Testing and Learning Convex Sets in the Ternary Hypercube

ITCS 24

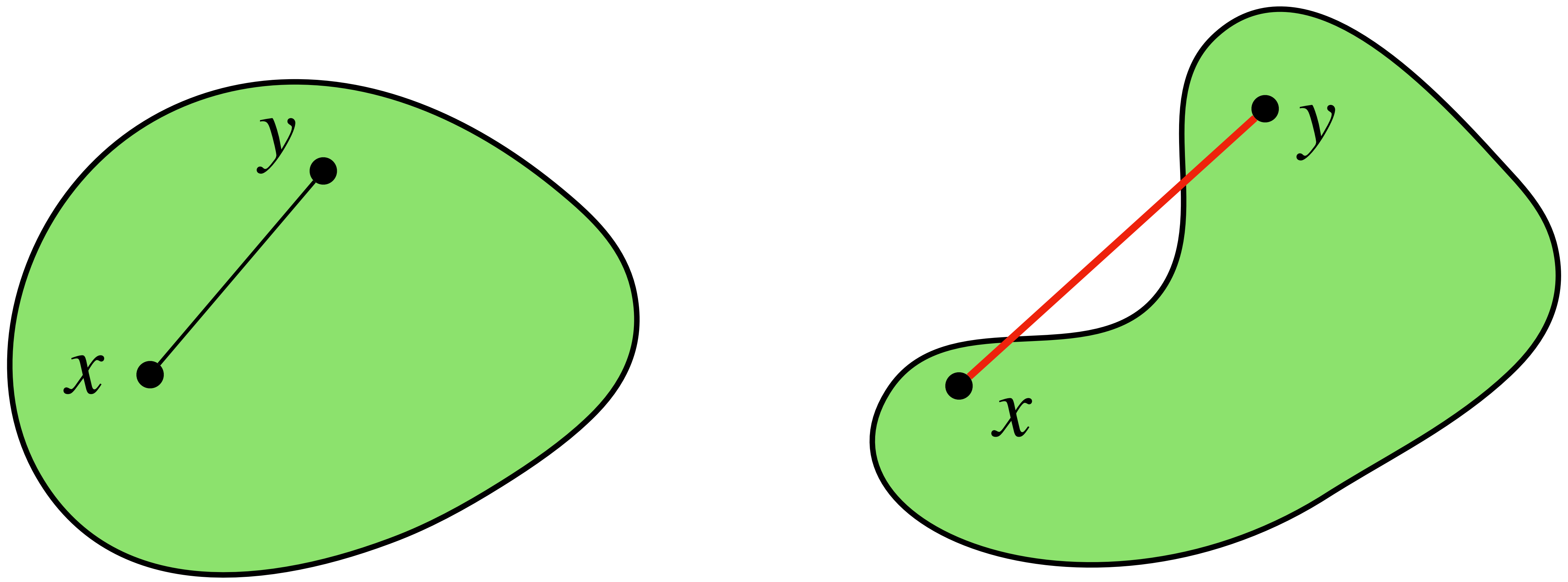


Hadley Black
UCSD

Eric Blais
University of Waterloo

Nathan Harms
EPFL

Convexity Testing



Def: $S \subseteq \mathbb{R}^n$ **convex** iff
$\forall x, y \in S, \lambda \in [0,1]: \lambda x + (1-\lambda)y \in S$

- Measure μ (Gaussian, uniform over $[-1,1]^n$, …)
- **Membership oracle access**: is $x \in S$?
- x given either as **samples** from μ, or chosen via **queries**

Distance to convexity

$$\varepsilon(S) = \min_{C \text{ convex}} \mu(S \Delta C)$$

Testing

Given S and $\varepsilon > 0$…

1. if S convex: **accept** w.p. $> 2/3$
2. if $\varepsilon(S) > \varepsilon$: **reject** w.p. $> 2/3$

$<_R$

Learning

Given convex S and $\varepsilon > 0$…

Output T such that $\mu(S \Delta T) \leq \varepsilon$ w.p. $> 2/3$

Question

How hard is it to test convexity with **queries** in n dimensions?

Prior Work on Convexity Testing

- Klivans-O'Donnell-Servedio [08]: $2^{\widetilde{O}(n^{1/2})}$ for **learning** with **samples** under Gaussian
- Chen-Freilich-Servedio-Sun [17]: $2^{\widetilde{\Theta}(n^{1/2})}$ for **testing** with **samples** under Gaussian
- Schmeltz [92], Raskhodnikova [03], Berman-Murzabulatov-Raskhodnikova [19],[19],[22]:
  Testing and learning over $[m]^2$ and $[0,1]^2$

What about queries in high dimensions?

- Rademacher-Vempala [04], Blais-Bommireddi [20]: testers that spot check for violations require $2^{\Omega(n)}$ queries

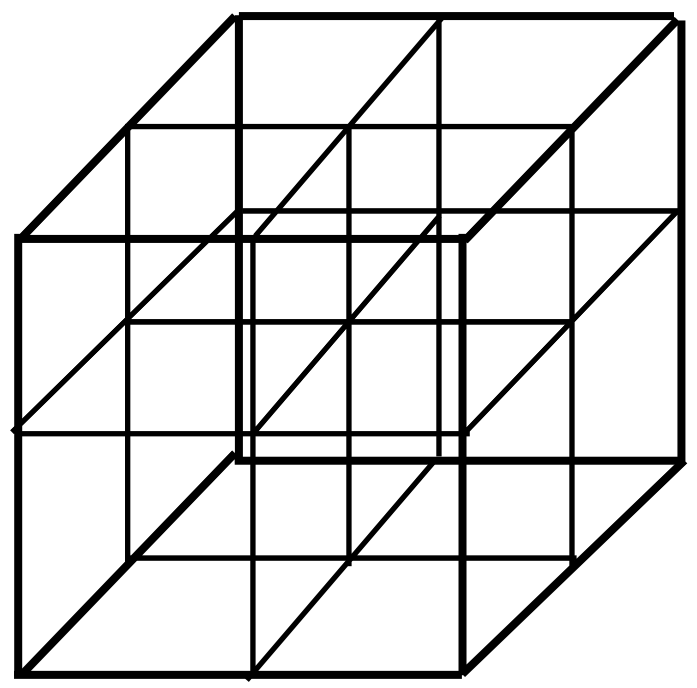
???

Query-based high-dimensional convexity testing is a wide open problem

The Ternary Hypercube

Black-Blais-Harms **[ITCS 24]**

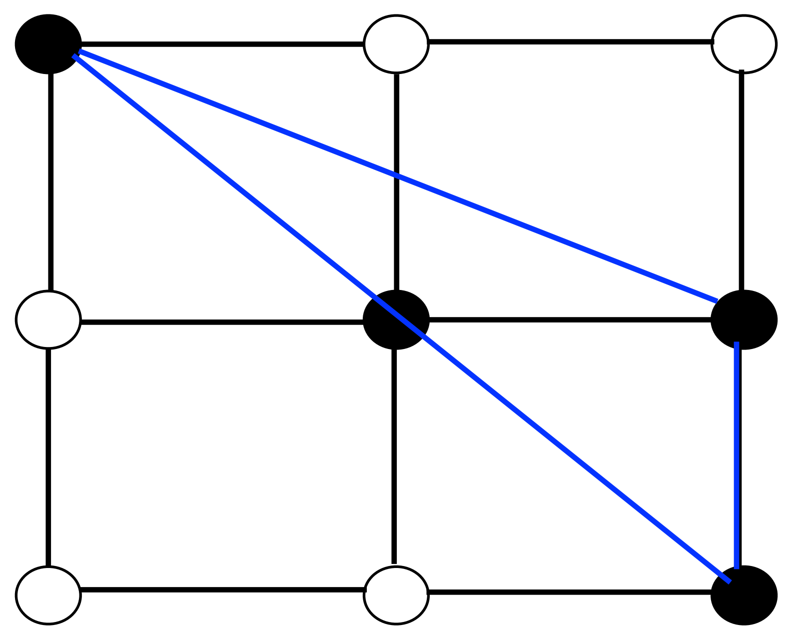
- We consider sets in the **ternary hypercube** $\{-1,0,1\}^n$



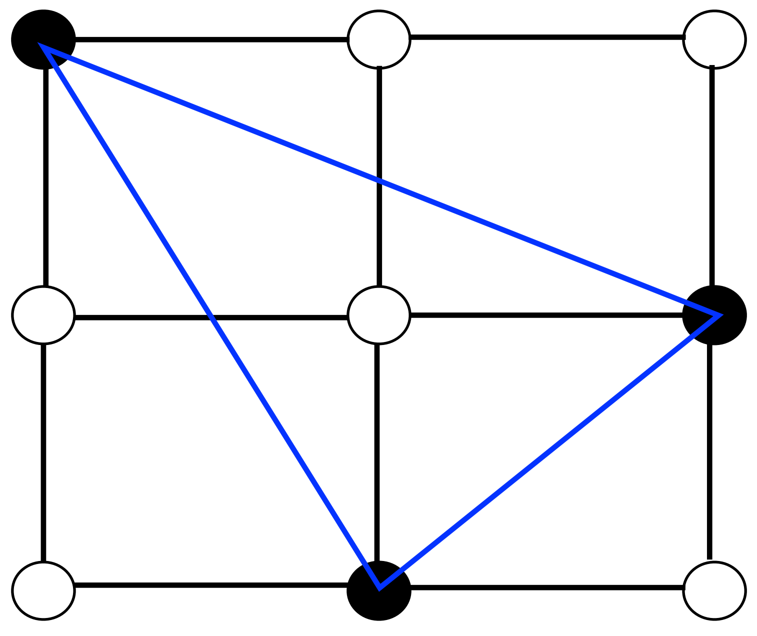
Def: $S \subseteq \{0, \pm 1\}^n$ **convex** if
$S = \text{Conv}(S) \cap \{0, \pm 1\}^n$

Distance to convexity

$$\varepsilon(S) = \min_{C \text{ convex}} 3^{-n} |S \Delta C|$$



convex



not convex

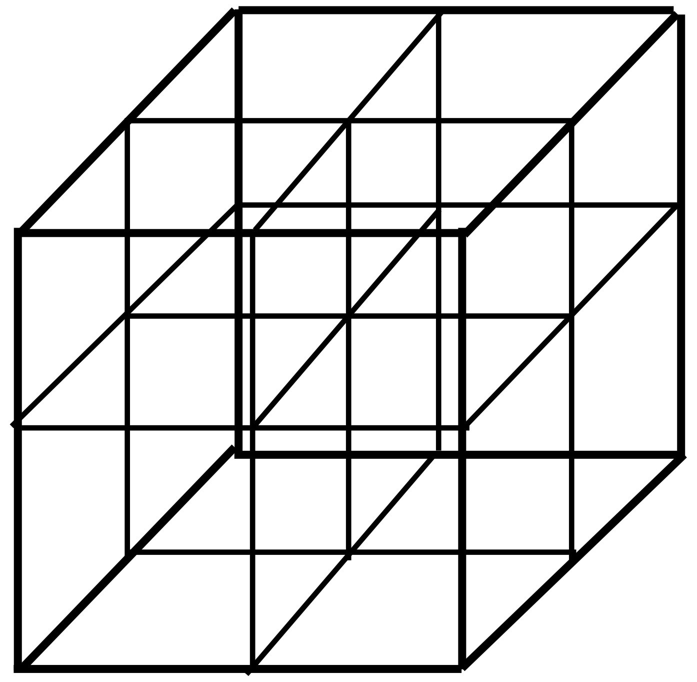
$\varepsilon(S) = 1/9$

Why the ternary cube?

- simplest high-dimensional domain where convexity is a non-trivial property (all sets in $\{\pm 1\}^n$ are convex)
- $Z \approx \frac{1}{\sqrt{k}} \sum_{i=1}^{k} X_i$ where $Z \sim \mathcal{N}(0,1)$ and $X_i \sim unif(\{-1,0,1\})$

Our Results

Black-Blais-Harms **[ITCS 24]**



Computational:

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$

Learning and testing with samples: $2^{\widetilde{O}(n^{3/4})}$

Learning and testing with samples: $2^{\Omega(n^{1/2})}$

Structural:

Def: $I(S) = 3^{-n} \cdot \#$ edges $(x, y) : x \in S, y \notin S$

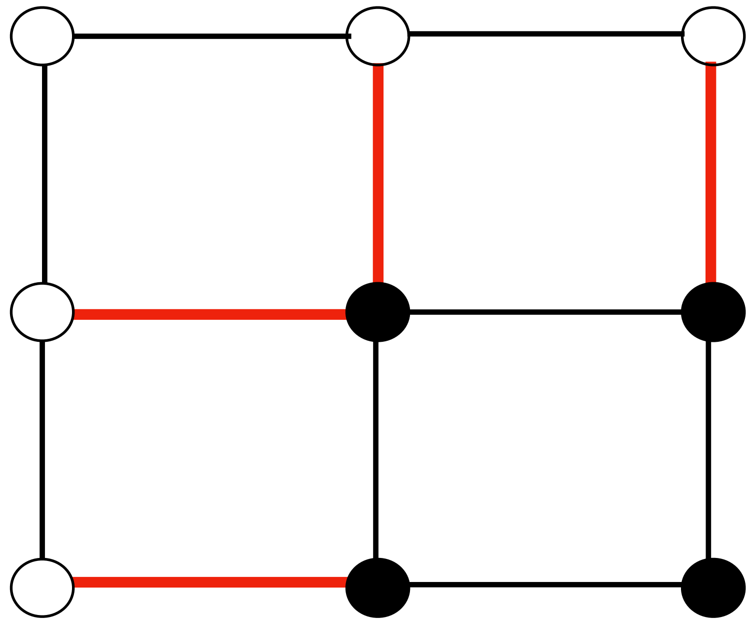
All convex sets satisfy $I(S) \leq \widetilde{O}(n^{3/4})$

