On the Power of Semidefinite Programming Hierarchies

Prasad Raghavendra
Georgia Institute of Technology,
Atlanta, GA.

David Steurer
Microsoft Research New England
Cambridge, MA

Overview

- Background and Motivation
- Introduction to SDP Hierarchies (Lasserre SDP hierarchy)
- Rounding SDP hierarchies via Global Correlation.

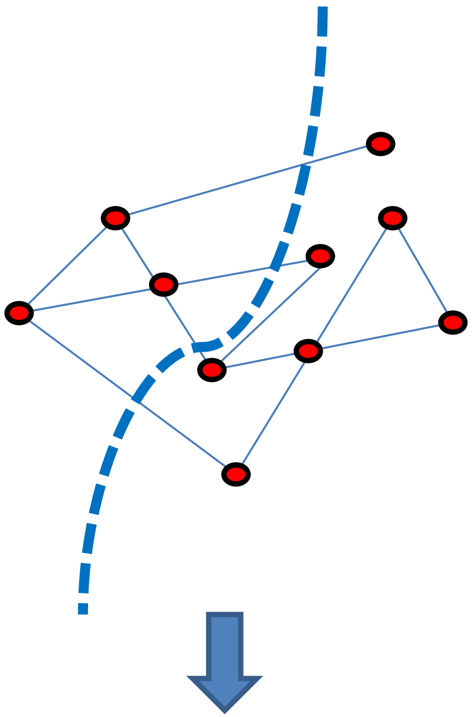
BREAK

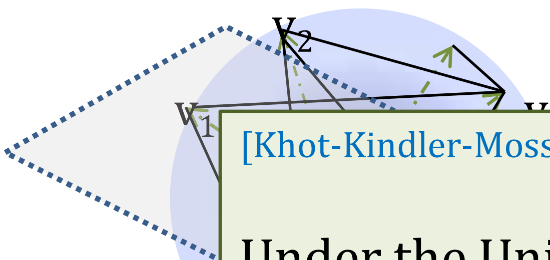
- Graph Spectrum and Small-Set Expansion.
- Sum of Squares Proofs.

Background and Motivation

Max-Cut Problem

Max-Cut





Input: A graph G

Find: A cut with maximum number of crossing edges

Semidefinite Program for MaxCut:

[Goemans-Williamson 94]

Embed the graph on the

N - dimensional unit ball,

Maximizing

¼ (Average Squared Length of the edges)

[Khot-Kindler-Mossel-O'Donnell]

Under the Unique Games Conjecture,

Goemans-Williamson SDP yields the optimal approximation ratio for MaxCut.

Motivation

Unique Games Conjecture (UGC) [Khot'02]

For every $\varepsilon > 0$, the following is **NP**-hard:

UG(ε)

Given: system of equations $x_i - x_j = c \bmod k$ (say $\mathrm{k} = \log n$)

Distinguish:

YES: at least $1 - \varepsilon$ of equations satisfiable

NO: at most ε of equations satisfiable

Assuming the Unique Games Conjecture,

A simple semidefinite program (Basic-SDP) yields the optimal approximation ratio for

Constraint Satisfaction Problems [Raghavendra`08][Austrin-Mossel]

MAX CUT [Khot-Kindler-Mossel-ODonnell][Odonnell-Wu]

MAX 2SAT [Austrin07][Austrin08]

Metric Labeling Problems [Manokaran-Naor-Raghavendra-Schwartz`08]

MULTIWAY CUT, 0-EXTENSION

Ordering CSPs [Charikar-Guruswami-Manokaran-Raghavendra-Hastad`08]

MAX ACYCLI

Is the conjecture true?

Many many ways to disprove the conjecture!

Find a better algorithm for any one of these problems.

Kernel Clustering Problems [Khot-Naor`08,10]

Grothendieck Problems [Khot-Naor, Raghavendra-Steurer]

Question I:

Could some small LINEAR PROGRAM
give a better approximation for MaxCut or Vertex Cover
thereby disproving the UGC?

Probably Not!

Question II:

Could some small SEMIDEFINITE PROGRAM
give a better approximation for MaxCut or Vertex Cover
thereby disproving the UGC?

We don’t know.

[Charikar-Makarychev-Makarychev][Schoenebeck-Tulsiani]

For MaxCut, for several classes of linear programs, exponential sized linear programs are necessary to even beat the trivial ½ approximation!

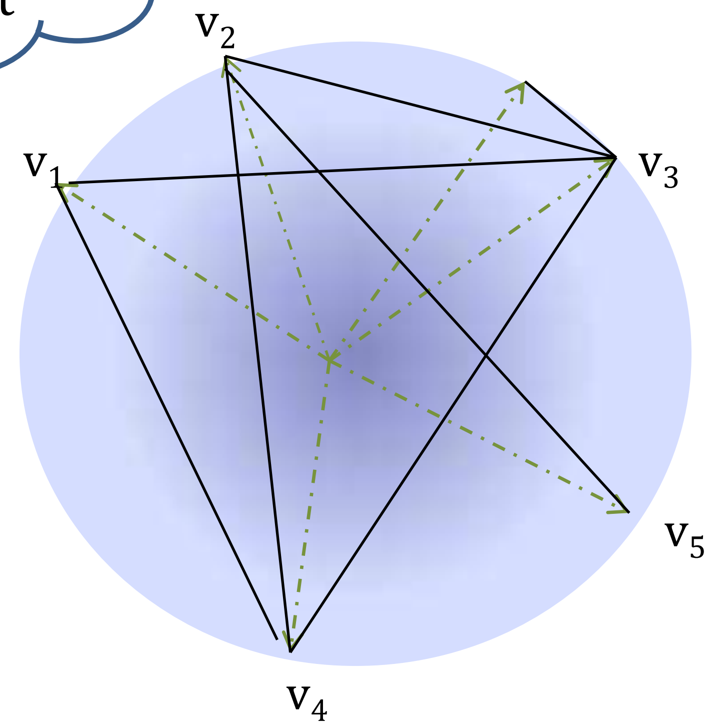
Max Cut SDP:

The Simplest Relaxation for MaxCut

Embedd the graph on the **N** - dimensional unit ball,

Maximizing

¼ (Average squared length of the edges)



In the integral solution, all the vectors v_i are 1,-1. Thus they satisfy additional constraints

For example : $(v_i - v_j)^2 + (v_j - v_k)^2 \geq (v_i - v_k)^2$

(the triangle inequality)

Does adding triangle inequalities improve approximation ratio?

(and thereby disprove UGC!)

[Arora-Rao-Vazirani 2002]

For Sparsest Cut,

SDP with triangle inequalities gives $O(\sqrt{\log n})$ approximation.

An $O(1)$-approximation would disprove the UGC!

[Goemans-Linial Conjecture 1997]

SDP with triangle inequalities would yield $O(1)$-approximation for Sparsest Cut.

[Khot-Vishnoi 2005]

SDP with triangle inequalities DOES NOT give O(1) approximation for Sparsest Cut

SDP with triangle inequalities DOES NOT beat the Goemans-Williamson 0.878 approximation for Max Cut

Until 2009:

Adding a simple constraint on every 5 vectors
could yield a better approximation for MaxCut, and disproves UGC!

Building on the work of [Khot-Vishnoi],

[Khot-Saket 2009][Raghavendra-Steurer 2009]

Adding all *valid local constraints* on at most $2^{(\log\log n)^\wedge 1/4}$ vectors to the simple SDP
DOES NOT improve the approximation ratio for MaxCut

[Barak-Gopalan-Hastad-Meka-Raghavendra-Steurer 2009]

As of Now:
Change $2^{(\log\log n)^\wedge 1/4}$ to $\exp(2^{(\text{poly}(\log\log n))})$ in the above result.

