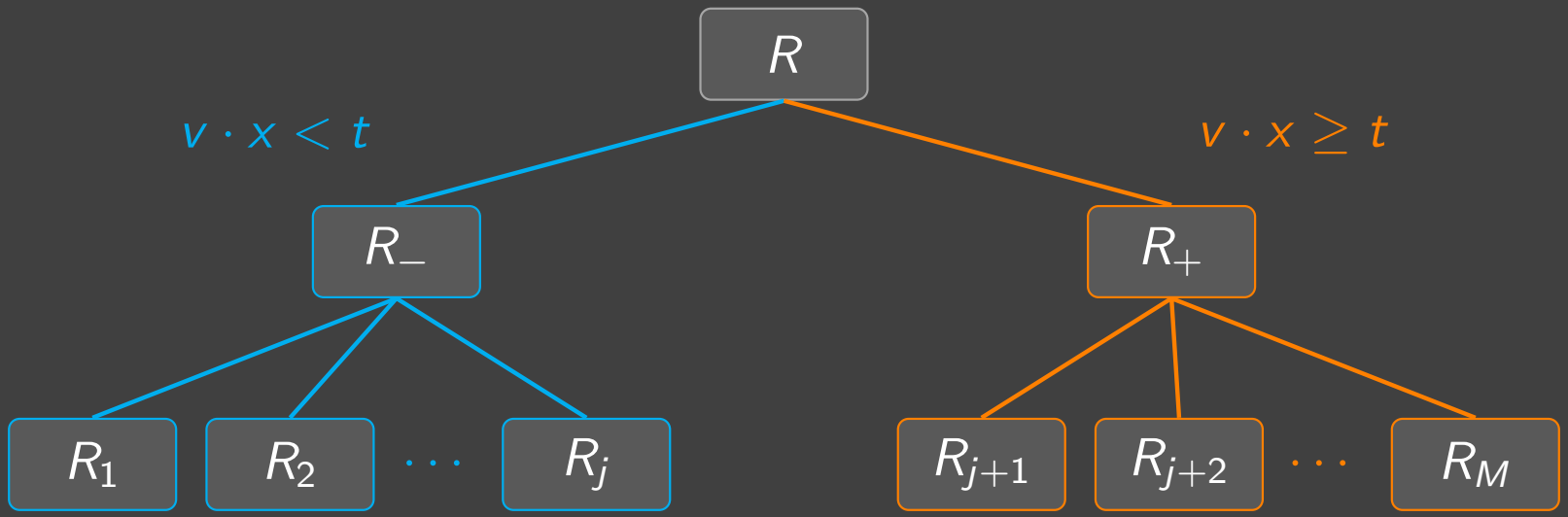


Inside $|v \cdot x| \le T$, use $M \le 1 + \lceil 2T/\delta_0 \rceil$ slabs of width at most δ_0 along v.

It suffices to bound the total Q-mass of tail cells, small-mass cells, and active cells at depth L.

Gaussian Partitioning

For and a witness $v \in V_R^{\perp}$, make the same cut as in the logconcave case.



Inside $|v \cdot x| \le T$, use $M \le 1 + \lceil 2T/\delta_0 \rceil$ slabs of width at most δ_0 along v.

It suffices to bound the total Q-mass of tail cells, small-mass cells, and active cells at depth L.

Choosing T sufficiently large and using the previous potential argument handle: tail cells, small-mass cells.
A depth bound controls mass still active at L and gives $\operatorname{diam}(R) \le \delta_0 \sqrt{L}$.

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.

↑ Hellinger affinity

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.
If $R \in \mathcal{A}_t$ is split to $R_\pm$ s.t. $|P(R_- \mid R) - Q(R_- \mid R)| > \gamma$, then
$\Phi(R_\pm|R) := \sqrt{P(R_-|R)Q(R_-|R)} + \sqrt{P(R_+|R)Q(R_+|R)} \leq e^{-\gamma^2/2}$.

