Semidefinite Programming Hierarchies and the Unique Games Conjecture

David Steurer

Microsoft Research New England

Cornell University

Algorithms Frontiers Workshop, EPFL, Lausanne, June 2012

Plan

Overview

aka Lasserre hierarchy

Sum-of-Squares (SoS) SDP Hierarchy

[Parrilo'00, Lasserre'01]

Rounding SDP Hierarchies via Global Correlation

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

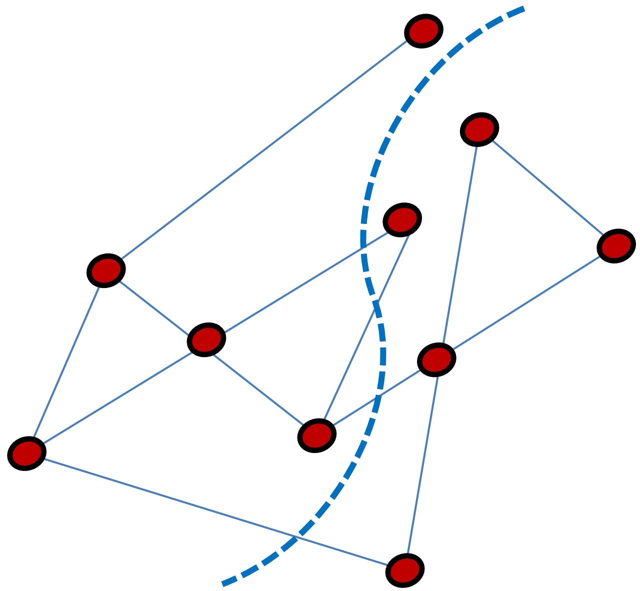
Power of Sum-of-Squares Proofs

[Barak-Brandão-Harrow-Kelner-S.-Zhou'12]

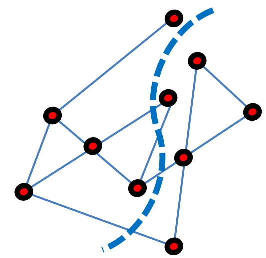
Max Cut

Given: undirected graph on n vertices

Find: bipartition that cuts as many edges as possible



Max Cut



Given: undirected graph on n vertices

Find: bipartition that cuts as many edges as possible

polynomial optimization problem

$$\max_{x\in\{\pm 1\}^n} \sum_{i\sim j} \tfrac{1}{4}\left(x_i - x_j\right)^2$$

simple space

can understand set of polynomials (ideal) vanishing on this set

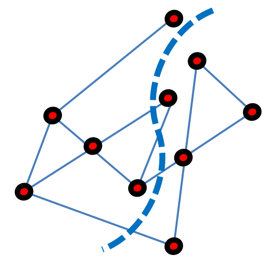
quadratic polynomial

sum of local terms

→ constraint satisfaction problem

(constraints of form $x_i \neq x_j$)

Max Cut

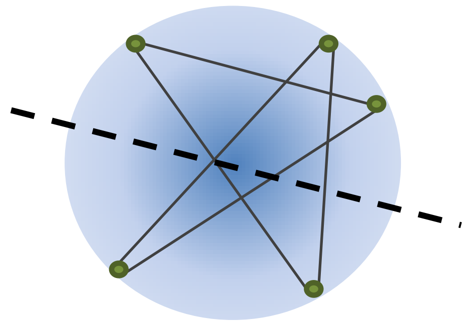


Given: undirected graph on n vertices

Find: bipartition that cuts as many edges as possible

best known approximation ratio: $\alpha_{\mathrm{GW}} \approx 0.878\ldots$ [Goemans-Williamson]

What does it take to beat this bound?

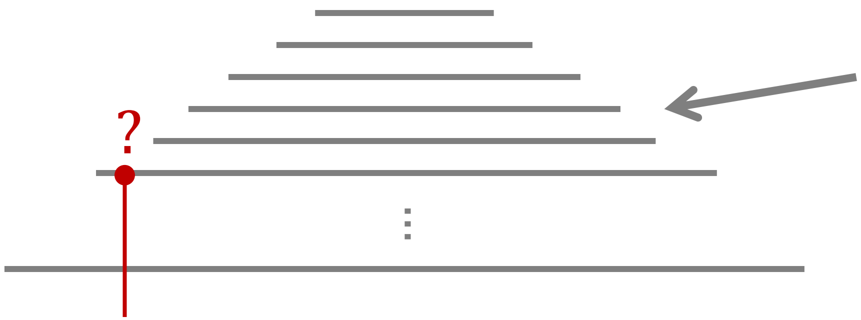


Semidefinite Programming (SDP) Hierarchies

[Sherali-Adams'90, Lovász-Schrijver'91,… Parrilo'00, Lasserre'01]

general approach for any combinatorial optimization problem

sequence of increasingly stronger SDP relaxations



level-k relaxation, time $n^{O(k)}$

(k is even)