A natural SDP of size $O(n^{16})$ (the 8^{th} round of Lasserre hierarchy) could disprove the UGC.

[Barak-Brandao-Harrow-Kelner-Steurer-Zhou 2012] (this conference)
8^{th} round of Laserre hierarchy solves all known instances of Unique Games.

Why play this game?

Connections between SDP hierarchies, Spectral Graph Theory and Graph Expansion.

New algorithms based on SDP hierarchies.

[Raghavendra-Tan]

Improved approximation for MaxBisection using SDP hierarchies

[Barak-Raghavendra-Steurer]

Algorithms for 2-CSPs on low-rank graphs.

New Gadgets for Hardness Reductions:

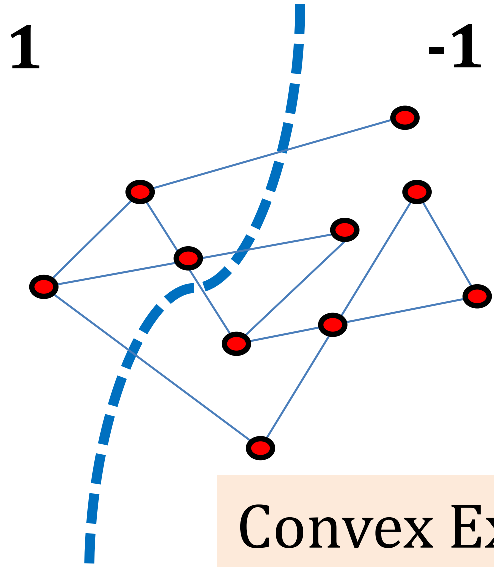
[Barak-Gopalan-Hastad-Meka-Raghavendra-Steurer]

A more efficient long code gadget.

Deeper understanding of the UGC – why it should be true if it is.

Introduction to SDP Hierarchies
(Lasserre SDP hierarchy)

Revisiting MaxCut Semidefinite Program



Integer Program:

Domain: $x_1, x_2, x_3, \ldots, x_n \in \{-1,1\}$

(x_i for vertex i)

Maximize:

$$\frac{1}{4}\sum_{(i,j)\in E}\left(x_i - x_j\right)^2$$

(Number of Edges Cut)

Convex Extension of Integer Program:

Domain: Probability distributions μ over assignments $x \in \{-1,1\}^n$

Maximize:

$$E_{x\sim\mu}\,\frac{1}{4}\sum_{(i,j)\in E}\left(x_i - x_j\right)^2$$

(Expected Number of Edges Cut under μ)

Good News: Convex program that *exactly* captures the MaxCut problem.

Bad News: Size of the convex extension is too large (exponential in n)

Representing a probability distribution μ over $\{-1,1\}^n$ requires exponentially many variables, μ_x for each $x \in \{-1,1\}^n$

Using Moments

Convex Extension of Integer Program:

Domain: Probability distributions μ over assignments $x \in \{-1,1\}^n$

Maximize:

$$E_{x\sim\mu} \frac{1}{4} \sum_{(i,j)\in E} \left(x_i - x_j\right)^2$$

(Expected Number of Edges Cut under μ)

Idea: Instead of finding the entire prob. distribution μ, just find its low degree moments

$$= \frac{1}{4} \sum_{(i,j)\in E} E_{x\sim\mu} \left(x_i - x_j\right)^2$$

$$= \frac{1}{4} \sum_{(i,j)\in E} \left(E_{x\sim\mu} x_i^2 + E_{x\sim\mu} x_j^2 - 2E_{x\sim\mu} x_i x_j\right)$$

$$= \frac{1}{4} \sum_{(i,j)\in E} \left(M_{ii} + M_{jj} - 2M_{ij}\right)$$

Moment Variables:

Let $M_i \stackrel{\text{def}}{=} E_{x\sim\mu}\left[x_i\right]$

$M_{ii} \stackrel{\text{def}}{=} E_{x\sim\mu}\left[x_i^2\right]$

$M_{ij} \stackrel{\text{def}}{=} E_{x\sim\mu}\left[x_i x_j\right]$

$M_{ijk} \stackrel{\text{def}}{=} E_{x\sim\mu}\left[x_i x_j x_k\right]$

…

…

…

$M_S \stackrel{\text{def}}{=} E_{x\sim\mu}\left[\prod_{i\in S} x_i\right]$

for a multiset $S \subseteq \{1,..,n\}$, $|S| \le d$

Constraints on Moments

For each i, since $x_i \in \{-1,1\}$,

$x_i^2 = 1$ always *so,*

$$E_{x\sim\mu}\left[x_i^2\right] = 1$$

$x_i^2 x_j x_k = x_j x_k$ always so,

$$E_{x\sim\mu}\left[x_i^2 x_j x_k\right] = x_j x_k$$

Constraint:
More generally, for every multiset $S, |S| \leq d$

$$M_S = M_{odd(S)} \text{ where}$$

$odd(S)$ = set of elements in S that appear an odd number of times.

Constraint: For each i,

$$M_{ii} = 1$$

For each i, j, k

$$M_{\{i,i,j,k\}} = M_{jk}$$

Constraint:

All valid moment equalities that hold for all distributions μ over $\{-1,1\}^n$

Constraints on Moments

d-round Lasserre SDP Hierarchy:

Variables: All moments

$$\{M_S\}$$

up to degree d of the unknown distribution μ over assignments $\{-1,1\}^n$

Maximize:

$$\frac{1}{4}\sum_{(i,j)\in E}\left(M_{ii}+M_{jj}-2M_{ij}\right)$$

$$= E_{x\sim\mu}\frac{1}{4}\sum_{(i,j)\in E}\left(x_i-x_j\right)^2$$

(Expected Number of Edges Cut under μ)

Constraint: For each i,

$$M_{ii} = 1$$

Constraint:
Use $x_i^2 = 1$ always for all i,

and include ALL valid equalities for moments M_S that hold for all distributions over $\{-1,1\}^n$

Constraint:

$$M_{\{1,1,2,2\}} - 6M_{\{1,2,3\}} + 9M_{33} \geq 0$$

Constraint: For every real polynomial $p(x_1, x_2, .., x_n)$ of degree at most $\frac{d}{2}$,

$$p^2 \circ M \geq 0$$

(basically $E_{x\sim\mu}p^2(x) \geq 0$)

Degree d = 2
(Goemans-Williamson SDP)

Degree 2 SOS SDP Hierarchy:

Variables:
Moments $\{M_{ij} | i,j \in \{1,\dots,n\}\}$
up to degree 2 of the unknown
distribution μ over assignments $\{-1,1\}^n$

Maximize:

$$\frac{1}{4}\sum_{(i,j)\in E}\left(M_{ii}+M_{jj}-2M_{ij}\right)$$

$$= E_{x\sim\mu}\frac{1}{4}\sum_{(i,j)\in E}\left(x_i-x_j\right)^2$$

(Expected Number of Edges Cut under μ)

Constraint: For each i,

$$M_{ii} = 1$$

Constraint: For every real *linear polynomial* $p(x_1, x_2, \dots, x_n)$,

$$p^2 \circ M \geq 0$$

So for all $p(x) = \sum_i c_i x_i$, we have,

$$\sum_{i,j} c_i c_j M_{ij} \geq 0$$

Constraint: Use $x_i^2 = 1$ always for all i, and include ALL valid equalities for moments M_S that hold for all distributions over $\{-1,1\}^n$