There exists a convex set with $I(S) \geq \Omega(n^{3/4})$

- Uses the "Low-Degree Algorithm" of Linial-Mansour-Nisan 93

$$I(S) \leq B \implies \sum_{T:\ |T| > B/\varepsilon} \widehat{S}(T)^2 \leq \varepsilon$$

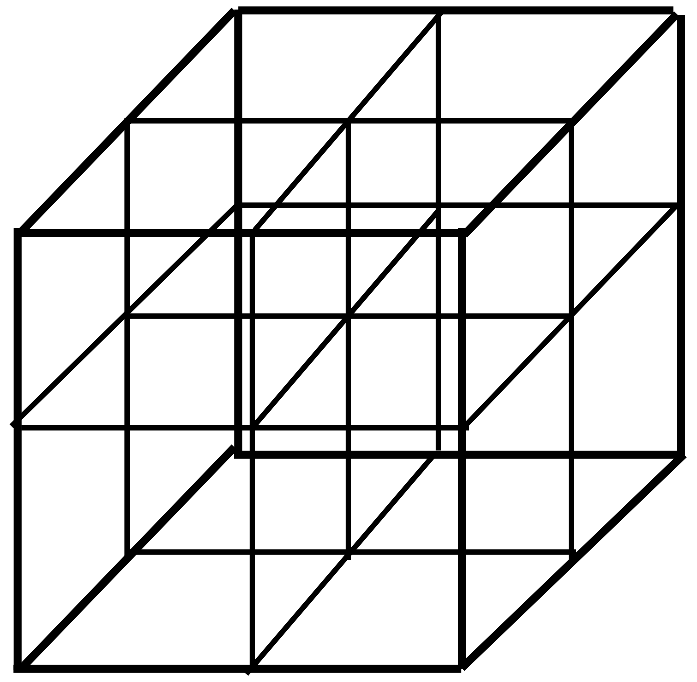
$\implies$ Can learn S to error ε with poly$(n^B, 1/\varepsilon)$ samples



$I(S) = \frac{4}{9}$

Our Results

Black-Blais-Harms **[ITCS 24]**



Computational:

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$

Learning and testing with samples: $2^{\widetilde{O}(n^{3/4})}$

Learning and testing with samples: $2^{\Omega(n^{1/2})}$

- Uses a version of *Talagrand's random DNF* (Talagrand [96]) adapted to $\{0, \pm 1\}^n$
- *Talagrand's random DNF* has been used to prove lower bounds for testing monotonicity, k-monotonicity, and unateness in $\{\pm 1\}^n$

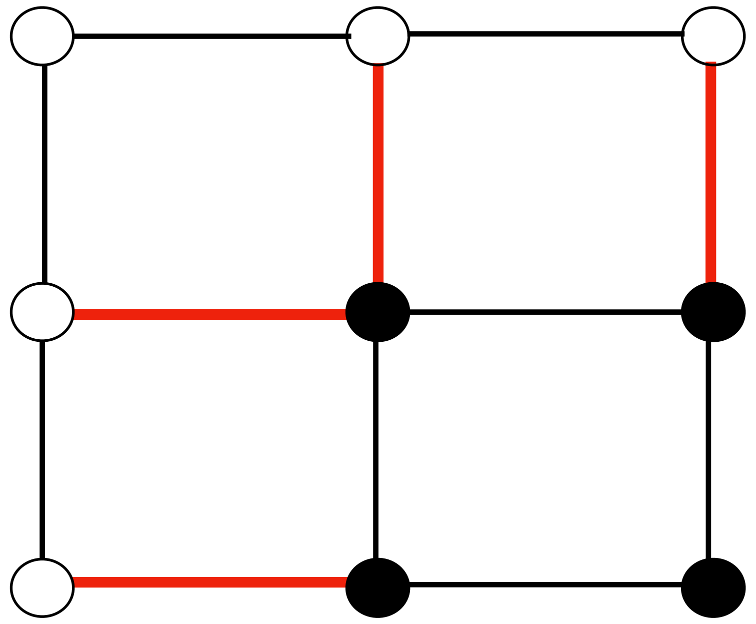
(Belovs-Blais [16], Chen-Waingarten-Xie [17], Chen-De-Li-Nadimpalli-Servedio [23], Black [23])

Structural:

Def: $I(S) = 3^{-n} \cdot \#$ edges $(x, y) : x \in S, y \notin S$

All convex sets satisfy $I(S) \leq \widetilde{O}(n^{3/4})$

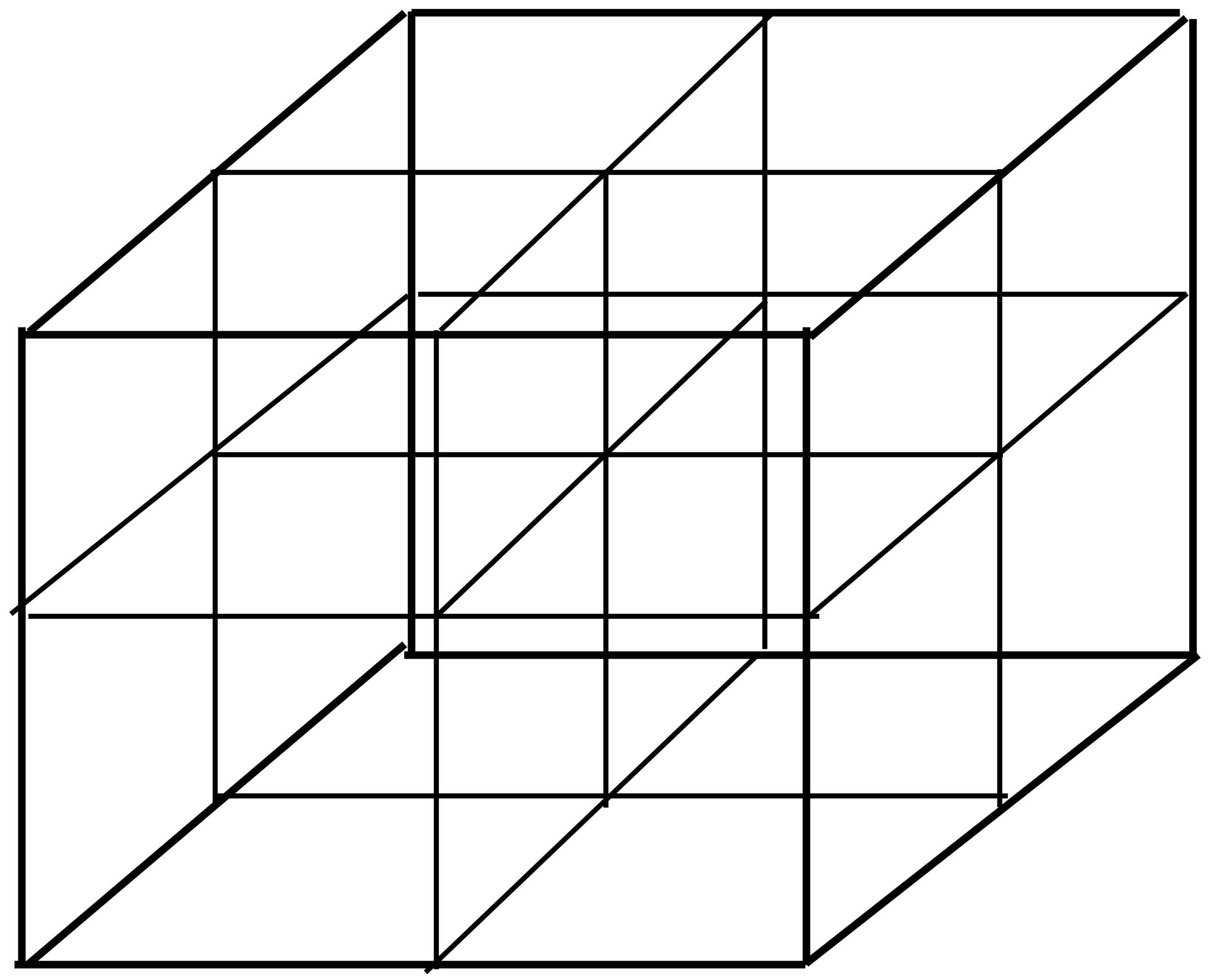
There exists a convex set with $I(S) \geq \Omega(n^{3/4})$



$I(S) = \frac{4}{9}$

The Influence of Convex Sets

The maximum influence of convex sets in $\{0, \pm 1\}^n$ is $\widetilde{\Theta}(n^{3/4})$



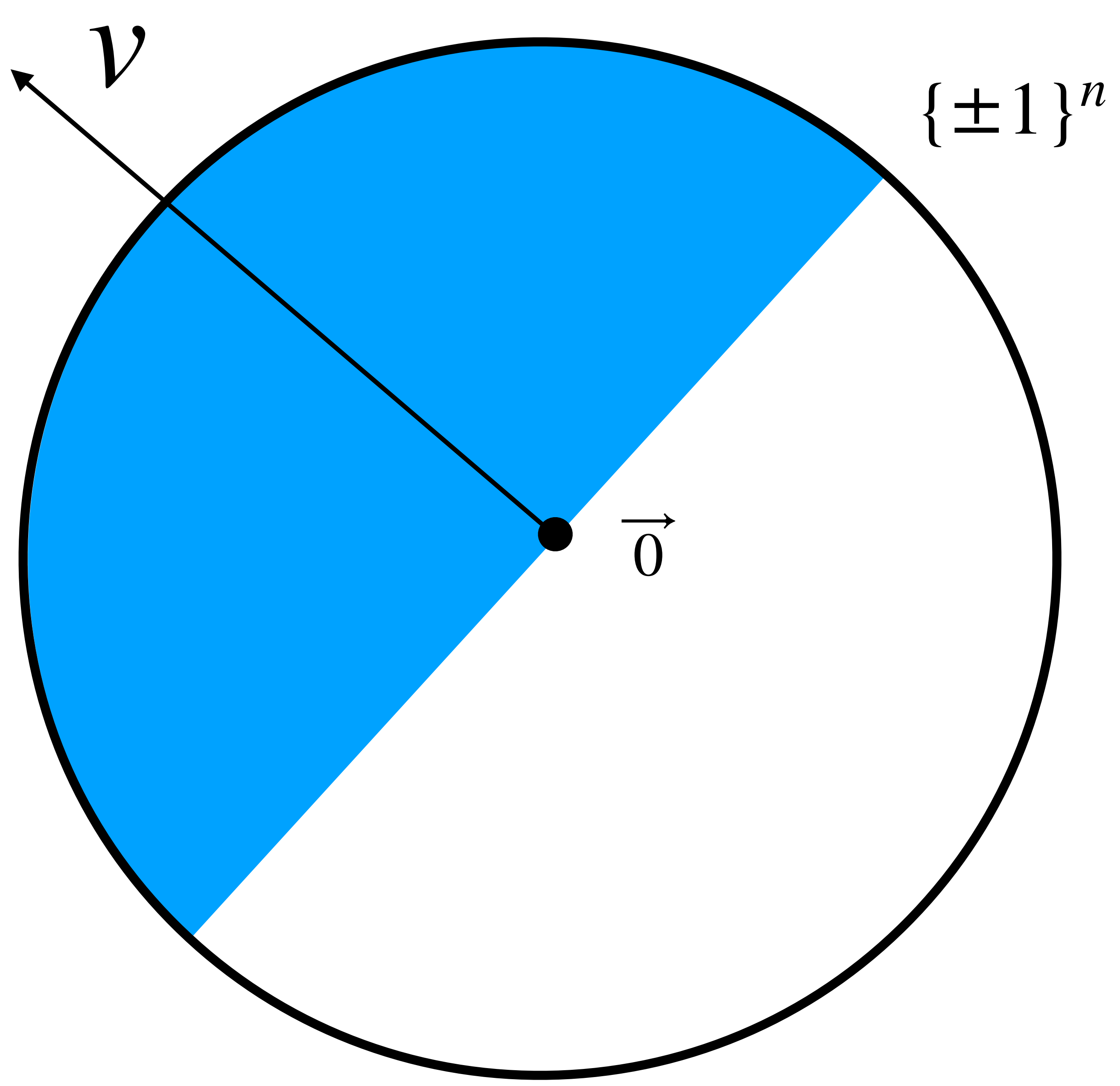
Examples

Def: $I(S) = 3^{-n} \cdot$ # edges (x, y) : $x \in S, y \notin S$

$$= \mathbb{E}_x[\#\ \textit{edges}\ (x, y) \colon x \in S, y \notin S]$$

Halfspace

$$H = \{x \colon \langle x, \overrightarrow{1} \rangle \geq 0\}$$



$$I(H) \approx \mathbb{P}_x[\langle x, \overrightarrow{1} \rangle = 0] \cdot \Theta(n)$$

