Slicing the hypercube is not easy

Amir Yehudayoff (Technion)

Gal Yehuda (Technion)

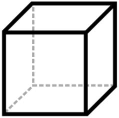
background

the Boolean hypercube can be embedded in Euclidean space

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

edges: line segments $[x, y]$

etc.



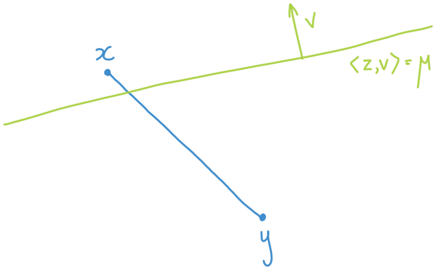
important, useful and interesting

question

how many hyperplanes are needed to slice all edges?

the edge $[x, y]$ is sliced by $\langle z, v\rangle = \mu$ if

$$(\langle x, v\rangle - \mu)(\langle y, v\rangle - \mu) < 0$$



motivations

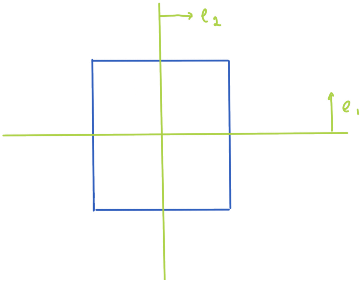
machine learning [O'Neil 70]

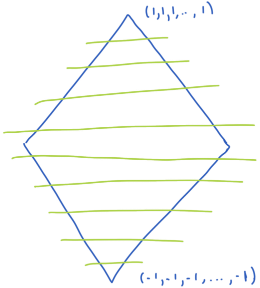
geometry [Grünbaum 72]

computational complexity [Håstad, Paturi-Saks 90, ...]

upper bounds

two constructions of n hyperplanes:





is this optimal?!

upper bounds

Paterson: there are 5 hyperplanes in dimension 6

subadditivity: there are $\lceil \frac{5n}{6} \rceil$ hyperplanes in dimension n

lower bounds

O'Neil:

at least $\Omega(n^{0.5})$ hyperplanes are needed to slice all edges

Emamy-Khansary: at least 4 in dimension 4

Ahlswede-Zhang: at least n if entries are positive

Alon-Bergmann-Coppersmith-Odlyzko, Saks: at least $\frac{n}{2}$ if entries are ± 1

main result

at least $\Omega(n^{0.57})$ hyperplanes are needed to slice all edges

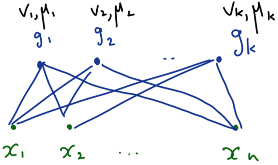
application: threshold circuits for parity

threshold gates compute $x \mapsto \mathrm{sign}(\langle x, v \rangle - \mu)$

threshold circuits are comprised of threshold gates

what is the minimum size of a threshold circuit for parity?

connection: the first layer yields a slicing family



application: threshold circuits for parity

O'Neil: size of first layer in any depth is $\Omega(n^{0.5})$

corollary: size of first layer in any depth is $\Omega(n^{0.57})$

Paturi-Saks, and Impagliazzo-P-S: number of wires in constant-depth

application: covering the cube

minimum number of hyperplanes needed to cover vertices?

what about skew (all entries non-zero) hyperplanes?

application: covering the cube

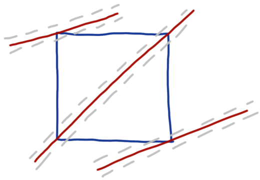
minimum number of skew hyperplanes needed to cover?

Littlewood-Offord, Erdös: at least $\Omega(n^{0.5})$

better lower bounds in special cases

corollary: at least $\Omega(n^{0.57})$

reason: a skew covering family yields a slicing family



plan

every $k \leq n^{0.57}$ hyperplanes must miss an edge

how to locate the missing edge?

randomness?

with $n^{0.51}$ hyperplanes we can slice vast majority of edges

algebra?

topology?

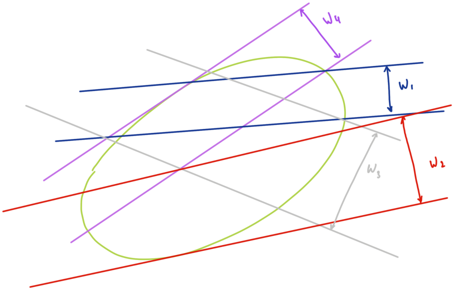
geometry?

how to capture slicing?

part I: geometry

opening move

Tarski's plank problem: what is minimum total width of planks that are needed to cover a convex body?



$w \leq w_1 + w_2 + w_3 + w_4$

opening move: Bang

Bang's theorem: if $p_1, p_2, \ldots$ cover a convex K then

$$\sum_i \text{width}(p_i) \geq \text{width}(K)$$

Ball isolated the following (and used it...)