Constraint: For every real polynomial $p(x_1, x_2, \dots, x_n)$ of degree at most $\frac{d}{2}$,

$$p^2 \circ M \geq 0$$

(basically $E_{x\sim\mu}p^2(x) \geq 0$)

(basically $E_{x\sim\mu}p^2(x) \geq 0$)

Goemans-Williamson SDP

Variables:
Moments $\{M_{ij} | i,j \in \{1,\ldots,n\}\}$
up to degree 2 of the unknown distribution μ over assignments $\{-1,1\}^n$

Maximize:

$$\frac{1}{4}\sum_{(i,j)\in E}\left(M_{ii}+M_{jj}-2M_{ij}\right)$$

$= E_{x\sim\mu}\frac{1}{4}\sum_{(i,j)\in E}\left(x_i - x_j\right)^2$

(Expected Number of Edges Cut under μ)

Arrange the variables in a matrix,

$$M = \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix}$$

Constraint: For each i,

$$M_{ii} = 1$$

``Diagonal entries of M are equal to 1''

Constraint: For every real *linear polynomial* $p(x_1, x_2, .., x_n)$,

$$p^2 \circ M \geq 0$$

So for all $p(x) = \sum_i c_i x_i$ we have,

$$\sum_{i,j} c_i c_j M_{ij} \geq 0$$

(basically $E_{x\sim\mu} p^2(x) \geq 0$)

``Matrix M is positive-semidefinite'

Positive Semidefiniteness (where are the vectors?)

Constraint: For every real *linear polynomial*

$$p(x_1, x_2, .., x_n), = \sum_i c_i x_i$$

we have,

$$\sum_{i,j} c_i c_j M_{ij} \geq 0$$

(basically $E_{x\sim\mu} p^2(x) \geq 0$)

$\Longleftrightarrow$

Positive Semidefiniteness:

With $M = \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix}$

For all real vectors $c \in \boldsymbol{R^n}$, we have,

$$c^T M c \geq 0$$

$$\Updownarrow$$

Cholesky Decomposition:

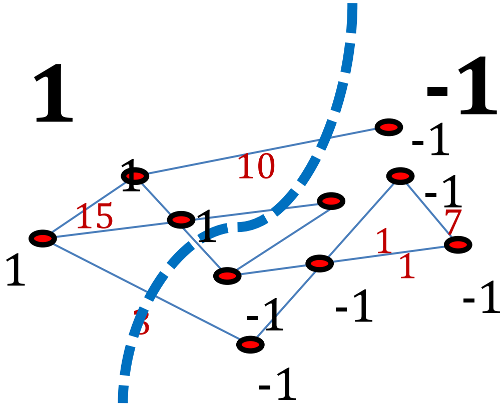
There exists vectors $\{v_i\}$ such that

$$\langle v_i, v_j \rangle = \mathrm{M_{ij}}$$

For degree d-Lasserre SDP,

the moments are appropriately arranged to give a p.s.d. matrix.

Cheat Sheet: d-round Lasserre SDP



X_1 X_2 X_3 X_4 .... X_{15} ....

1 -1 1 -1 .... 1 1 1 -1 1 1 -

1 -1 -1 -1 .... 1 1 1 -1 1 1 1

1 -1 -1 -1 .... 1 1 1 -1 1 1 -

1 -1 1 -1 .... 1 1 1 -1 1 1 -

1

....

1 1 1 -1 .... 1 1 1 -1 1 1 -1

...

1 1 1 -1 .... 1 1 1 -1 1 1 -1

Fictitious Distribution over assignments

Local distribution μ_S

For any subset S of $\leq d$ vertices,

A local distribution μ_S over {+1,-1} assignments to the set S

All moments up to degree d

➔ Specify every marginal on up to d variables.

Conditioned SDP Solution

For any subset S of $k \leq d$ vertices, and an assignment α in $\{-1,1\}^k$,

We can condition the SDP solution to the event that S is assigned α and get a d-k round SDP solution.

Rounding SDP Hierarchies

Subexponential Algorithm for Unique Games

UG(ε) in time $\exp\left(n^{\varepsilon^{1/3}}\right)$ via level-$n^{\varepsilon^{1/3}}$ SDP relaxation

[Arora-Barak-S.'10, Barak-Raghavendra-S.'11]

Contrast

many NP-hard approximation problems require exponential time

(assuming 3-SAT does)

[...,Moshkovitz-Raz]

often these lower bounds are known *unconditionally* for SDP hierarchies

[Schoenebeck, Tulsiani]

→ separation of UG from known NP-hard approximation problems

Subexponential Algorithm for Unique Games

UG(ε) in time $\exp\left(n^{\varepsilon^{1/3}}\right)$ via level-$n^{\varepsilon^{1/3}}$ SDP relaxation

General framework for rounding SDP hierarchies (not restricted to Unique Games)

[Barak-Raghavendra-S.'11, Guruswami-Sinop'11]

Potentially applies to wide range of "graph problems"

Examples: MAX CUT, SPARSEST CUT, COLORING, MAX 2-CSP

Some more successes (polynomial time algorithms)

Approximation scheme for general MAX 2-CSP [Barak-Raghavendra-S.'11]

on constraint graphs with $O(1)$ significant eigenvalues

Better 3-COLORING approximation for some graph families [Arora-Ge'11]

Better approximation for MAX BISECTION (general graphs) [Raghavendra-Tan'12]

[Austrin-Benabbas-Georgiou'12]

Subexponential Algorithm for Unique Games

UG(ε) in time $\exp\left(n^{\varepsilon^{1/3}}\right)$ via level-$n^{\varepsilon^{1/3}}$ SDP relaxation

General framework for rounding SDP hierarchies (not restricted to Unique Games)

[Barak-Raghavendra-S.'11, Guruswami-Sinop'11]

Potentially applies to wide range of "graph problems"

Examples: MAX CUT, SPARSEST CUT, COLORING, MAX 2-CSP

Key concept: global correlation

Interlude: Pairwise Correlation

Two jointly distributed random variables X and Y

Correlation measures dependence between X and Y

Does the distribution of X change if we condition Y?

Examples:

(Statistical) distance between $\{X, Y\}$ and $\{X\}\{Y\}$

Covariance $\mathbf{E}\, XY - (\mathbf{E}\, X)(\mathbf{E}\, Y)$ (if X and Y are real-valued)

Mutual Information $\mathrm{I}(X, Y) = H(X) - H(X|Y)$

entropy lost due to conditioning

random variables $X_1, \ldots, X_n$ over $\mathbb{Z}_k$

$\Pr\left(X_i - X_j = c\right) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Sampling
~~Rounding~~ problem

degree-ℓ moments of a distribution over assignments with expected value $\geq 1 - \varepsilon$

Given

UG instance + ~~level-ℓ SDP solution with value $\geq 1 - \varepsilon$~~ $(\ell = n^{O(\varepsilon^{1/3})})$

Sample

distribution over assignments with expected value $\geq \varepsilon$

similar (?)

More convenient to think about actual distributions instead of SDP solutions

But: proof should only "use" linear equalities satisfied by these moments and *certain* linear inequalities, namely non-negativity of squares

(Can formalize this restriction as proof system → next talk)

Sampling by conditioning

Pick an index j

Sample assignment a for index j from its marginal distribution $\{X_j\}$

Condition distribution on this assignment, $X'_i \coloneqq \{ X_i \mid X_j = a \}$

If we condition n times, we correctly sample the underlying distribution

Issue: after conditioning step, know only degree $\ell - 1$ moments (instead of degree ℓ)

Hope: need to condition only a small number of times; then do something else

How can conditioning help?