$$\approx \mathbb{P}_x\left[\sum_i x_i = 0\right] \cdot \Theta(n)$$

$$\approx \Theta(n^{1/2})$$

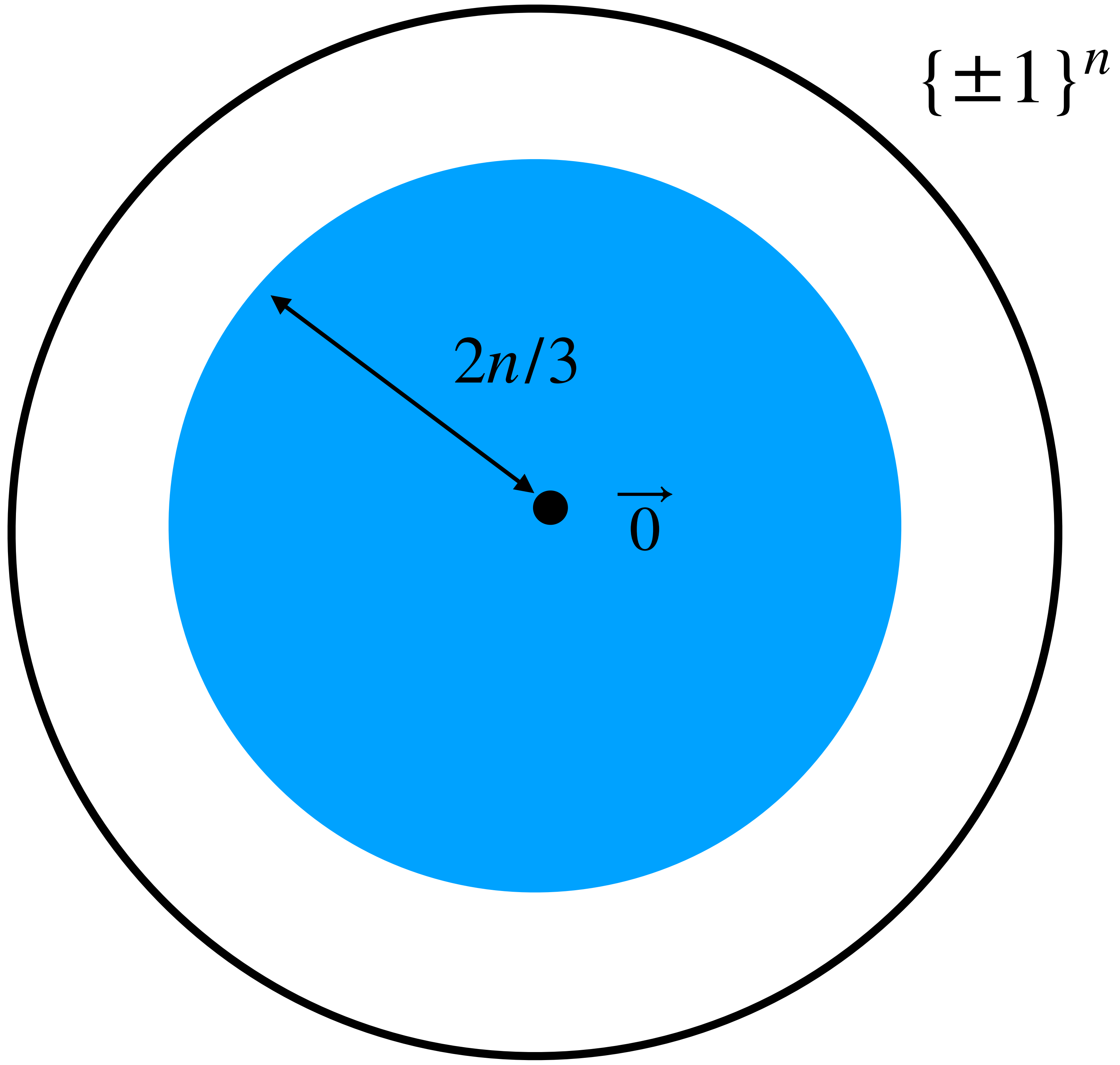
Examples

Def: $I(S) = 3^{-n} \cdot \#$ edges $(x, y) : x \in S, y \notin S$

$$= \mathbb{E}_x[\# \textit{ edges } (x, y) : x \in S, y \notin S]$$

Ball

$$B_{2n/3} = \{x : \|x\|_1 \leq 2n/3\}$$



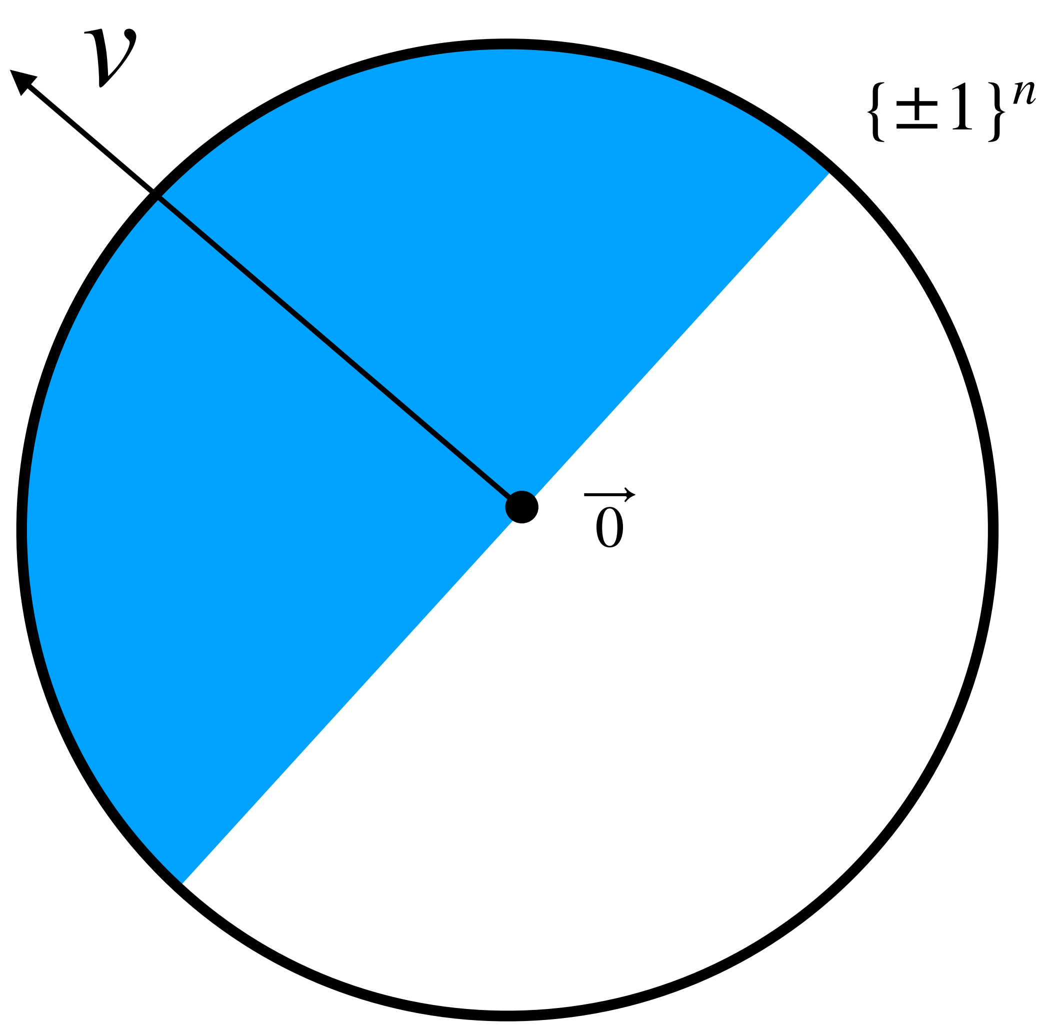
$$I(B_{2n/3}) = \mathbb{P}_x[\|x\|_1 = 2n/3] \cdot 2n/3$$