Bang's lemma: for all $k \times k$ symmetric matrices M with ones on diagonal and $\mu \in \mathbb{R}^k$ and $\theta \in \mathbb{R}$, there **exists** $\epsilon \in \{\pm 1\}^k$ so that for all $i \in [k]$,

$$|\theta(M\epsilon)_i - \mu_i| \geq \theta$$

part II: antichains

antichains of edges

if the entries in $v \in \mathbb{R}^n$ are positive then the set of edges that are sliced by $\langle z, v \rangle = \mu$ is an antichain

for every chain

$$(c_0, c_1), (c_1, c_2), \ldots, (c_{n-1}, c_n)$$

of edges with $c_0 = (-1, -1, \ldots, -1)$ and $c_n = (1, 1, \ldots, 1)$, there is at most a single edge (c_j, c_{j+1}) that is sliced

O'Neil's bound

theorem

the fraction of edges that are sliced by a hyperplane is $O(\frac{1}{\sqrt{n}})$

proof

the edges sliced by a hyperplane form an (oriented) antichain

Baker proved that antichains are small

antichains of vertices

vertex antichain = no strict pairwise inclusions

identify $\{\pm 1\}^n$ and $\{0,1\}^n$

Sperner's theorem:

the maximum size of an antichain is $\max_\ell \binom{n}{\ell} \leq O(\frac{2^n}{\sqrt{n}})$

fundamental in extremal set theory with many applications

antichains of vertices

the Lubell-Yamamoto-Meshalkin inequality:

if A is an antichain then

$$\sum_{\ell} \frac{|A_\ell|}{\binom{n}{\ell}} \leq 1$$

where $A_\ell = \{a \in A : |a| = \ell\}$

stronger and more useful than Sperner's theorem

generalizations

there are many ...

need: general products measures

Aizenman-Germinet-Klein-Warzel generlized

Bernouli decomposition

bound is not sharp

measures of antichains

theorem

for every non trivial product measure P on $\{0,1\}^n$ and for every antichain A

$$\Pr[z \in A] \leq \max_{\ell} \Pr[|z| = \ell]$$

and

$$\sum_{\ell} \Pr[z \in A | |z| = \ell] \leq 1$$

where $z \sim P$

generalizes both Sperner and LYM

what is special about product measures?

lemma

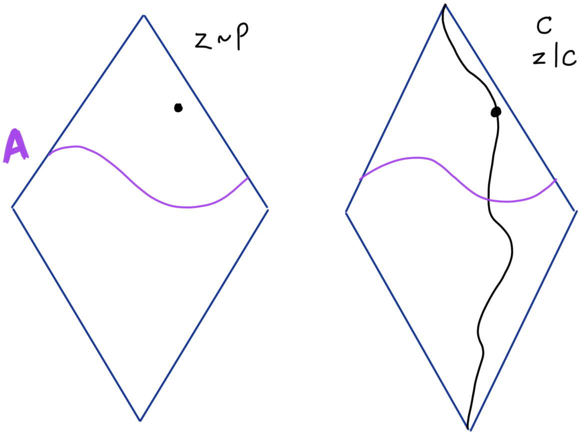
for every non trivial product measure P there is a way to choose a chain

$$\emptyset = c_0 \subset c_1 \subset c_2 \subset \cdots \subset c_n = [n]$$

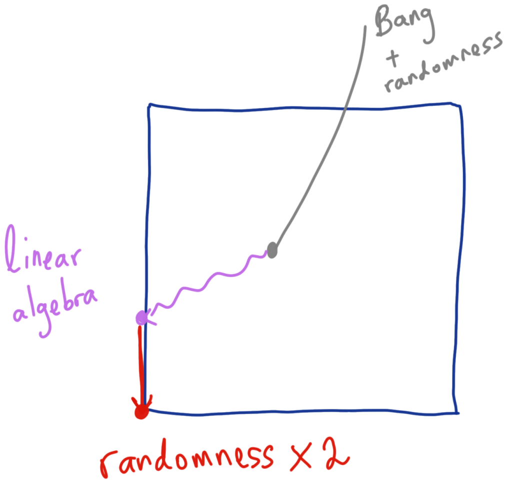
so that $c_\ell \sim P|\{\text{size} = \ell\}$ for all ℓ

our proof is technical

sketch



outline



summary

theorem

at least $\Omega(n^{0.57})$ hyperplanes are needed to slice all edges

main ideas

Bang's lemma

context for Sperner's theorem and LYM inequality

strong anti-concentration for many scales

structure of normal vectors

rounding using linear algebra

concentration of measure

part III: strong anti-concentration

many scales

the vector $v \in \mathbb{R}^n$ has **many scales** if it can be partitioned to $v^{(1)}, v^{(2)}, \ldots, v^{(S)}$ with $S = n^{0.001}$ so that for all s