How can conditioning help?

Allows us to assume: distribution has *low global correlation*

$$\mathbf{E}_{i,j}\mathrm{I}\left(X_i, X_j\right) \leq O_k(1) \cdot {}^{1}\!/_{\ell}$$

typical pair of variables
almost pairwise independent

Claim: general cases reduces to case of *low global correlation*

Proof:

Idea: significant global correlation → conditioning decreases entropy

Potential function $\Phi = \mathbf{E}_i H(X_i)$

Can always find index j such that for $X_i' := \{X_i | X_j\}$

$$\Phi - \Phi' \geq \mathbf{E}_i\, H(X_i) - \mathbf{E}_i\, H\left(X_i \middle| X_j\right) = \mathbf{E}_i\, I\left(X_i, X_j\right) \geq \mathbf{E}_{i,j}\, I\left(X_i, X_j\right)$$

Potential can decrease $\leq \ell/2$ times by more than $O_k(1/\ell)$ ■

How can conditioning help?

Allows us to assume: distribution has *low global correlation*

$$\mathbf{E}_{i,j} \mathrm{I}\left(X_i, X_j\right) \leq O_k(1) \cdot {}^{1}\!/_{\ell}$$

typical pair of variables
almost pairwise independent

How can low global correlation help?

How can low global correlation help?

$$\mathbf{E}_{i,j}\mathrm{I}(X_i, X_j) \leq {}^1/_\ell$$

For some problems, this condition alone gives improvement over BASIC SDP

Example: MAX BISECTION [Raghavendra-Tan'12, Austrin-Benabbas-Georgiou'12]

hyperplane rounding gives near-bisection if global correlation is low

How can low global correlation help?

$\mathbf{E}_{i,j} \mathrm{I}(X_i, X_j) \le {}^{1}/_{\ell}$

For Unique Games

random variables $X_1, \dots, X_n$ over $\mathbb{Z}_k$

$\Pr(X_i - X_j = c) \ge 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Extreme cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

all variables have uniform marginal distribution & $\exists$ partition:

ℓ equal-sized components

X_p X_q ...

$\mathrm{I}(X_p, X_q) = 0$

→ inter-component constraint cannot be typical
→ $\le \varepsilon$ fraction of constraints are inter-component

How can low global correlation help?

$\mathbf{E}_{i,j}\mathrm{I}(X_i, X_j) \le {}^{1}/_{\ell}$

For Unique Games

random variables $X_1, \dots, X_n$ over $\mathbb{Z}_k$

$\Pr(X_i - X_j = c) \ge 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Only
~~Extreme~~ cases with low global correlation

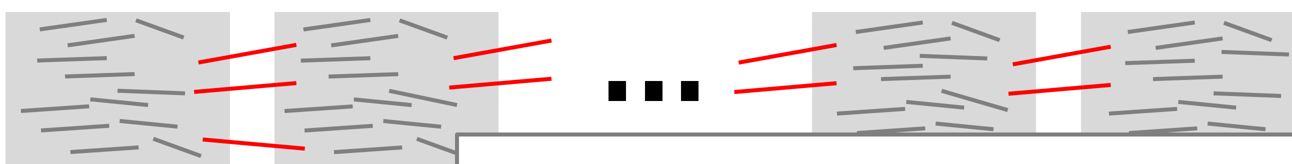
1) no entropy: all variables are fixed

2) many small independent components:

Show: no other cases are possible! (informal)

all variables have uniform marginal distribution & $\exists$ partition:

ℓ equal-sized components



$\mathrm{I}(X_p, X_q) = 0$

→ inter-component constraint cannot be typical
→ $\le \varepsilon$ fraction of constraints are inter-component

Idea: round components independently & recurse on them

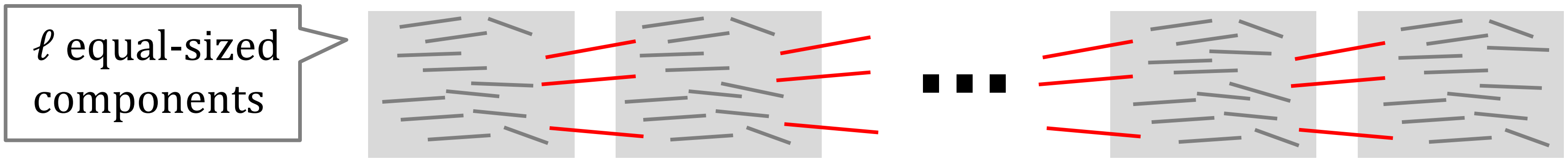
How many edges ignored in total? (between different components)

We chose $\ell = n^{\beta}$ for $\beta \gg \varepsilon$
→ each level of recursion decrease component size by factor $\geq n^{\beta}$
→ at most $1/\beta$ levels of recursion
→ total fraction of ignored edges $\leq \varepsilon/\beta \ll 1$

→ $2^{n^{\beta}}$-time algorithm for UG(ε)

2) many small independent components:

all variables have uniform marginal distribution & ∃ partition:



How can low global correlation help?

$\mathbf{E}_{i,j} \mathrm{I}(X_i, X_j) \leq 1/\ell$

For Unique Games

random variables $X_1, \ldots, X_n$ over $\mathbb{Z}_k$

$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Only

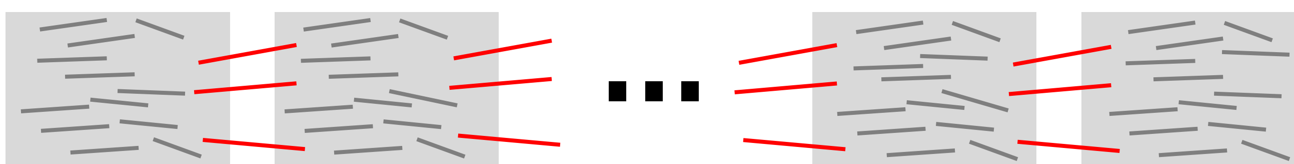
~~Extreme~~ cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

all variables have uniform marginal distribution & $\exists$ partition:

ℓ equal-sized components



Suppose: random variables $X_1, \dots, X_n$ over $\mathbb{Z}_k$ with uniform marginals

$\Pr\left(X_i - X_j = c\right) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

global correlation $\leq 1/n^{2\beta}$

Then: $\exists\, S \subseteq [n]$. $|S| \leq n^{1-\beta}$ & all constraints touching S stay inside of S except for an $O\left(\sqrt{\varepsilon/\beta}\right)$ fraction

(in constraint graph, S has low expansion)

Proof: Define $\mathrm{Corr}\left(X_i, X_j\right) = \max_c \Pr\left(X_i - X_j = c\right)$

Correlation Propagation

For random walk $i \sim j_1 \sim \cdots \sim j_t$ of length t in constraint graph

$\mathrm{Corr}(X_i, X_{j_t}) \geq (1-\varepsilon)^t$

$\mathrm{Corr}\left(X_i, X_{j_t}\right) \gtrsim \Pr\left(X_i - X_{j_1} = c_1\right) \cdots \Pr\left(X_i - X_{j_t} = c_t\right)$

proof uses non-negativity of squares (sum-of-squares proof)

→ works also for SDP hierarchy

■

Suppose: random variables $X_1, \dots, X_n$ over $\mathbb{Z}_k$ with uniform marginals