$$= \frac{1}{3^n}\binom{n}{2n/3} \cdot 2^{2n/3} \cdot \Theta(n)$$

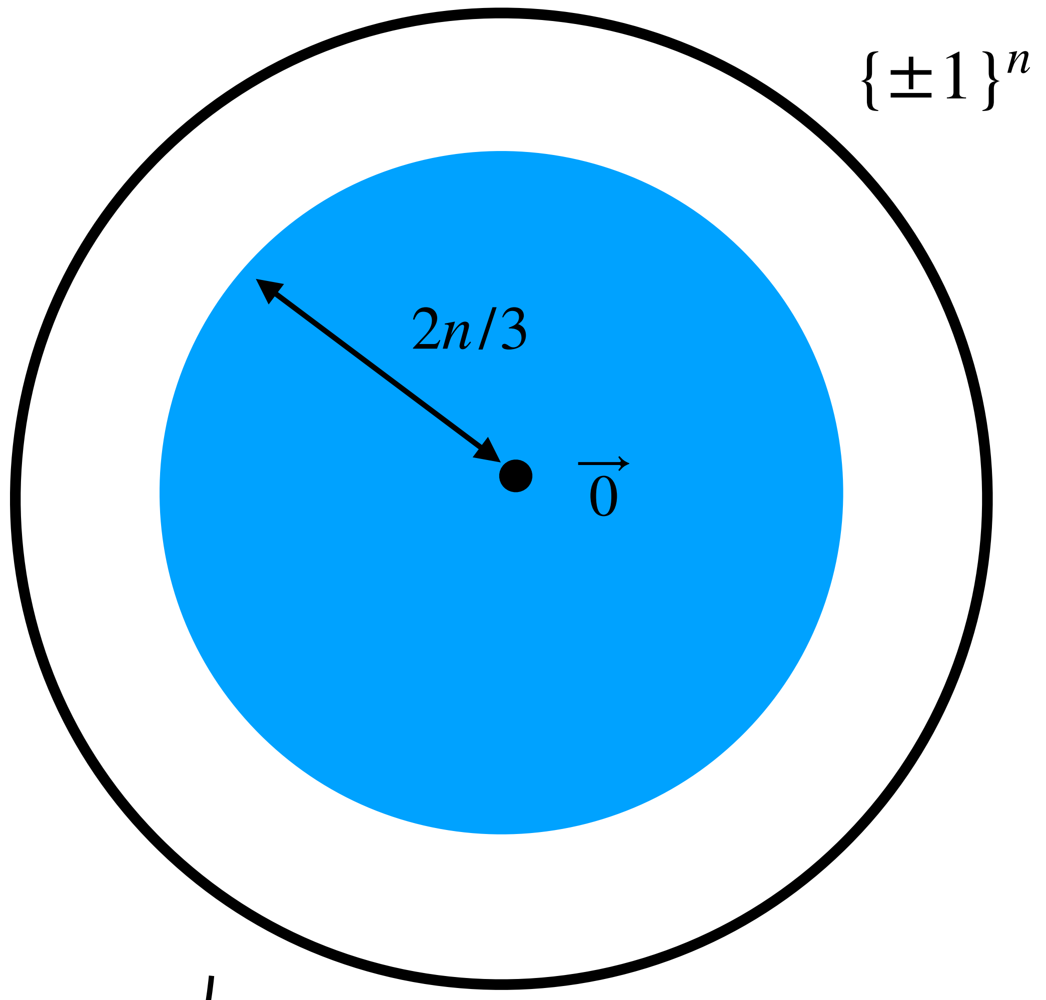
$$\approx \Theta(n^{1/2})$$

High Influence Set

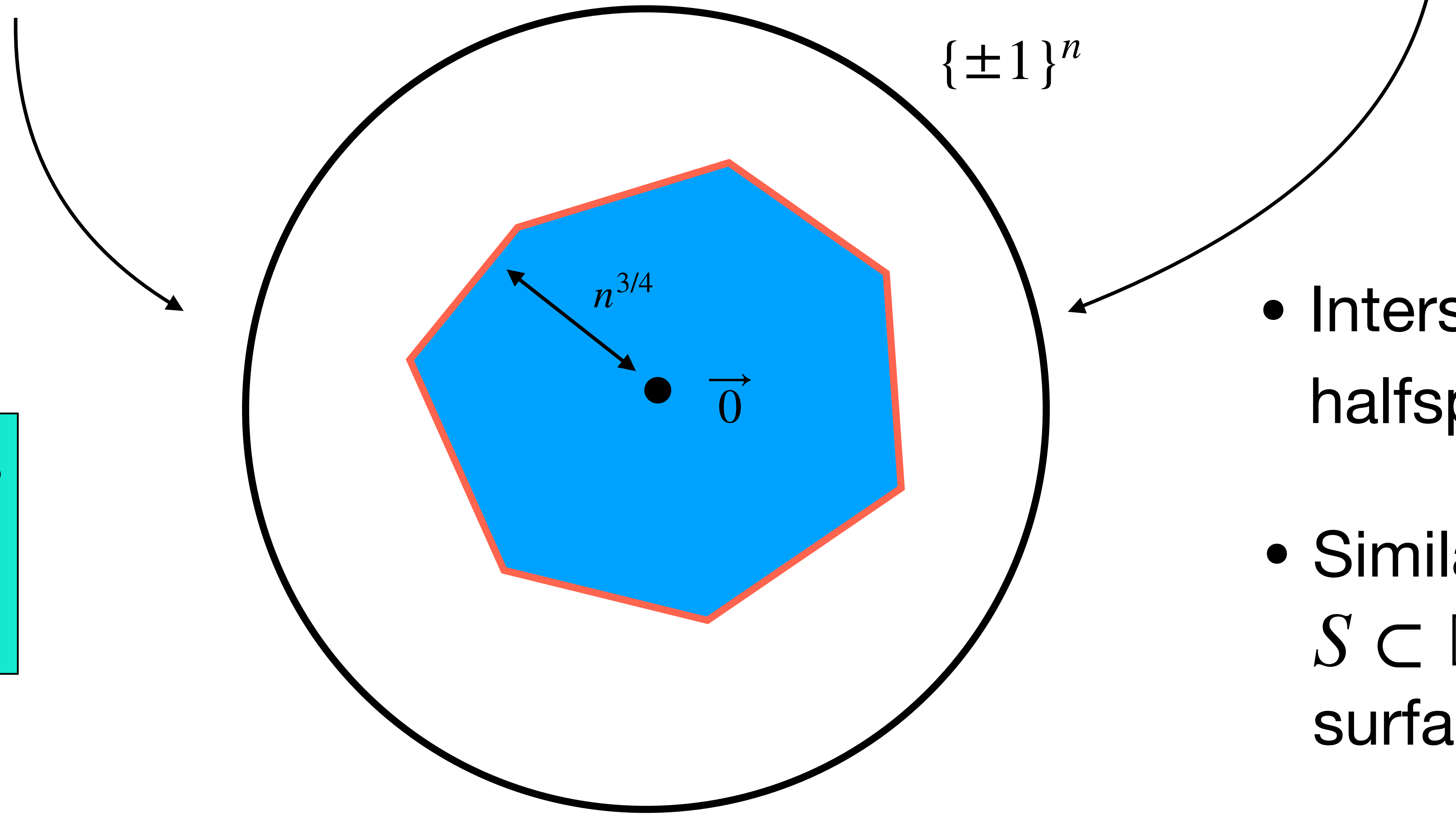
Halfspace



Influence $\approx n^{1/2}$



Ball

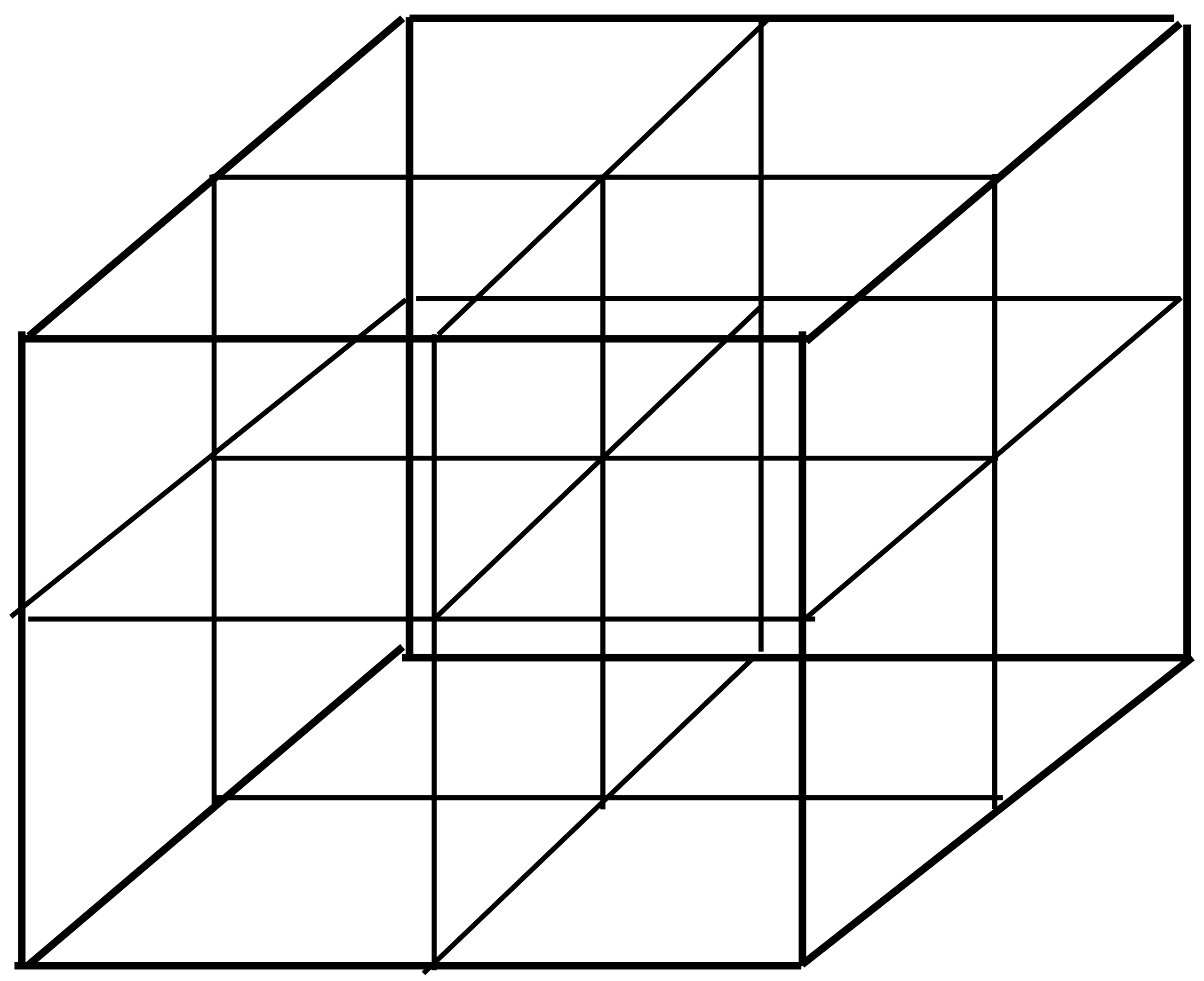


Theorem: There exists a convex set with $I(S) \geq \Omega(n^{3/4})$

- Intersection of $3\sqrt{n}$ random halfspaces with distance $n^{3/4}$
- Similar construction obtains convex $S \subset \mathbb{R}^n$ with maximum Gaussian surface area (Nazarov [03])

Proof Sketch

All convex sets satisfy
$$I(S) \leq \widetilde{O}(n^{3/4})$$



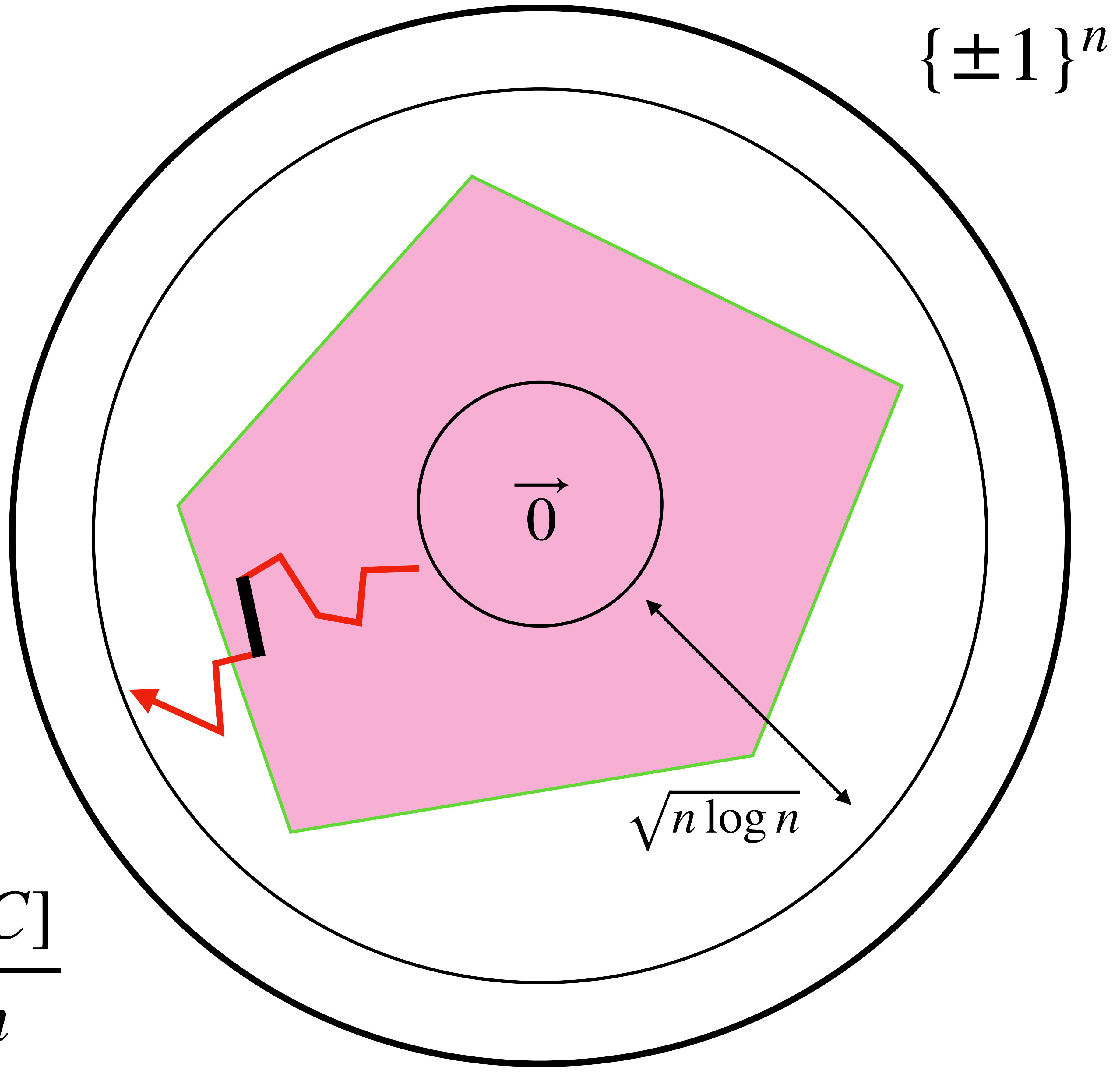
The Edge-Boundary of Convex Sets

Def: $I(S) = 3^{-n} \cdot \#$ edges $(x, y) : S(x) \neq S(y)$ $\approx n \cdot \mathbb{P}_{(x,y)}[S(x) \neq S(y)]$

$I(S) \leq \widetilde{O}(n^{3/4})$ $\iff$ $\mathbb{P}_{(u,v)}[S(u) \neq S(v)] \leq \widetilde{O}(n^{-1/4})$

Distribution D over edges:

- Perform a **directed** random walk of length $m \approx \sqrt{n/\log n}$ in the middle $\sqrt{n \log n}$ layers
- Return a random edge (x, y) from the walk

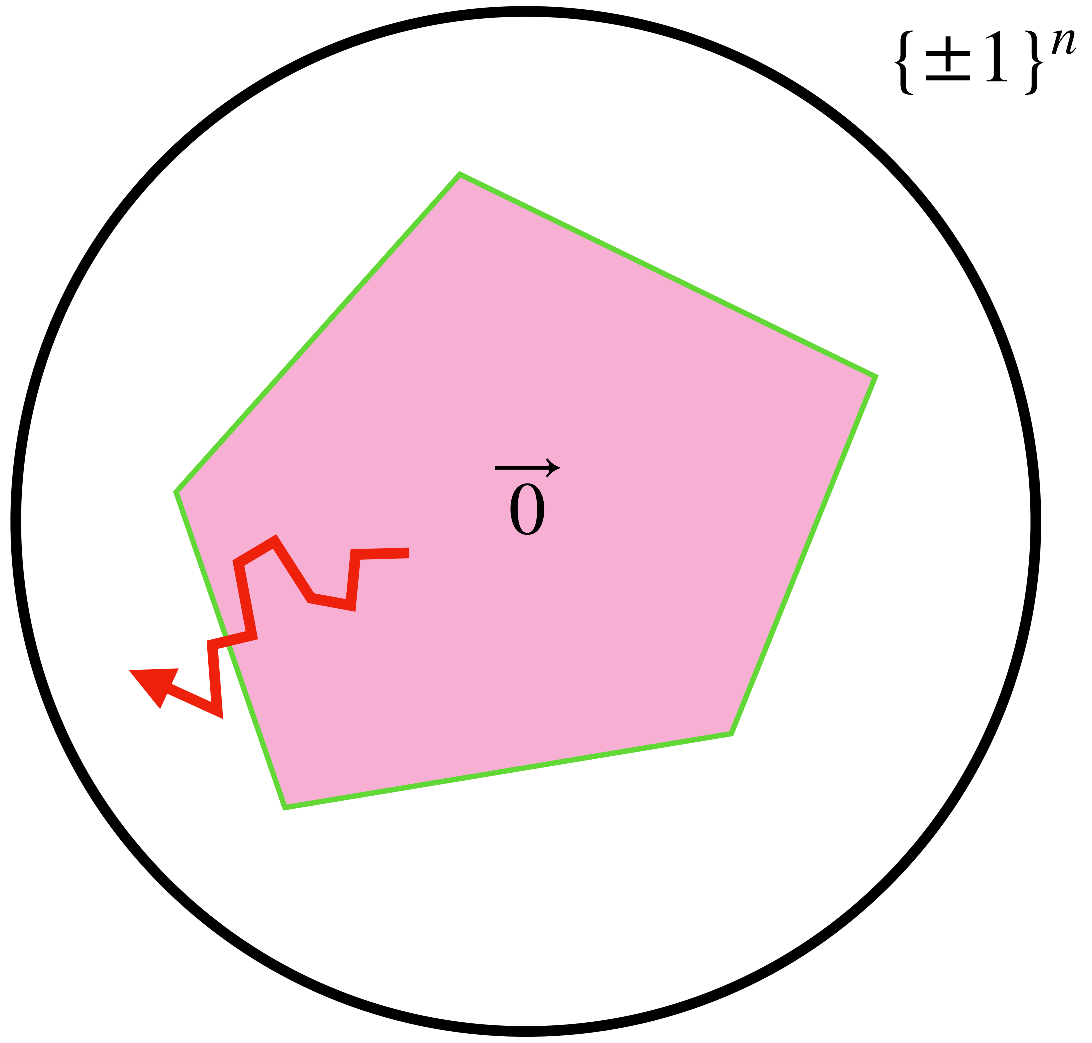


Let $C = \#$ times walk crosses boundary of S

1) D is "close" to uniform $\implies$ $\mathbb{P}_{(x,y)}[S(x) \neq S(y)] \approx \dfrac{\mathbb{E}[C]}{m}$