$$\|v^{(s+1)}\| \leq \frac{\|v^{(s)}\|}{100}$$

the **minimum scale** of v is $\delta = \|v^{(S)}\|$

lemma

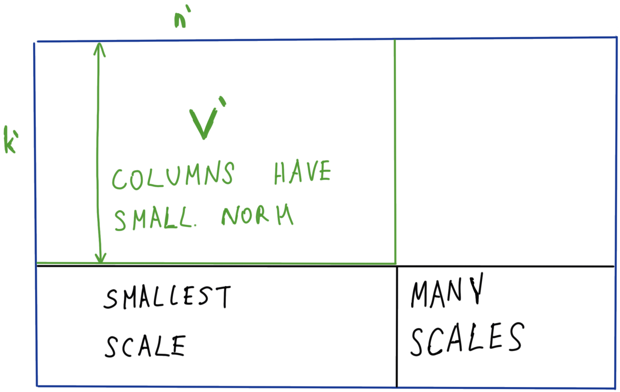
if v has many scales then for all a

$$\Pr_{x \sim \{\pm 1\}^n}[|\langle x, v\rangle - a| < n\delta] \leq \exp(-\Omega(S))$$

part IV: structure

structure

arrange the normals as rows of $k \times n$ matrix V



part V: rounding

rounding

Bang's lemma:

$\exists\ u \in \mathbb{R}^n$ with $\|u\|_\infty \leq 1$ so that $\langle u, v_i \rangle$ is far from μ_i for all i

need to round u to a vertex

lemma

there is $w \in \mathbb{R}^n$ so that

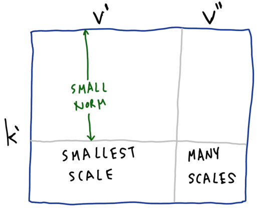
— $\langle w, v_i \rangle = \langle u, v_i \rangle$ for all i

— $\|w\|_\infty \leq 1$

— $|w_j| = 1$ for at least $n - k$ values of j

proof outline

let V be the matrix whose rows are the k normals



structure: write $v_i = (v_i', v_i'')$

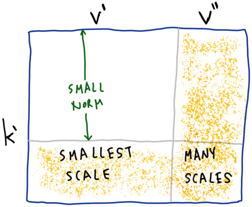
many scales: there is $x'' \sim \{\pm 1\}^{n''}$ so that for all $i > k'$,

$$\langle x'', v_i'' \rangle \text{ is very far from } \mu_i$$

done with rows $i > k'$

chose $x'' \sim \{\pm 1\}^{n''}$

done with rows $i > k'$

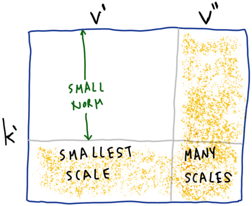


for $i \leq k'$, set $\sigma_i = \mu_i - \langle x'', v_i'' \rangle$

Bang's lemma: $\exists u \in \mathbb{R}^{n'}$ so that $\|u\|_\infty \leq 1$ and

$$\langle u, v_i' \rangle \text{ is far from } \sigma_i \text{ for all } i \leq k'$$

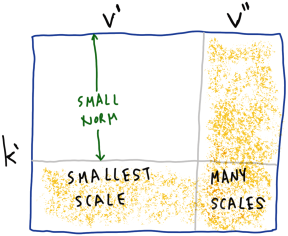
$M = V'V'^T$ and $u = \theta V' \epsilon$ with $\epsilon \in \{\pm 1\}^{k'}$ and $\theta \approx n^{-0.01}$



$\exists u \in \mathbb{R}^{n'}$ so that $\|u\|_\infty \leq 1$ and

$$\langle u, v_i' \rangle \text{ is far from } \sigma_i \text{ for all } i \leq k'$$

problems: u is not a vertex & need an edge



rounding: round u to an almost vertex w

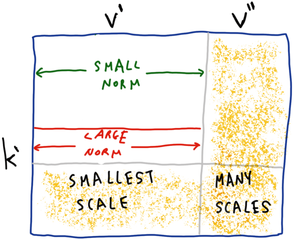
choose $x' \sim P$ so that $\mathbb{E}\, x'_j = w_j$ for all j
(at most k entries has randomness)

let y be a random* neighbor of $x = (x', x'')$

chose $[x, y]$ carefully at random

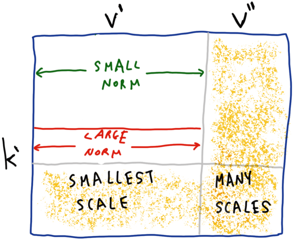
structure: most rows $i \leq k'$ have small norm

all columns have small norm



chose $[x, y]$ carefully at random

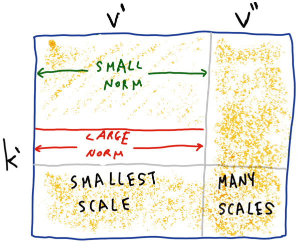
structure: most rows $i \leq k'$ have small norm



Bernstein: $\langle x', v_i \rangle$ is far from σ_i for most rows $i \leq k'$

$\mathbb{E}\langle x', v_i' \rangle = \langle w, v_i' \rangle = \langle u, v_i' \rangle$ is far from σ_i

done with most rows $i \leq k'$



measure of antichains: deal with the few final rows

done

thank you!