$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

global correlation $\leq 1/n^{2\beta}$

Then: $\exists\, S \subseteq [n]$. $|S| \leq n^{1-\beta}$ & all constraints touching S stay inside of S except for an $O\left(\sqrt{\varepsilon/\beta}\right)$ fraction

(in constraint graph, S has low expansion)

Proof: Define $\mathrm{Corr}(X_i, X_j) = \max_c \Pr(X_i - X_j = c)$

Correlation Propagation $t = {}^{\beta}/_{\varepsilon} \cdot \log n$

For random walk $i \sim j_1 \sim \cdots \sim j_t$ of length t in constraint graph

$\mathrm{Corr}(X_i, X_{j_t}) \geq (1-\varepsilon)^t \quad \geq 1/n^{\beta}$

low global correlation

On the other hand, $\mathrm{Corr}(X_i, X_j) \leq 1/n^{2\beta}$ for typical j

→ random walk from i doesn't mix in t-steps (actually far from mixing)

→ exist small set S around i with low expansion

■

Suppose: random variables $X_1, \ldots, X_n$ over $\mathbb{Z}_k$ with uniform marginals

$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

global correlation $\leq \cancel{1/n^{2\beta}}$ $1/\ell$

Then: constraint graph has ℓ eigenvalues $\geq 1 - \varepsilon$

Proof: a graph has ℓ eigenvalues $\geq \lambda$ $\Leftrightarrow$ $\exists$ vectors $v_1, \ldots, v_n$

(local: typical edge) $\mathbf{E}_{i \sim j} \langle v_i, v_j \rangle \geq \lambda$

(global: typical pair) $\mathbf{E}_{p,q} \langle v_p, v_q \rangle^2 \leq 1/\ell$

$\mathbf{E}_i \|v_i\|^2 = 1$

→ For graphs with $< \ell$ such eigenvalues, algorithm runs in time n^{ℓ}

o(n) ≤ 0.1

How large does ℓ have to be to guarantee a very small set with low expansion ?

Improving $\ell = n^{\varepsilon}$ to $\ell = n^{o(1)}$ would refute Small-Set Expansion Hypothesis

(closely related to UGC)

Thanks!

On the Power of Semidefinite Programming Hierarchies

Prasad Raghavendra
Georgia Institute of Technology,
Atlanta, GA.

David Steurer
Microsoft Research New England
Cambridge, MA

Overview

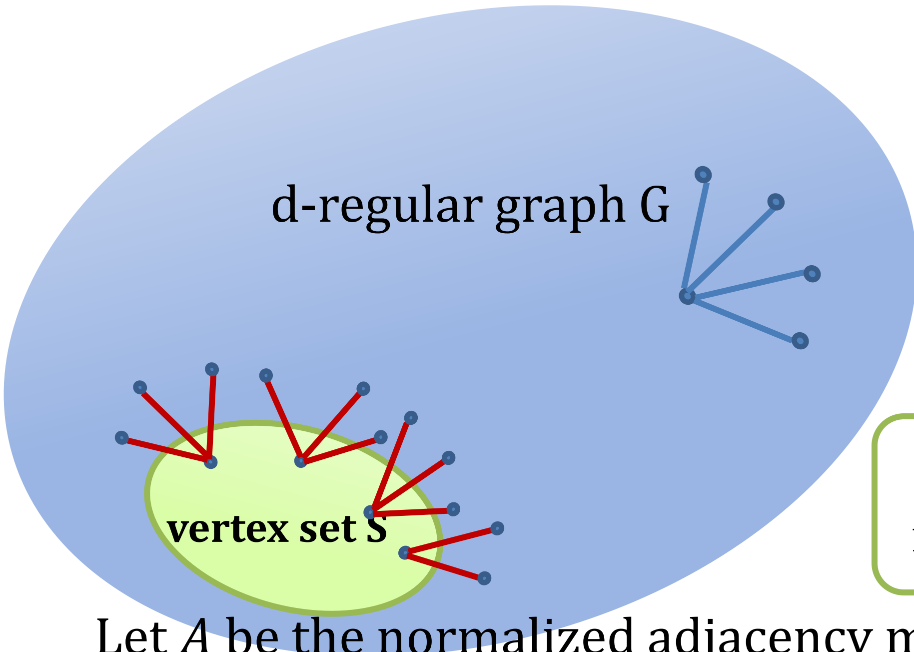
- Background and Motivation
- Introduction to SDP Hierarchies (Lasserre SDP hierarchy)
- Rounding SDP hierarchies via Global Correlation.

BREAK

- Graph Spectrum and Small-Set Expansion.
- Sum of Squares Proofs.

Graph Spectrum
&
Small-Set Expansion

A Question in Spectral Graph Theory



$$expansion(S) = \frac{\#\ edges\ leaving\ S}{d\ |S|}$$

A random neighbor of a random vertex in S is outside of S with probability $expansion(S)$

Let A be the normalized adjacency matrix of the d-regular graph G,

(all entries are 0 or $\frac{1}{d}$)

A has n eigenvalues $\lambda_1 = 1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N$

Qu

matrix A have?

Def (Small Set Expander):

(say $\delta = 10d^6$)

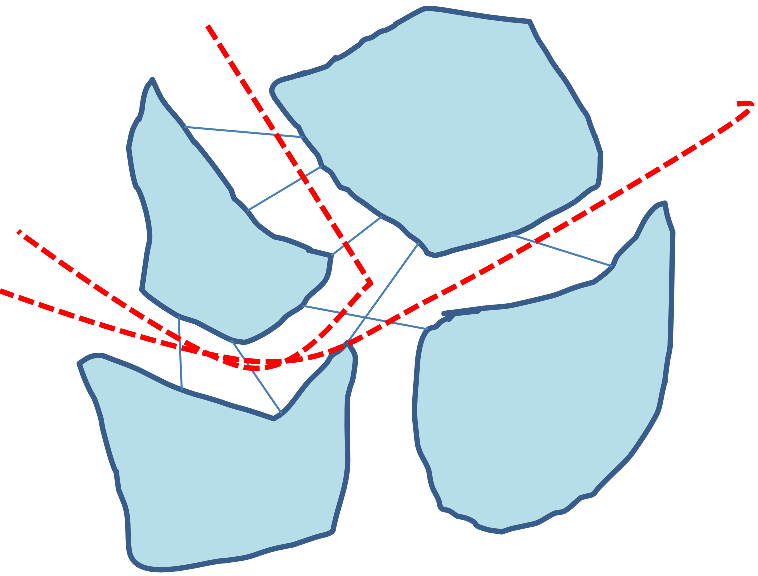
A regular graph G is a δ-small set expander

if for every set $S \subset V$,

$$|S| \leq \delta N \qquad \Rightarrow \qquad expansion(S) \geq \frac{1}{2}$$

A Question in Spectral Graph Theory

d-regular graph G



Def (Small Set Expander):
A regular graph G is a δ-small set expander if for every set $S \subset V$

$$|S| \leq \delta N \quad \Rightarrow \quad expansion(S) \geq \frac{1}{2}$$

Question:
If G is a δ-small set expander,
How many eigenvalues larger than $1-\epsilon$ can the normalized adjancency matrix A have?

Intuitively,
How many sparse cuts can a graph G have without having a unbalanced sparse cut?