2) $\mathbb{E}[C] \leq O(\sqrt{m})$

Halfspaces and 1-D Walks



Lemma: $\mathbb{E}[C] \leq O(\sqrt{m})$

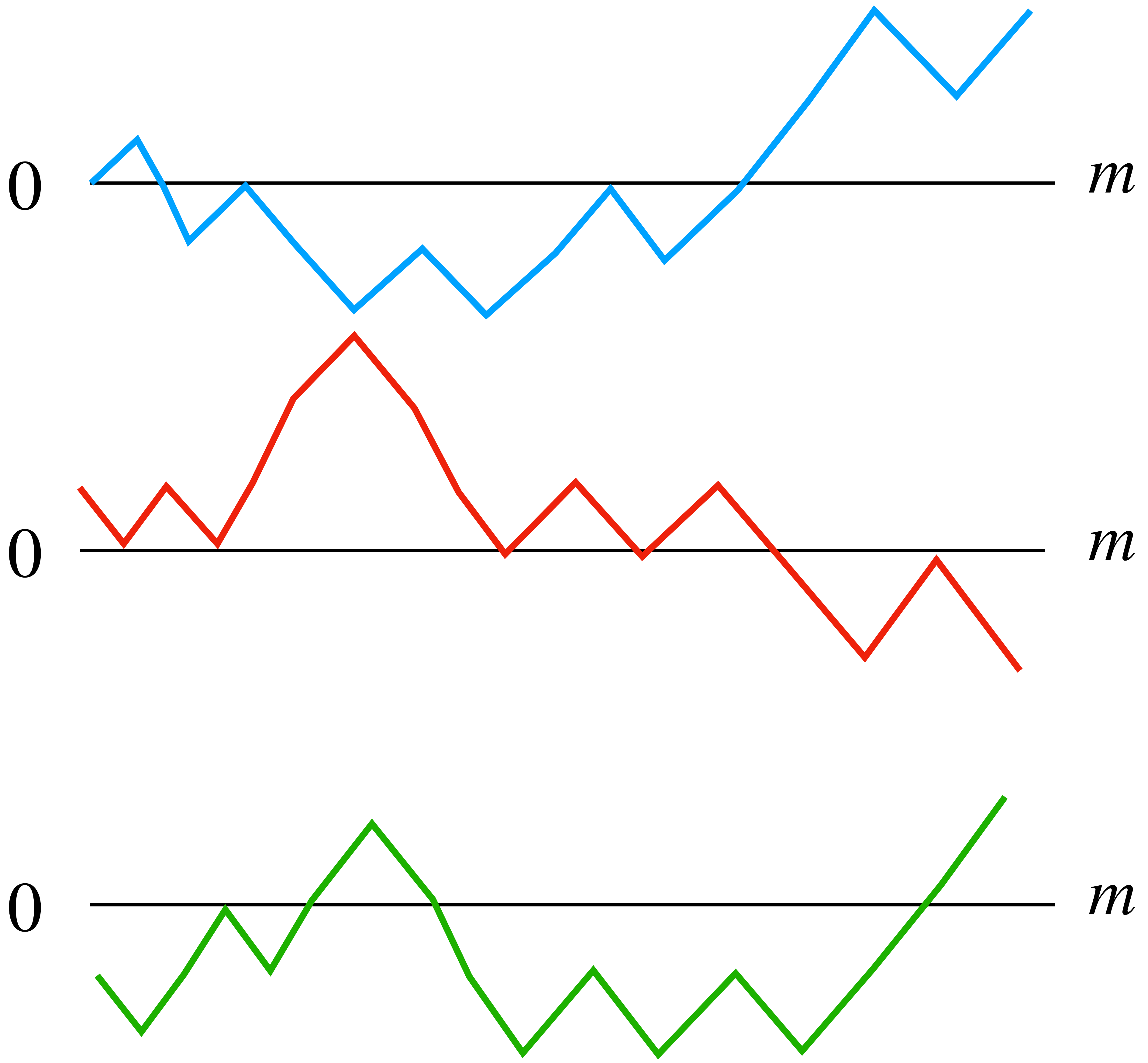
S convex $\implies S = \bigcap_{i=1}^{k} H_i$ where $H_i = \{x : \langle x, v_i \rangle < \tau_i\}$

$w_1(t) = \langle z^{(t)}, v_1 \rangle - \tau_1$

$w_2(t) = \langle z^{(t)}, v_2 \rangle - \tau_2$

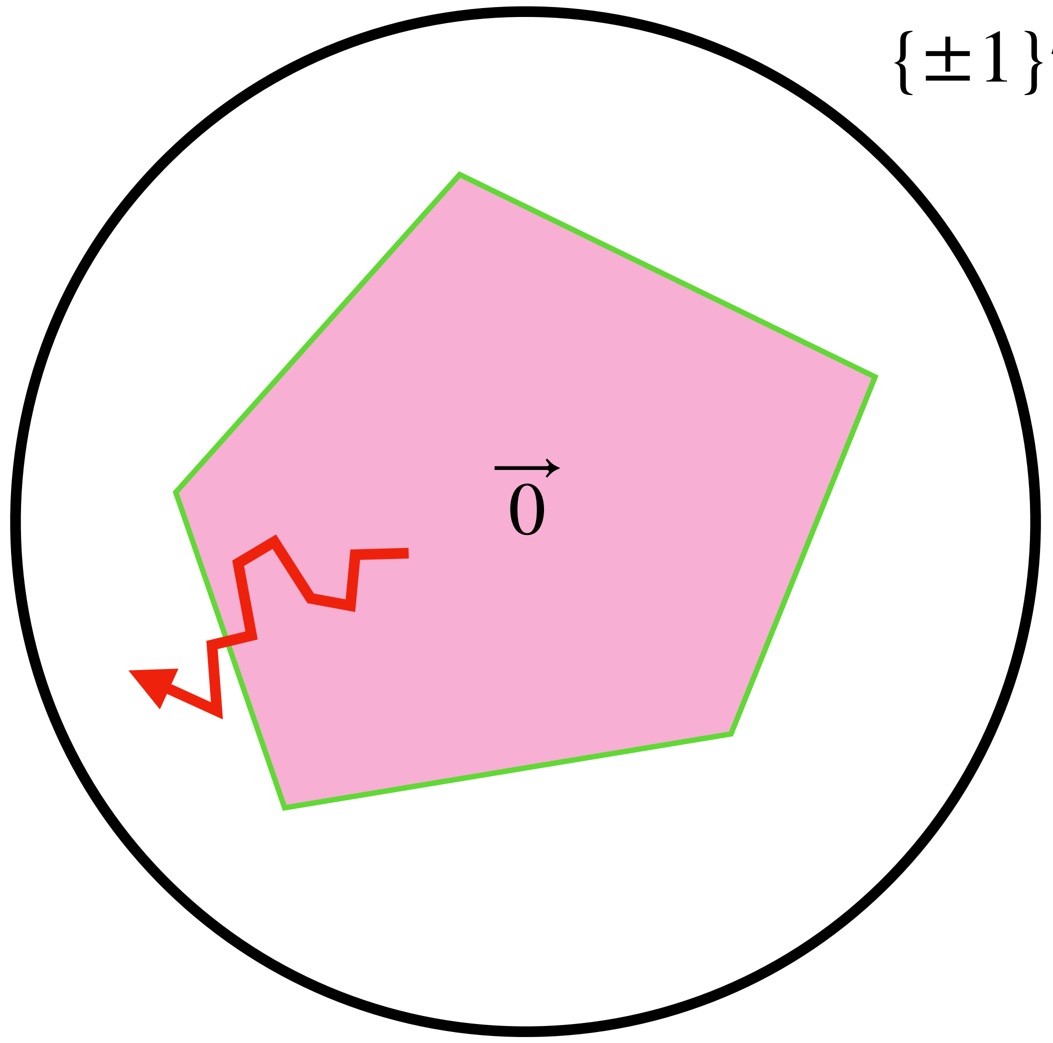
$\vdots$

$w_k(t) = \langle z^{(t)}, v_k \rangle - \tau_k$



Intuition:
If $v \in \{\pm 1\}^n$, then expected # crossings for a single walk is $\leq O(\sqrt{m})$

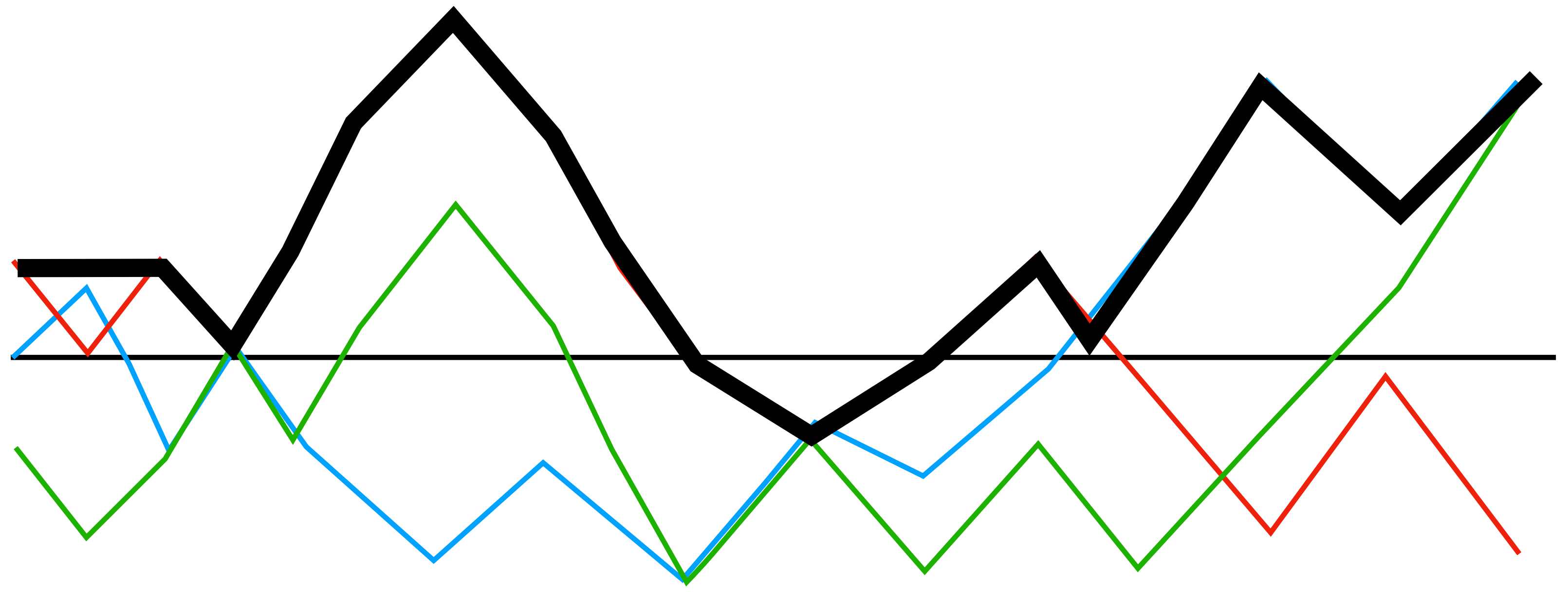
The Max-Walk



Lemma: $\mathbb{E}[C] \leq O(\sqrt{m})$

S convex $\implies S = \bigcap_{i=1}^{k} H_i$ where $H_i = \{x : \langle x, v_i \rangle < \tau_i\}$

$$W(t) = \max_{i \in [k]} \langle z^{(t)}, v_i \rangle - \tau_i$$



$C =$ # times max walk crosses the origin

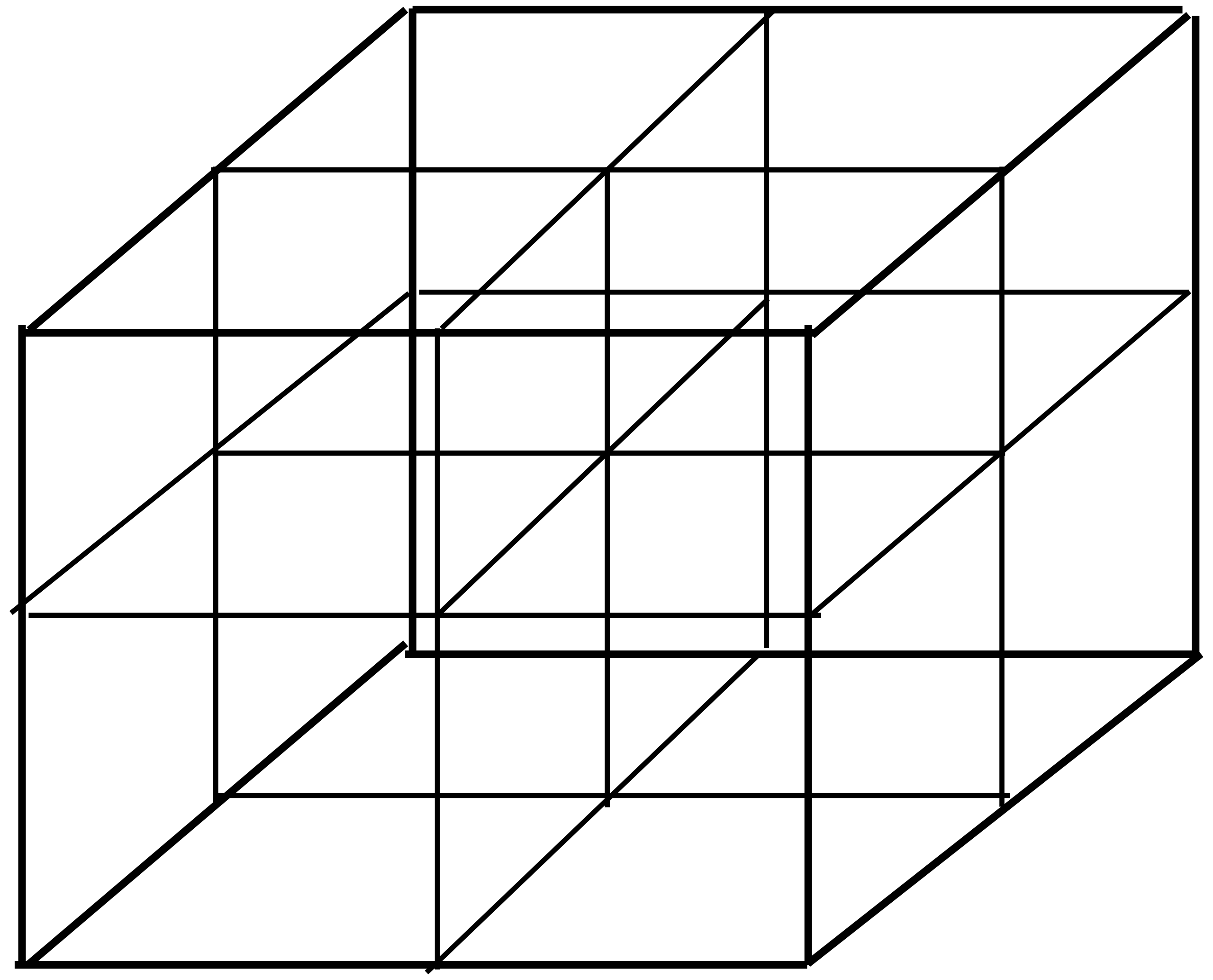
Challenges:

- v_i's can be arbitrary vectors in $\mathbb{R}^n$
- Analyzing max-walk for arbitrary real vectors is tricky

Key tool: Sparre Andersen's fluctuation theorem [Sparre '54]

Testing with Queries

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$



Testing with Queries

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$

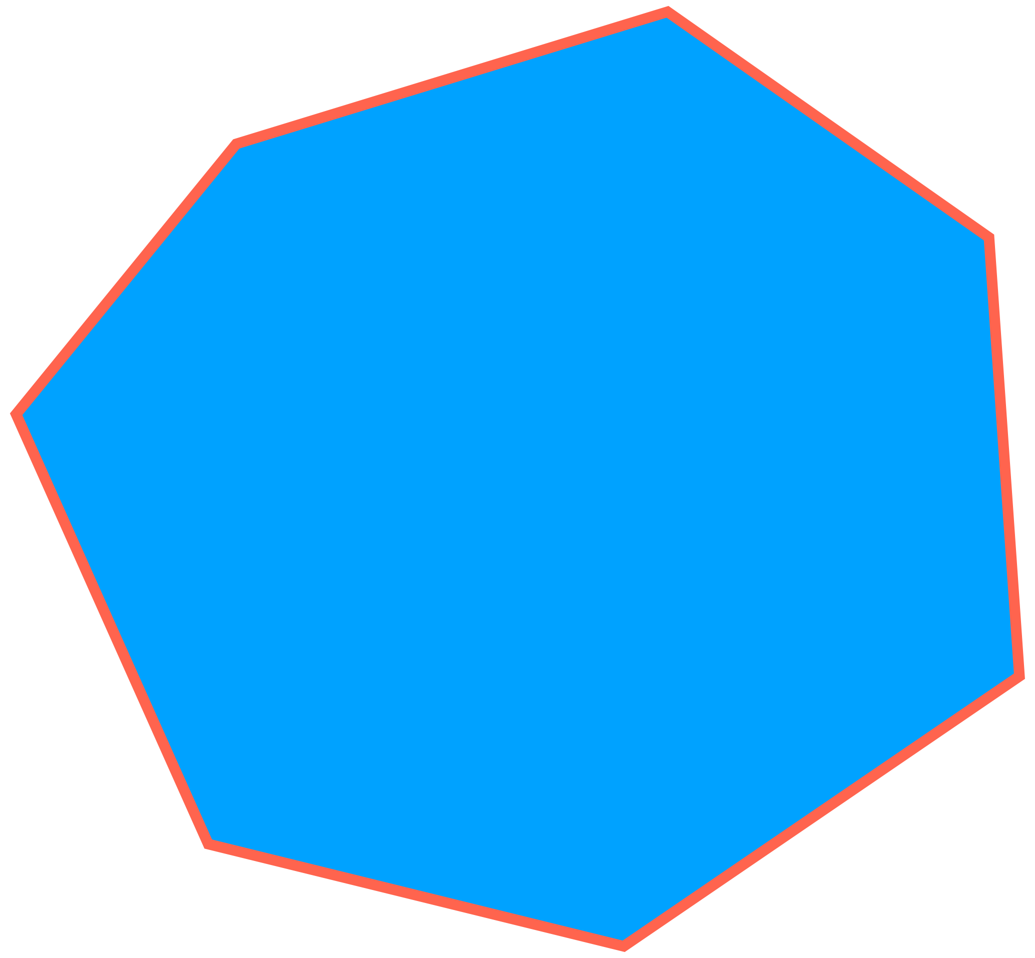
Testing

Given S and $\varepsilon > 0$...

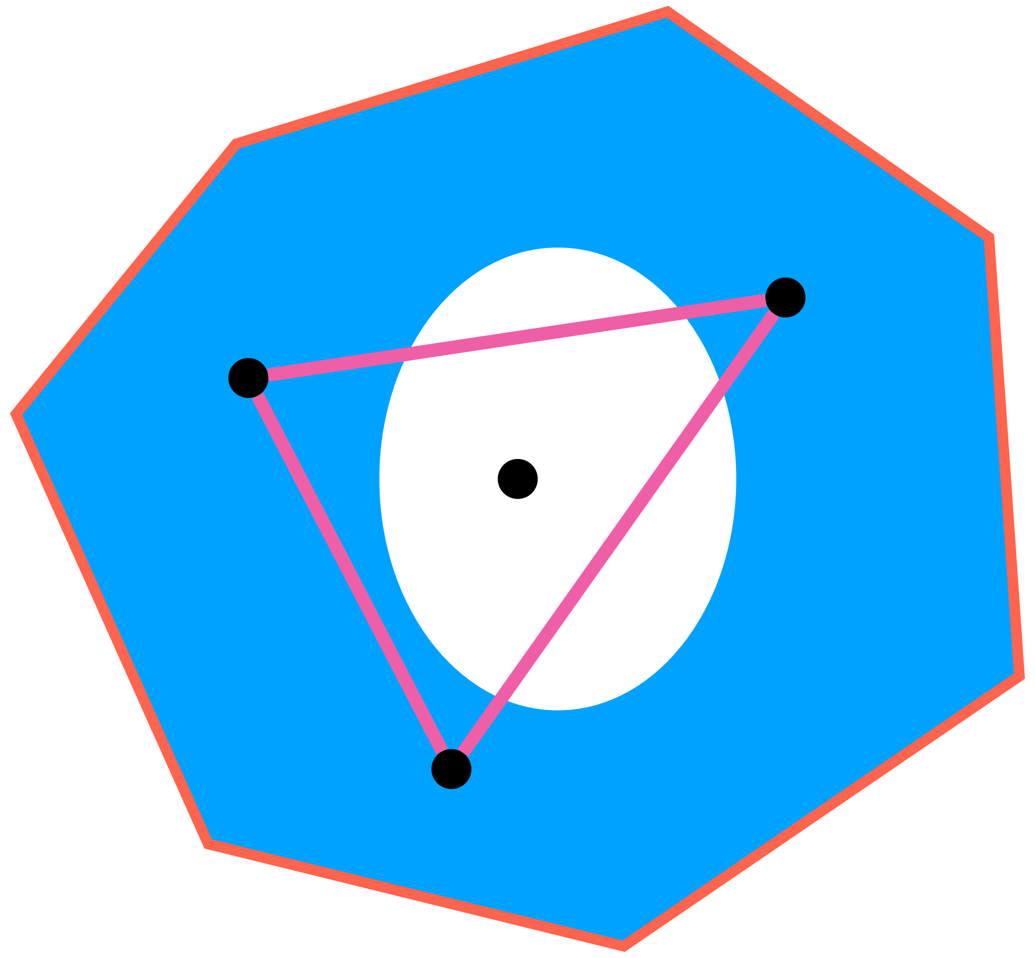
1. if S convex: **accept** w.p. 1
2. if $\varepsilon(S) > \varepsilon$: **reject** w.p. $> 2/3$

Testing with Queries

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$



Always accept



Find a witness of non-convexity
with probability $> 2/3$

Witness

$y \notin S$

$X \subseteq S$: $y \in \text{Conv}(X)$

Question

How many queries to find a witness when S if ε-far from convex?

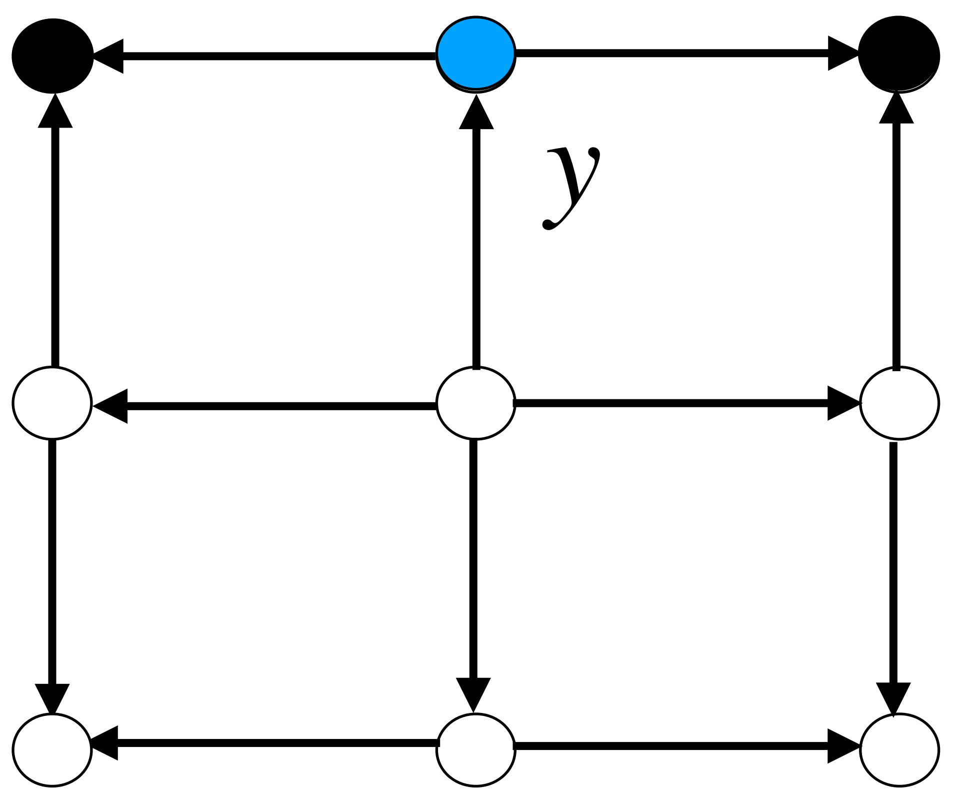
Special Structure of $\{0, \pm 1\}^n$

Def: Outward-oriented poset on $\{0, \pm 1\}^n$

$y \prec x$ iff $y_i \neq 0 \implies x_i = y_i$

Minimal point: $\vec{0}$

Maximal points: $\{\pm 1\}^n$



Fact: Suppose $y \in \mathrm{Conv}(X)$ and X is **minimal.** Then $y \prec x$ for all $x \in X$

Proof: $y = \sum_{x \in X} \lambda_x x$ *Suppose* $y_i = 1$*...* *i.e.,* $\sum_{x \in X} \lambda_x x_i = 1$

X *minimal* $\implies \lambda_x > 0$ *for all* x

If $x_i \neq y_i$ *for some* x*, then* $\sum_{x \in X} \lambda_x x_i < 1$***. Contradiction.*** $\square$