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.
If $R \in \mathcal{A}_t$ is split to $R_\pm$ s.t. $|P(R_- \mid R) - Q(R_- \mid R)| > \gamma$, then
$\Phi(R_\pm|R) := \sqrt{P(R_-|R)Q(R_-|R)} + \sqrt{P(R_+|R)Q(R_+|R)} \leq e^{-\gamma^2/2}$.
Refining to $R_\pm$ to $\{R_i\}_{i\in[M]}$ gives

$$\Phi(\{R_i\}_{i\in[M]}|R) = \sum_{i\in[M]} \sqrt{Q(R_i|R)P(R_i|R)} \overset{\text{C–S}}{\leq} \Phi(R_\pm|R)$$

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.
If $R \in \mathcal{A}_t$ is split to $R_\pm$ s.t. $|P(R_- \mid R) - Q(R_- \mid R)| > \gamma$, then
$\Phi(R_\pm|R) := \sqrt{P(R_-|R)Q(R_-|R)} + \sqrt{P(R_+|R)Q(R_+|R)} \le e^{-\gamma^2/2}$.
Refining to $R_\pm$ to $\{R_i\}_{i\in[M]}$ gives

$$\Phi(\{R_i\}_{i\in[M]}|R) = \sum_{i\in[M]} \sqrt{Q(R_i|R)P(R_i|R)} \overset{\text{C–S}}{\le} \Phi(R_\pm|R)$$

$$\sum_{i\in[M]} \sqrt{P(R_i|R)Q(R_i|R)} = \Phi(\{R_i\}_{i\in[M]}|R)\sqrt{P(R)Q(R)} \le e^{-\gamma^2/2}\sqrt{P(R)Q(R)}$$

Summing $R \in \mathcal{A}_t$ gives $\Phi_{t+1} \le e^{-\gamma^2/2}\Phi_t$, $\Phi_0 \le 1$.

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.
If $R \in \mathcal{A}_t$ is split to $R_\pm$ s.t. $|P(R_- \mid R) - Q(R_- \mid R)| > \gamma$, then
$\Phi(R_\pm|R) := \sqrt{P(R_-|R)Q(R_-|R)} + \sqrt{P(R_+|R)Q(R_+|R)} \le e^{-\gamma^2/2}$.
Refining to $R_\pm$ to $\{R_i\}_{i\in[M]}$ gives

$$\Phi(\{R_i\}_{i\in[M]}|R) = \sum_{i\in[M]} \sqrt{Q(R_i|R)P(R_i|R)} \overset{\text{C–S}}{\le} \Phi(R_\pm|R)$$

$$\sum_{i\in[M]} \sqrt{P(R_i|R)Q(R_i|R)} = \Phi(\{R_i\}_{i\in[M]}|R)\sqrt{P(R)Q(R)} \le e^{-\gamma^2/2}\sqrt{P(R)Q(R)}$$

Summing $R \in \mathcal{A}_t$ gives $\Phi_{t+1} \le e^{-\gamma^2/2}\Phi_t$, $\Phi_0 \le 1$.
Since $Q(R) \le P(R)/\sigma$, $Q(\mathcal{A}_t) \le (1/\sqrt{\sigma}) \sum_{R\in\mathcal{A}_t} \sqrt{P(R)Q(R)} \le e^{-\gamma^2 t/2}/\sqrt{\sigma}$.

Depth bound

For $q = 1$, let $\mathcal{A}_t$ be the active cells at depth t. It holds that $Q(\mathcal{A}_t) \lesssim e^{-\Omega(t)}$.