Cheeger's inequality:
every large eigenvalue → a sparse cut in G
$(\lambda_i \geq 1-\epsilon)$ (cut of sparsity $O(\sqrt{\epsilon})$)

Significance of the Question

Def (Small Set Expander)**:**
A regular graph G is a δ-small set expander if for every set $S \subset V$

$|S| \leq \delta N \quad \Rightarrow \quad expansion(S) \geq \frac{1}{2}$

Question:
If G is a δ-small set expander,
How many eigenvalues larger than $1 - \epsilon$ can the normalized adjacency matrix A have?

$$threshold\ rank_{1-\epsilon}(G) \stackrel{\text{def}}{=} \#\text{ of eigenvalues of graph G that are } \geq 1 - \epsilon$$

[Barak-Raghavendra-Steurer][Guruswami-Sinop]

k-round Lasserre SDP solves $UG(\epsilon)$ on graphs with low threshold rank $\leq k$.

Graph G has small non-expanding sets,

→

decompose G in to smaller pieces and solve each piece.

If UGC is true, there must be hard instances of Unique Games that

a) Have high threshold rank b) Are Small-Set Expanders.

Significance of the Question

Def (Small Set Expander)**:**
A regular graph G is a δ-small set expander if for every set $S \subset V$

$|S| \leq \delta N \quad \Rightarrow \quad expansion(S) \geq \frac{1}{2}$

Question:
If G is a δ-small set expander,
How many eigenvalues larger than $1-\epsilon$ can the normalized adjacency matrix A have?

$R(\epsilon,\delta) \stackrel{\text{def}}{=}$ Answer to the above question (a function of N, ϵ, δ)

[Arora-Barak-Steurer 2010]

There is a $N^{R(\epsilon,\delta)}$ -time algorithm for GAP-SMALL-SET-EXPANSION problem – a problem closely related to UNIQUE GAMES.

Is $\Phi(\delta) < \epsilon$ OR $\Phi(\delta) > 1-\epsilon$?
where $\Phi(\delta) =$ minimum expansion of sets of size $\leq \delta$

At the time,
Best known lower bound for $R(\epsilon,\delta) = \log N \rightarrow$
(GAP-SMALL-SET-EXPANSION problem could be solved in quasipolynomial time.)

[Arora-Barak-Steurer 2010] $R(\epsilon,\delta) \leq N^{\epsilon}$
A subexponential-time algorithm for GAP-SMALL-SET-EXPANSION problem

Short Code Graph

[Barak-Gopalan-Hastad-Meka-Raghavendra-Steurer]

For all small constant δ,

There exists a graph (the Short Code Graph) that is a δ −small set expander with $\exp(\log^{\beta} n)$ eigenvalues $\geq 1 - \epsilon., i.e.,$

$$threshold\ rank_{1-\epsilon}(G) \geq \exp(\log^{\beta} N)$$

for some β depending on ϵ.

[BGHMRS 11] $\exp(\log^{\beta} N) \leq R(\epsilon, \delta) \leq N^{\epsilon}$ [ABS]

- Led to new gadgets for hardness reductions (derandomized Majority is Stablest) and new SDP integrality gaps.

- An interesting mix of techniques from, Discrete Fourier analysis, Locally Testable Codes and Derandomization.

Overview

Long Code Graph

- Eigenvectors
- Small Set Expansion

Short Code Graph

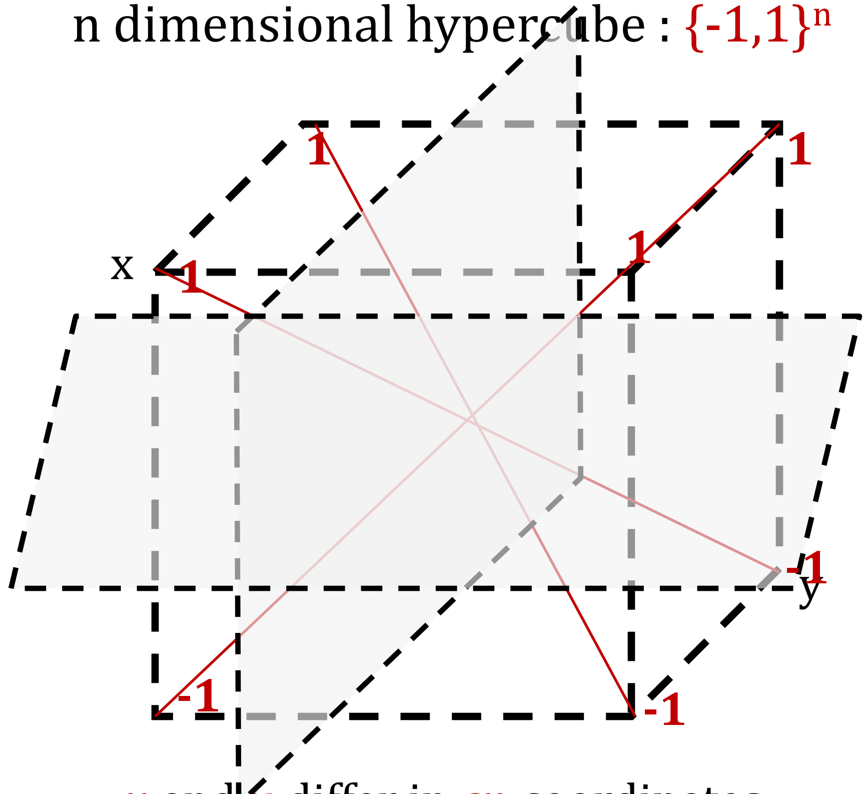
The Long Code Graph aka Noisy Hypercube

Noise Graph: H_ϵ

Vertices: $\{-1,1\}^n$

Edges: Connect every pair of points in hypercube separated by a Hamming distance of ϵn

n dimensional hypercube : $\{-1,1\}^n$



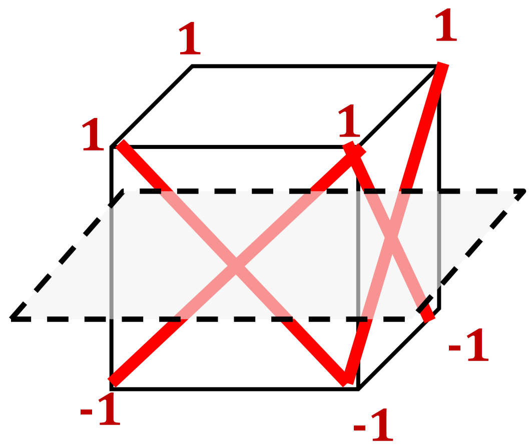
x and y differ in ϵn coordinates.

Eigenvectors are functions on $\{-1,1\}^n$

Dictator cuts: Cuts parallel to the Axis
(given by $F(x) = x_i$)

The dictator cuts yield n sparse cuts in graph H_ϵ

Sparsity of Dictator Cuts



n dimensional hypercube

Connect every pair of vertices in hypercube separated by Hamming distance of ϵn

$$\begin{aligned} &\mathit{Fraction\ of\ edges\ cut\ by\ first\ dictator} \\ &= \Pr_{random\ edge\ (x,y)}[(x,y)\ is\ cut] \\ &= \Pr_{random\ edge\ (x,y)}[x_1 \neq y_1] \\ &= \epsilon \end{aligned}$$

Dictator cuts: n-eigenvectors with eigenvalues $1-\epsilon$ for graph H_ϵ
(Number of vertices N = 2^n, so #of eigenvalues = $\log N$)

Eigenfunctions for Noisy Hypercube Graph

Eigenfunctions for the Noisy hypercube graph are multilinear polynomials of fixed degree!