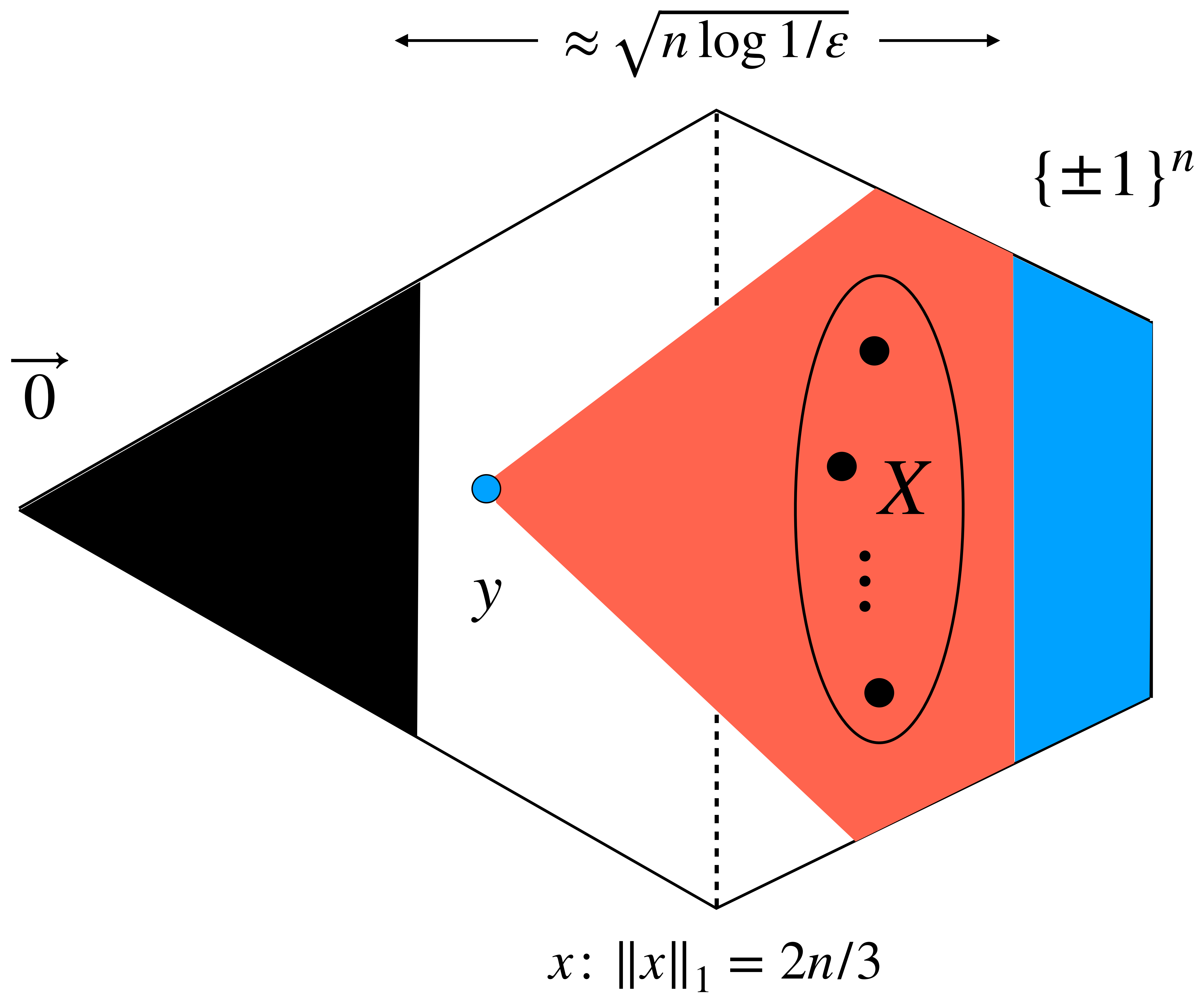
Testing Upper Bound

Witness

$y \notin S$

$X \subseteq S \colon y \in \mathrm{Conv}(X)$

Fact: Suppose $y \in \mathrm{Conv}(X)$ and X is **minimal.** Then $y \prec x$ for all $x \in X$



$$3^{-n} |\mathrm{Conv}(S) \backslash S| \geq \varepsilon(S) > \varepsilon$$

$$\implies \Pr_y[y \in \mathrm{Conv}(S) \backslash S] \geq \varepsilon$$

If we can find $X \subset S$ such that $y \in \mathrm{Conv}(X)$, then we win

Obs 1: By Fact, $\exists$ such an X where $y \prec x \ \forall x \in X$

Obs 2: By concentration bounds, it suffices to query only x such that $\|x\|_1 \leq 2n/3 + O(\sqrt{n \log 1/\varepsilon})$

# of points satisfying (1) and (2)

$$\approx \sum_{\ell=1}^{\sqrt{n \log 1/\varepsilon}} \binom{\#i \colon y_i = 0}{\ell} \cdot 2^{\ell} \leq 2^{\widetilde{O}(\sqrt{n \log 1/\varepsilon})}$$

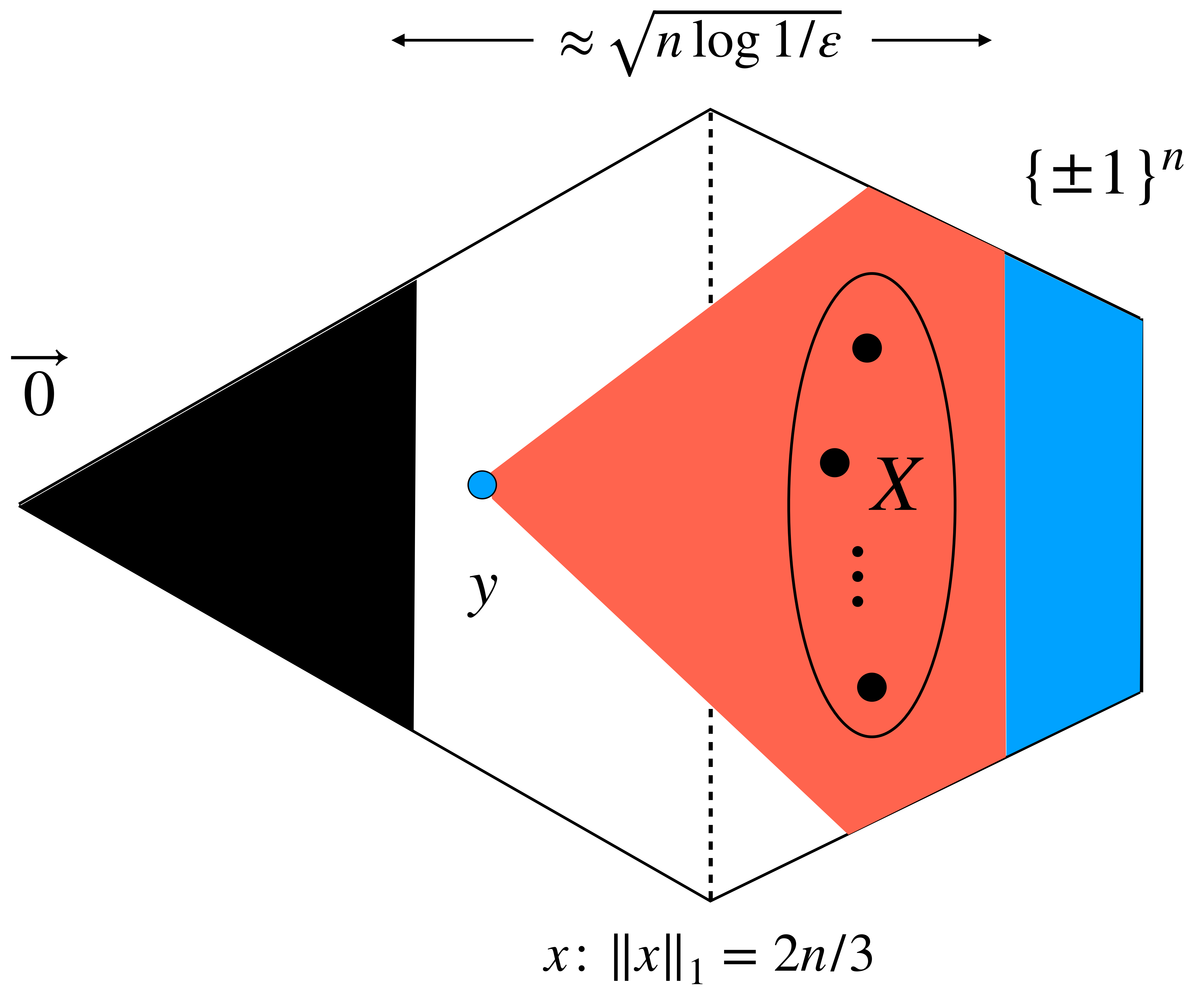
Testing Upper Bound

Witness

$y \notin S$

$X \subseteq S \colon y \in \mathrm{Conv}(X)$

Fact: Suppose $y \in \mathrm{Conv}(X)$ and X is **minimal.** Then $y \prec x$ for all $x \in X$



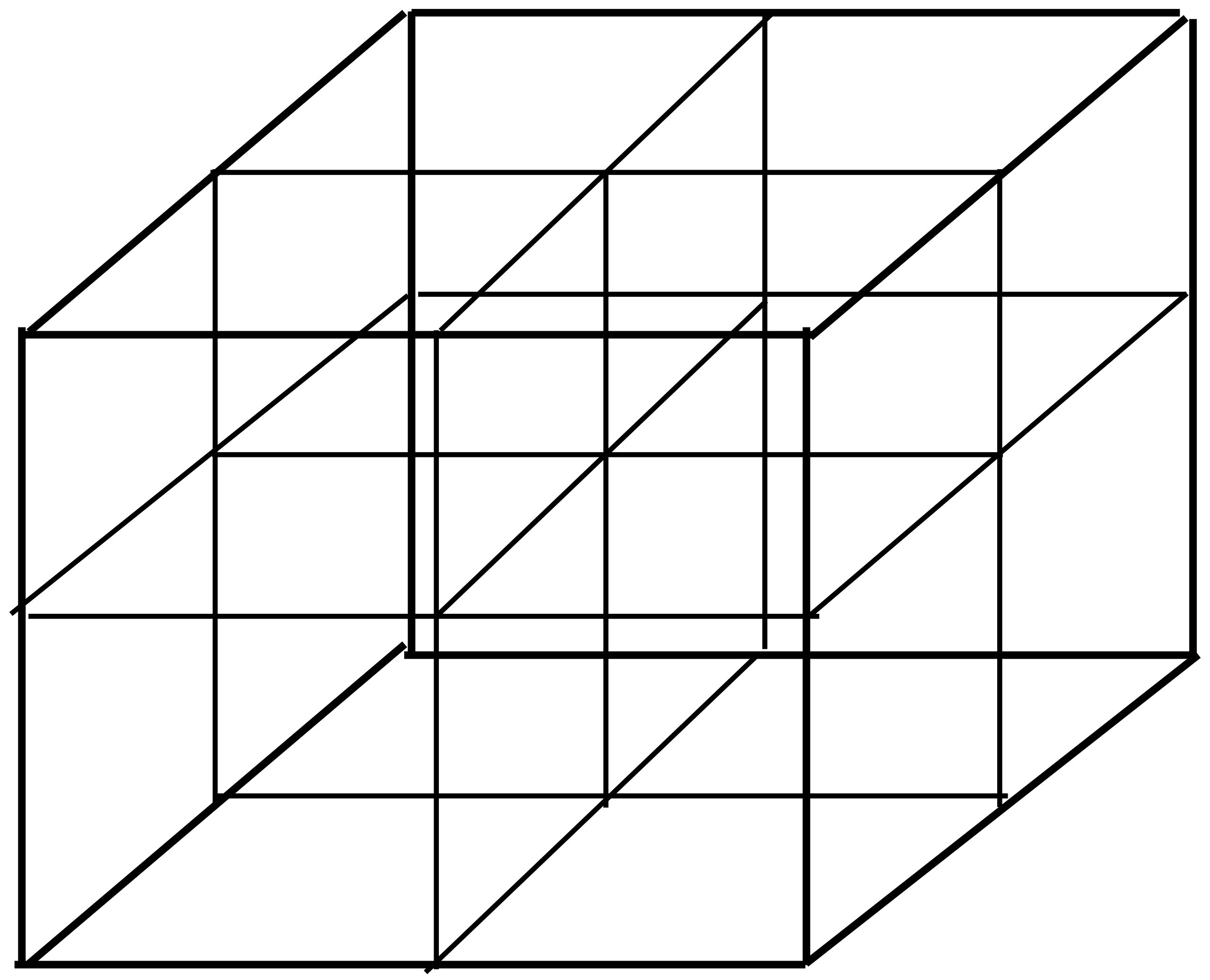
Tester

Repeat $O(1/\varepsilon)$ times:

- Query $y \in \{0, \pm 1\}^n$ uniformly at random
- If $\|y\|_1 > 2n/3 - \widetilde{O}(\sqrt{n})$, then query all of $U_y = \{x \succ y \colon \|x\|_1 \leq 2n/3 + \widetilde{O}(\sqrt{n})\}$
- If $y \notin S$ and there exists $X \subset U_y \cap S$ such that $y \in \mathrm{Conv}(X)$, then **reject**

Proof Sketch

1-sided non-adaptive **query**-based testing: $2^{\Omega(n^{1/2})}$



Hard Family of Sets: Truncated Anti-Slabs

Def: Given $v \in \{\pm 1\}^n$, let
$\mathrm{Slab}_v = \{x : |\langle x, v\rangle| < n^{1/2}\}$