Proof Sketch: Define the potential function $\Phi_t = \sum_{R \in \mathcal{A}_t} \sqrt{P(R)Q(R)}$.
If $R \in \mathcal{A}_t$ is split to $R_\pm$ s.t. $|P(R_- \mid R) - Q(R_- \mid R)| > \gamma$, then
$\Phi(R_\pm|R) := \sqrt{P(R_-|R)Q(R_-|R)} + \sqrt{P(R_+|R)Q(R_+|R)} \le e^{-\gamma^2/2}$.
Refining to $R_\pm$ to $\{R_i\}_{i\in[M]}$ gives

$$\Phi(\{R_i\}_{i\in[M]}|R) = \sum_{i\in[M]} \sqrt{Q(R_i|R)P(R_i|R)} \overset{\text{C–S}}{\le} \Phi(R_\pm|R)$$

$$\sum_{i\in[M]} \sqrt{P(R_i|R)Q(R_i|R)} = \Phi(\{R_i\}_{i\in[M]}|R)\sqrt{P(R)Q(R)} \le e^{-\gamma^2/2}\sqrt{P(R)Q(R)}$$

Summing $R \in \mathcal{A}_t$ gives $\Phi_{t+1} \le e^{-\gamma^2/2}\Phi_t$, $\Phi_0 \le 1$.
Since $Q(R) \le P(R)/\sigma$, $Q(\mathcal{A}_t) \le (1/\sqrt{\sigma}) \sum_{R\in\mathcal{A}_t} \sqrt{P(R)Q(R)} \le e^{-\gamma^2 t/2}/\sqrt{\sigma}$.
For $L \ge \frac{2}{\gamma^2} \log(\frac{1}{\eta\sqrt{\sigma}})$, $Q(\mathcal{A}_L) \le \eta$.

Depth Bound

$$\boxed{\text{For } q = 1, \text{ if } L \geq \frac{2}{\gamma^2} \log\left(\frac{1}{\eta\sqrt{\sigma}}\right), \text{ then } Q(\mathcal{A}_L) \leq \eta.}$$

- Quasipolynomial cell count: Each refinement step produces at most polynomially in η and σ many cells, and the depth is logarithimc. Hence, the total number of cells is quasipolynomial.

Depth Bound

For $q = 1$, if $L \geq \frac{2}{\gamma^2} \log\left(\frac{1}{\eta\sqrt{\sigma}}\right)$, then $Q(\mathcal{A}_L) \leq \eta$.

- Quasipolynomial cell count: Each refinement step produces at most polynomially in η and σ many cells, and the depth is logarithimc. Hence, the total number of cells is quasipolynomial.
- Depth bound $q \geq 1$: Smoothness gives $P(R) \geq \sigma' Q(R)$ outside cells of total Q-mass at most $\eta/8$, with $\sigma' = \sigma^q(\eta/8)^{q-1}$.

Depth Bound

$$\boxed{\text{For } q = 1, \text{ if } L \geq \frac{2}{\gamma^2} \log\left(\frac{1}{\eta\sqrt{\sigma}}\right), \text{ then } Q(\mathcal{A}_L) \leq \eta.}$$

- Quasipolynomial cell count: Each refinement step produces at most polynomially in η and σ many cells, and the depth is logarithimc. Hence, the total number of cells is quasipolynomial.
- Depth bound $q \geq 1$: Smoothness gives $P(R) \geq \sigma' Q(R)$ outside cells of total Q-mass at most $\eta/8$, with $\sigma' = \sigma^q(\eta/8)^{q-1}$.
- Moment Tester: The tester also needs smoothness control. But if Q is smooth w.r.t. $\mathcal{N}(0, I_d)$, $Q|R$ may not be. We show that this holds for almost all of the Q-mass.

Open Problems

Open Problem 1: Other Reference Distributions

Extend our results to other reference distributions whose approximation rates improve on the worst-case logconcave rate.

Open Problem 2: PTAS Goal

Improve the quasipolynomial dependence on $1/\epsilon$ to polynomial.

Open Problems

Open Problem 1: Other Reference Distributions

Extend our results to other reference distributions whose approximation rates improve on the worst-case logconcave rate.

Open Problem 2: PTAS Goal

Improve the quasipolynomial dependence on $1/\epsilon$ to polynomial.

Thank you!