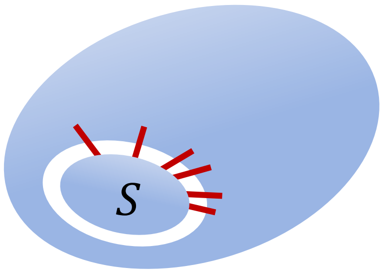
(Noisy hypercube is a Cayley graph on $Z_2{}^n$ *, therefore its eigen functions are characters of the group)*

Eigenfunction	**Eigenvalue**
$F_1(x) = x_1, F_2(x) = x_2, \quad \ldots , F_n(x) = x_n$	$1 - \epsilon$
$F_{12}(x) = x_1 x_2, F_{23}(x) = x_2 x_3, \quad \ldots F_{n-1n}(x) = x_{n-1}\, x_n$	$(1-\epsilon)^2$
……………………….	
Degree d multilinear polynomials	$(1-\epsilon)^d$

Small-Set Expansion (SSE)

Given: regular graph G with vertex set V, parameter $\delta > 0$

Suppose $f = \mathbb{1}_S$ indicator function of a small non-expanding set.



S has δ-fraction of vertices ➔ $\|f\|^2{}_2 = E[f^2] = \delta$

Fraction of edges inside S = $E_{\text{random edge }(x,y)} f(x)f(y)$

$= \langle f, Gf \rangle$

If $expansion(S) \leq 0.001$ then, at least a 0.999δ-fraction of edges are inside S.
So,

$$\langle f, Gf \rangle \geq 0.999 \|f\|_2{}^2$$

(f is close to the span of eigenvectors of G with eigenvalue ≥ 0.99)

Conclusion:
Indicator function of a small non-expanding set f = $\mathbb{1}_S$ is a
- sparse vector
- close to the span of the large eigenvectors of G

Hypercontractivity

Definition: (Hypercontractivity)

A subspace $S \in R^N$ is *hypercontractive* if for all $w \in S$

$$\|w\|_4 \leq C\|w\|_2$$

Projector P_S in to the subspace S, also called hypercontractive.

(No-Sparse-Vectors)

Roughly, No sparse vectors in a hypercontractive subspace S because,

w is δ-sparse "⇔" $\|w\|_4/\|w\|_2 > 1/\delta^{1/4}$

Hypercontractivity implies Small-Set Expansion

$P_{1-\epsilon} =$ projector into span of eigenvectors of G with eigenvalue $\geq 1-\epsilon$

$P_{1-\epsilon}$ is hypercontractive

→

No sparse vector in span of top eigenvectors of G

→

No small non-expanding set in G. (G is a small set expander)

Hypercontractivity for Noisy Hypercube

Top eigenfunctions of noisy hypercube are low degree polynomials.

(Hypercontractivity of Low Degree Polynomials)
For a degree d multilinear polynomial f on $\{-1,1\}^n$,

$$\|f\|_4 \leq 9^d \|f\|_2$$

(Noisy Hypercube is a Small-Set Expander)
For constant ϵ, the noisy hypercube is a small-set expander.
Moreover, the noisy hypercube has $N = 2^n$ vertices and n eigenvalues larger than $1 - \epsilon$.

Short Code Graph

[Barak-Gopalan-Hastad-Meka-Raghavendra-Steurer 2009]

For all small constant δ,

There exists a graph (the Short Code Graph) that is a δ −small set expander with $\exp(\log^{\beta} n)$ eigenvalues $\geq 1 - \epsilon., i.e.,$

$$threshold\ rank_{1-\epsilon}(G) \geq \exp(\log^{\beta} N)$$

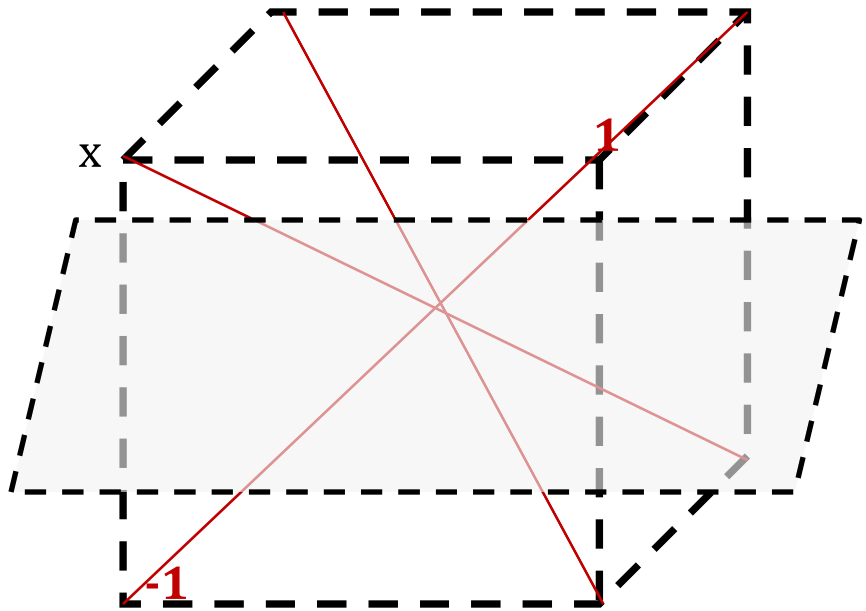
for some β depending on ϵ.

Short Code Graph

Noise Graph: H_ϵ

Vertices: $\{-1,1\}^n$

Edges: Connect every pair of points in hypercube separated by a Hamming distance of ϵn



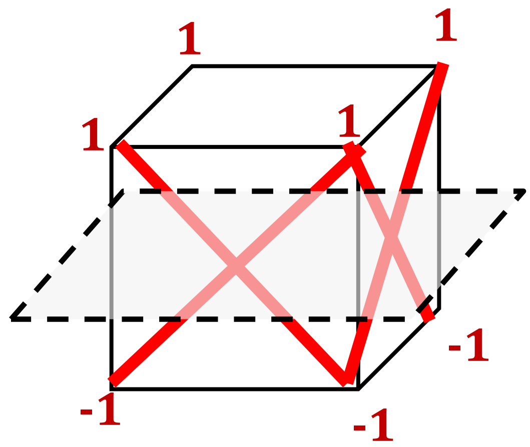
Has n sparse cuts, but $N = 2^n$ vertices -- too many vertices!

Idea:
Pick a subset of vertices of the long code graph, and their induced subgraph.

1. The dictator cuts still yield n-sparse cuts
2. The subgraph is a small-set expander!

Choice: Reed Muller Codewords of large constant degree.

If



n dimensional hypercube

Sparsity of Dictator Cuts

Connect every pair of vertices in hypercube separated by Hamming distance of ϵn

$$\begin{aligned} &Fraction\ of\ edges\ cut\ by\ first\ dictator \\ &= \Pr_{random\ edge\ (x,y)}[(x,y)\ is\ cut] \\ &= \Pr_{random\ edge\ (x,y)}[x_1 \neq y_1] \\ &= \epsilon \end{aligned}$$

Easy: Pretty much for any reasonable subset of vertices, dictators will be sparse cuts.

Preserving Small Set Expansion

Top eigenfunctions of noisy hypercube are low degree polynomials.

+

(Hypercontractivity of Low Degree Polynomials)
For a degree d multilinear polynomial f on $\{-1,1\}^n$,

$$\|f\|_4 \leq 9^d \|f\|_2$$

⬇

(Noisy Hypercube is a Small-Set Expander)
For constant ϵ, the noisy hypercube is a small-set expander.