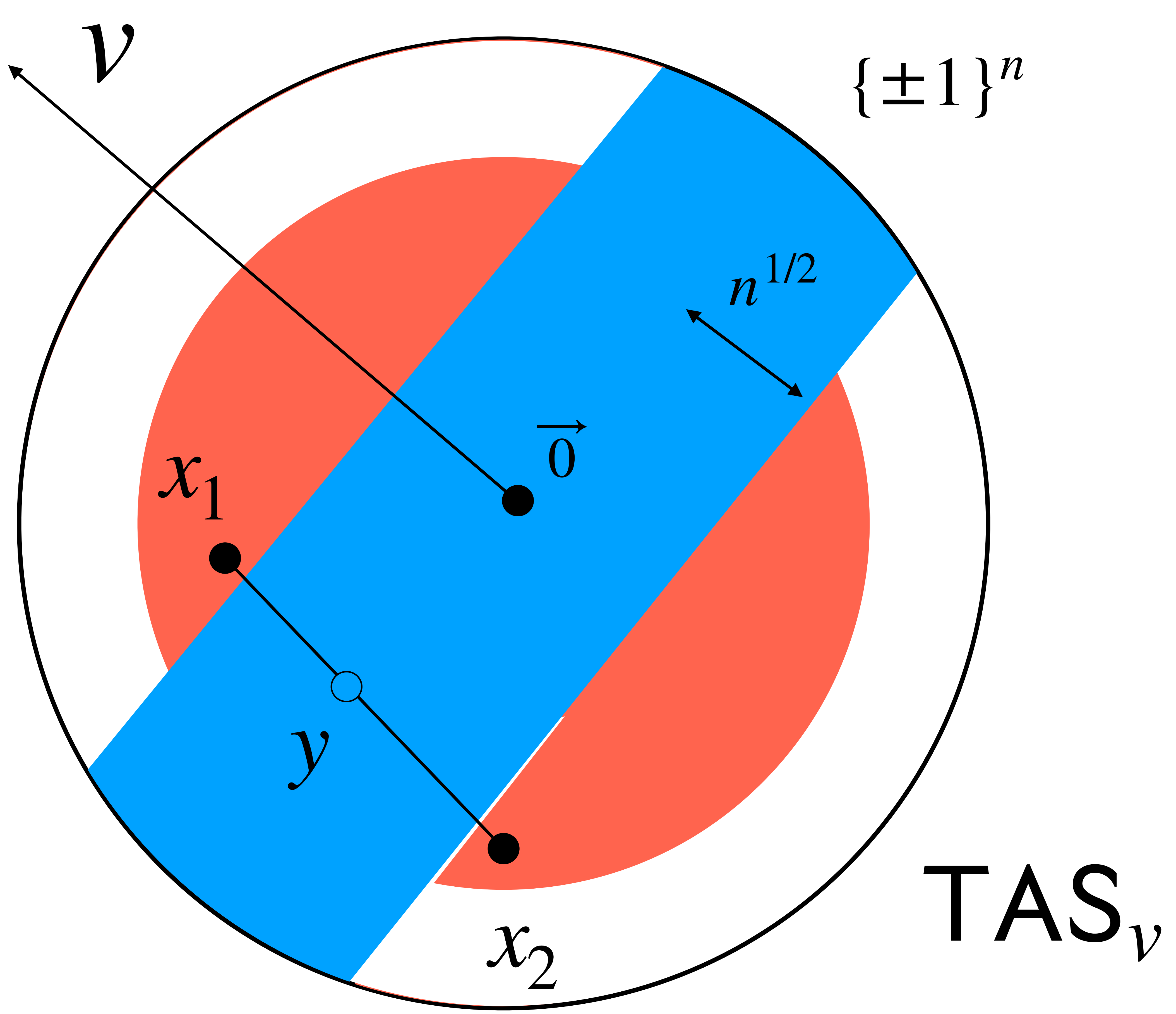
$$\mathrm{TAS}_v = \overline{\mathrm{Slab}_v} \cup \left\{x : \|x\|_1 < \frac{2n}{3} - 0.6n^{1/2}\right\} \setminus \left\{x : \|x\|_1 > \frac{2n}{3} + 0.6n^{1/2}\right\}$$

Fact: $\varepsilon(\mathrm{TAS}_v) = \Omega(1)$

$\exists x \in \{x_1, x_2\}$:

(A) $|\langle x - y, v\rangle| > n^{1/2}$

(B) $y < x \implies \|x - y\|_1 < 1.2n^{1/2}$



Witnesses of Non-Convexity

Let T be a 1-sided non-adaptive tester

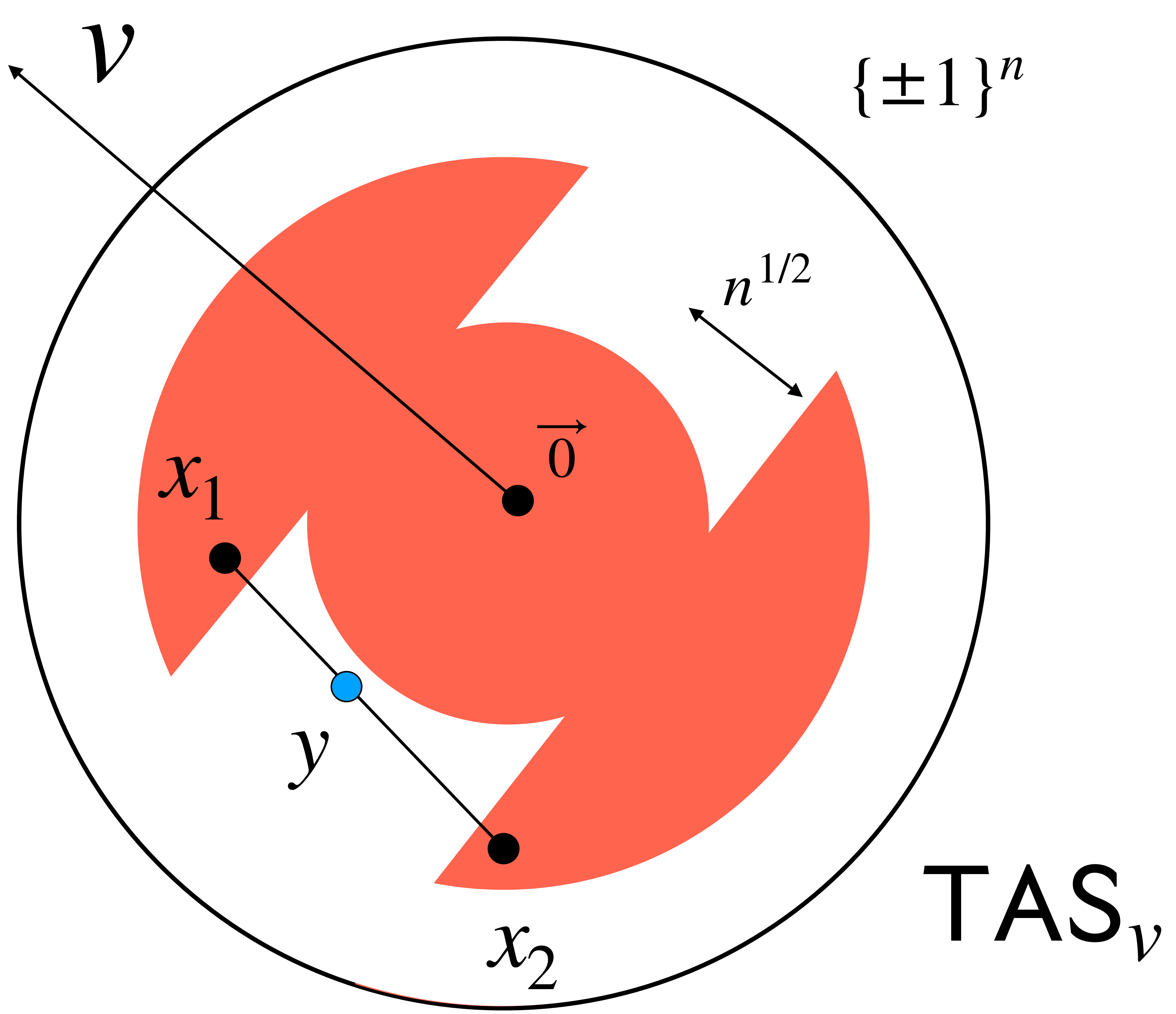
- Query set: $Q \subset \{0, \pm 1\}^n$
- T rejects $\text{TAS}_v \implies \exists x, y \in Q$

(A) $|\langle x - y, v \rangle| > n^{1/2}$

(B) $\|x - y\|_1 < 1.2n^{1/2}$

Question:
If **(B)** holds for x, y, then for how many $v \in \{\pm 1\}^n$ can **(A)** hold?

Answer:
at most $2^{n - 0.08n^{1/2}}$



Lower Bound Proof

Let T be a 1-sided non-adaptive tester

$w(Q) = \#\, v : Q$ witnesses non-convexity of TAS_v

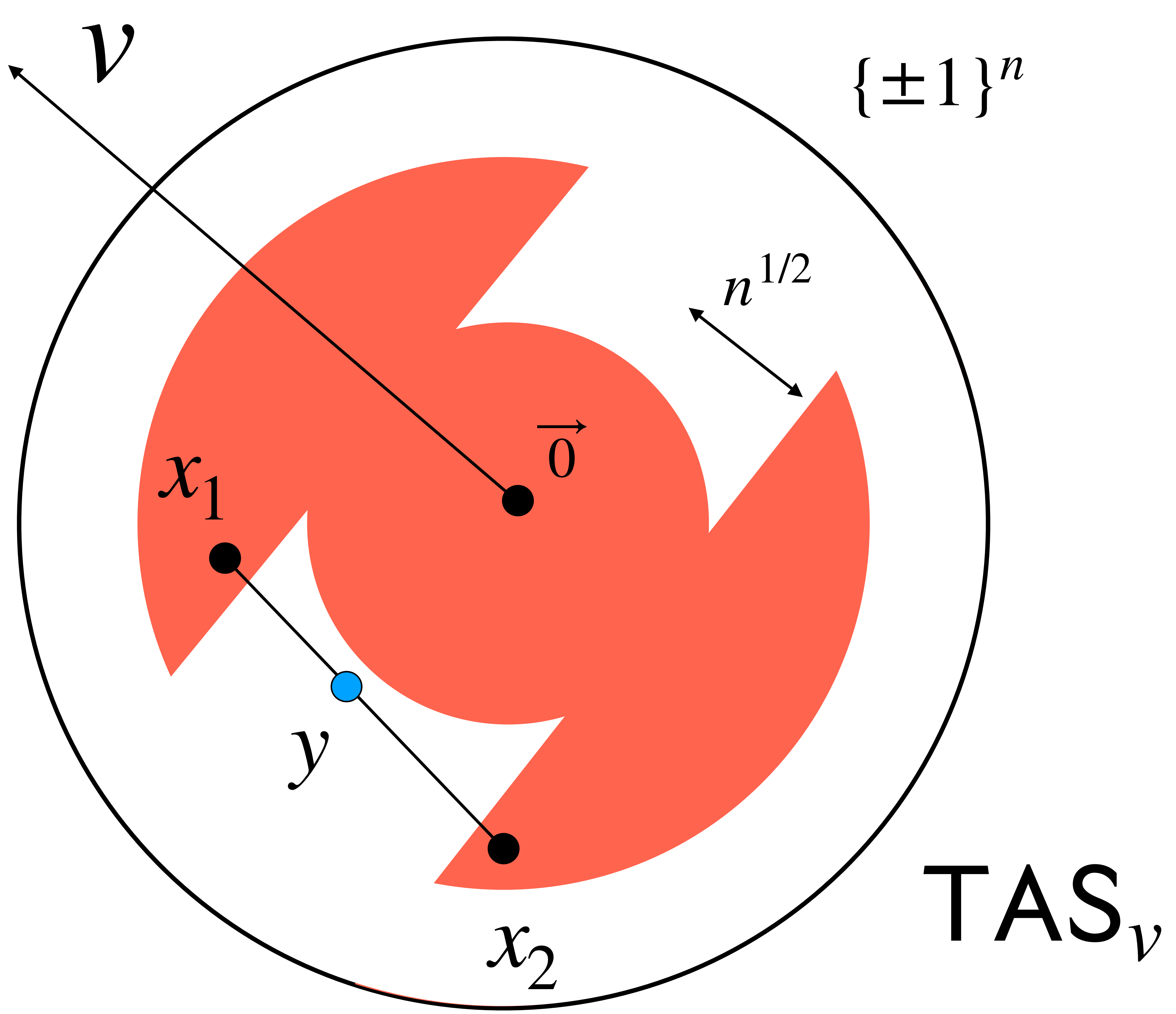
$$\leq |Q|^2 \cdot 2^{n-0.08n^{1/2}}$$

$$\frac{2}{3} \cdot 2^n \leq \sum_{v \in \{\pm 1\}^n} \mathbb{P}_Q[T \text{ rejects } \mathrm{TAS}_v]$$

$$\leq \sum_{v \in \{\pm 1\}^n} \mathbb{P}_Q[T \text{ contains a witness for } \mathrm{TAS}_v]$$

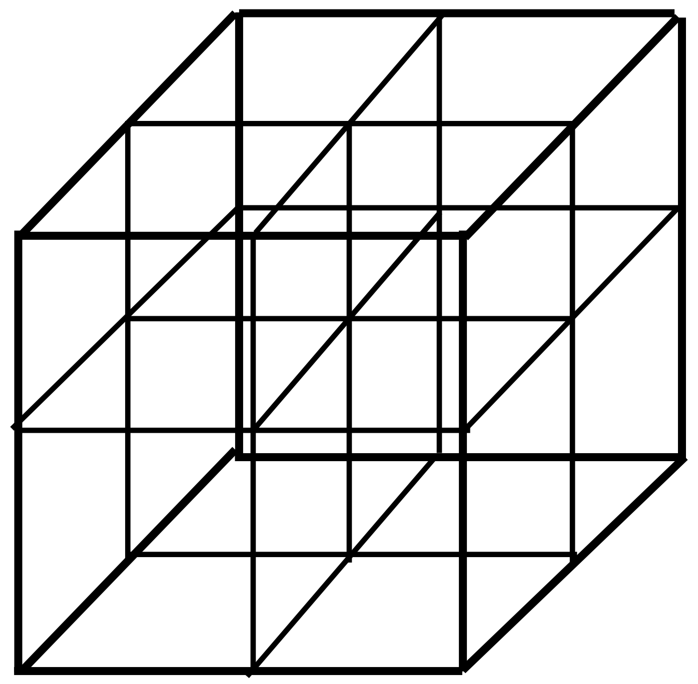
$$= \mathbb{E}_Q[w(Q)] \leq |Q|^2 \cdot 2^{n-0.08n^{1/2}}$$

$$\implies |Q| > 2^{0.03n^{1/2}} \quad \square$$



Future Directions

Black-Blais-Harms **[ITCS 24]**



Computational:

1-sided non-adaptive **query**-based testing: $2^{\widetilde{\Theta}(n^{1/2})}$

Learning and testing with samples: $2^{\widetilde{O}(n^{3/4})}$

Learning and testing with samples: $2^{\Omega(n^{1/2})}$

Structural:

Def: $I(S) = 3^{-n} \cdot$ # edges $(u, v) : u \in S, v \notin S$

All convex sets satisfy $I(S) \leq \widetilde{O}(n^{3/4})$

There exists a convex set with $I(S) \geq \Omega(n^{3/4})$

Thank you!

Questions:

2-sided and/or adaptive testing?

Close this gap

Noise sensitivity of convex sets?

Larger hypergrids?

Is there a formal connection to testing over **continuous** domains?