Preserving Small Set Expansion

(Hypercontractivity of Low Degree Polynomials)
For a degree d multilinear polynomial f on $\{-1,1\}^n$,

$$\|f\|_4 \leq 9^d \|f\|_2$$

For a degree d polynomial f,

By hypercontractivity over hypercube,

$$E_{x\in\{-1,1\}^n}[f(x)^4] \leq 9^{4d}\left(E_{x\in\{-1,1\}^n}[f(x)]\right)^2$$

We picked a subset $S \subset \{-1,1\}^n$ and so we want,

$$E_{x\in S}[f(x)^4] \leq 9^{4d}(E_{x\in S}[f(x)])^2$$

f is degree d, so f^4 and f^2 are degree $\leq 4d$.

If S is a 4d-wise independent set then,

$$\begin{aligned} E_{x\in S}[f(x)^4] &= E_{x\in\{-1,1\}^n}[f(x)^4] \\ &\leq 9^{4d}\left(E_{x\in\{-1,1\}^n}[f(x)]\right)^2 = 9^{4d}(E_{x\in S}[f(x)])^2 \end{aligned}$$

Preserving Small Set Expansion

Top eigenfunctions of noisy hypercube are low degree polynomials.

We Want:

Only top eigenfunctions on the subgraph of noisy hypercube are also low degree polynomials.

Connected to **local-testability** of the dual of the underlying code S!

We appeal to local testability result of Reed-Muller codes

[Bhattacharya-Kopparty-Schoenebeck-Sudan-Zuckermann]

Applications of Short Code

[Barak-Gopalan-Hastad-Meka-Raghavendra-Steurer 2009]

`Majority is Stablest' theorem holds for the short code.

-- More efficient gadgets for hardness reductions.

-- Stronger integrality gaps for SDP relaxations.

[Kane-Meka]

A $2^{(\log\log n)^{1/3}}$ SDP gap with triangle inequalities for BALANCED SEPARATOR

Recap:

``How many large eigenvalues can a small set expander have?''

-- A graph construction led to better hardness gadgets and SDP integrality gaps.

On the Power of
Sum-of-Squares Proof

SoS hierarchy is a natural candidate algorithm for refuting UGC

Should try to prove that this algorithm fails on *some* instances

Only candidate instances were based on long-code or short-code graph

Result:

Level-8 SoS relaxation refutes UG instances
based on *long-code* and *short-code* graphs

We don't know any instances on which this algorithm could potentially fail!

Result:

Level-8 SoS relaxation refutes UG instances
based on *long-code* and *short-code* graphs

How to prove it? (rounding algorithm?)

Interpret dual as proof system

Show in this proof system that no assignments for these instances exist

We already know "regular" proof of this fact! (soundness proof)

Try to lift this proof to the proof system

qualitative difference to other hierarchies: basis independence

Sum-of-Squares Proof System (informal)

Axioms

$P_1(z) \geq 0$

$\vdots$

$P_m(z) \geq 0$

derive $\Rightarrow$ $Q(z) \leq c$

($P_1, \ldots, P_m, Q$ bounded-degree polynomials)

Rules

$\left.\begin{array}{l}\text{Polynomial operations} \\ R(z)^2 \geq 0 \text{ for any polynomial } R\end{array}\right\}$ "Positivstellensatz" [Stengel'74]

Intermediate polynomials have *bounded degree*

(c.f. bounded-width resolution, but basis independent)

Example

Axiom: $z^2 \leq z$ Derive: $z \leq 1$

$$1 - z = z - z^2 + (1 - z)^2$$

$$\geq z - z^2 \qquad \text{(non-negativity of squares)}$$

$$\geq 0 \qquad \text{(axiom)}$$

Components of soundness proof (for known UG instances)

Non-serious issues:

Cauchy–Schwarz / Hölder

Influence decoding

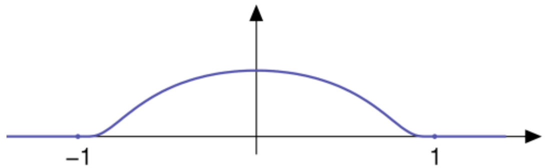
Serious issues:

Hypercontractivity

can use variant of inductive proof, works in *Fourier basis*

Invariance Principle

typically uses *bump functions*, but for UG, polynomials suffice



G $=$ long-code graph $\mathrm{Cay}(\mathbb{F}_2^m, T)$ where $T = \{$points with Hamming weight $\varepsilon m\}$

$P =$ projector into span of eigenfunctions of G with eigenvalue $\geq \lambda = 0.1$

SoS proof of hypercontractivity:

$2^{O(1/\varepsilon)}\|f\|_2^4 - \|Pf\|_4^4$ is a sum of squares

G $=$ long-code graph $\mathrm{Cay}(\mathbb{F}_2^m, T)$ where T $= \{$points with Hamming weight $\varepsilon m\}$

P $=$ projector into span of eigenfunctions of G with eigenvalue $\geq \lambda = 0.1$

SoS proof of hypercontractivity:

difference is sum of squares

$$\|Pf\|_4^4 \preccurlyeq 2^{O(1/\varepsilon)} \|f\|_2^4$$

For long-code graph, P projects into *Fourier polynomials* with degree $O(1/\varepsilon)$

Stronger ind. Hyp.:

$\mathbf{E}\, f^2 g^2 \preccurlyeq 3^{d+e} \mathbf{E} f^2 \cdot \mathbf{E} g^2$ where f is a generic degree-d Fourier polynomial and g is a generic degree-e Fourier polynomial

$$\mathbf{E} f^2 = \sum_{S, |S| \leq d} \hat{f}_S^2$$

G = long-code graph $\mathrm{Cay}(\mathbb{F}_2^m, T)$ where $T = \{\text{points with Hamming weight } \varepsilon m\}$

$P =$ projector into span of eigenfunctions of G with eigenvalue $\geq \lambda = 0.1$

SoS proof of hypercontractivity:

$$\boxed{\|Pf\|_4^4 \preccurlyeq 2^{O(1/\varepsilon)} \|f\|_2^4}$$

For long-code graph, P projects into *Fourier polynomials* with degree $O(1/\varepsilon)$

Stronger ind. Hyp.:

$\mathbf{E}\, f^2 g^2 \preccurlyeq 3^{d+e} \mathbf{E} f^2 \cdot \mathbf{E} g^2$ where f is a generic degree-d Fourier polynomial and g is a generic degree-e Fourier polynomial

Write $f = f_0 + x_1 \cdot f_1$ and $g = g_0 + x_1 \cdot g_1$ (degrees of f_1, g_1 smaller than d, e)

$$\begin{aligned}
\mathbf{E}\, f^2 g^2 &= \mathbf{E}\, f_0^2 g_0^2 + \mathbf{E}\, f_1^2 g_0^2 + \mathbf{E}\, f_0^2 g_1^2 + \mathbf{E}\, f_1^2 g_1^2 + 4\mathbf{E}\, f_0 f_1 g_0 g_1 \\
&\preccurlyeq \quad \ldots \quad + 2\mathbf{E}\, f_0^2 g_1^2 + 2\mathbf{E}\, f_1^2 g_0^2 \\
&\preccurlyeq 3^{d+e} (\mathbf{E} f_0^2 + \mathbf{E} f_1^2) \cdot (\mathbf{E} g_0^2 + \mathbf{E} g_1^2) \quad \text{(ind. hyp.)} \quad \blacksquare
\end{aligned}$$

Open Questions

Does 8 rounds of Lasserre hierarchy disprove UGC?

Can we make the short code, any shorter?
(applications to hardness gadgets)

Subexponential time algorithms for MaxCut or Vertex Cover
(beating the current ratios)

Thanks!