MAX CUT$(\alpha_{\mathrm{GW}} + \varepsilon)$

Appeal:

systematic
+
powerful

no creativity required to come up with relaxations!
(*contrast*: ad-hoc relaxations)

plausibly optimal approximation-time tradeoff for large class of problems, including MAX CUT

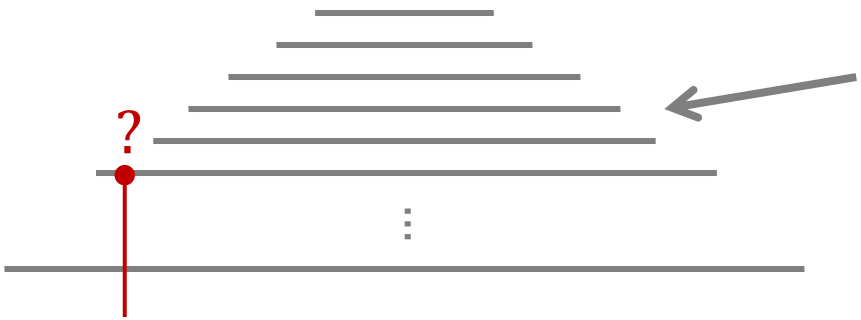
Grand Unified Theory for optimization?

Semidefinite Programming (SDP) Hierarchies

[Sherali-Adams'90, Lovász-Schrijver'91,... Parrilo'00, Lasserre'01]

general approach for any combinatorial optimization problem

sequence of increasingly stronger SDP relaxations



level-k relaxation, time $n^{O(k)}$

(k is even)

MAX CUT($\alpha_{\text{GW}} + \varepsilon$)

level-2 ratio = α_{GW}
level-4 ratio = ??

Appeal: ***systematic + powerful***

Grand Unified Theory for optimization?

So far: for many problems, e.g., MAX CUT, understanding is poor

But: there is progress

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 $k = \log n$)

Distinguish:

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

NO: at most ε of equations satisfiable

Unique Games Conjecture (UGC)

[Khot'02]

Implications of UGC

= level-2 SDP relaxation

For large class of problems, BASIC SDP achieves optimal approximation

Examples: MAX CUT, VERTEX COVER, any MAX CSP

[Khot-Regev'03, Khot-Kindler-Mossel-O'Donnell'04, Mossel-O'Donnell-Oleszkiewicz'05, Raghavendra'08]

Is the conjecture true?

Unique Games Conjecture (UGC) [Khot'02]

Implications of UGC

Is the conjecture true?

Difference to other complexity conjectures

difficulty: seems only barely out of reach (good bang-for-buck!)

plausibility: relatively weak evidence (might very well be false!)

Framework: general & simple approach for analyzing SDP hierarchies

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

gives subexponential algorithm for UNIQUE GAMES [Arora-Barak-S.'10]

contrast: many NP-hard approximation problems require exponential time (assuming 3-SAT does)

some other application (later in talk)

Limitations: this approach cannot give much faster algorithms

construction of small-set expanders with many large eigenvalues

[Barak-Gopalan-Håstad-Meka-Raghavendra-S.'11]

leads to hard instances for weaker SDP hierarchies

New approach: these instances are not hard for (stronger) SDP hierarchies

can be solved in a constant number of levels

Plan

Overview

Sum-of-Squares (SoS) SDP Hierarchy

Rounding SDP Hierarchies via Global Correlation

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

Power of Sum-of-Squares Proofs

[Barak-Brandão-Harrow-Kelner-S.-Zhou'12]

Three equivalent formulations of Max Cut

assignments $x \in \{\pm 1\}^n$

$$\max_{x} \sum_{i \sim j} ¼\left(x_i - x_j\right)^2$$

not convex!

distributions μ over $\{\pm 1\}^n$

$$\max_{\mu} \mathbb{E}_{x \sim \mu} \sum_{i \sim j} ¼\left(x_i - x_j\right)^2$$

too large to represent!

low-degree moments M of distributions over $\{\pm 1\}^n$

$$\max_{M} \sum_{i \sim j} ¼\left(M_{ii} - 2M_{ij} - M_{jj}\right)^2$$

$M_{ij} = \mathbb{E}_{x \sim \mu} x_i x_j$

low-degree moments M of distributions over $\{\pm 1\}^n$

$$\max_M \sum_{i \sim j} ¼\left(M_{ii} - 2M_{ij} - M_{jj}\right)^2$$

$M\left(\sum_{i\sim j} ¼(x_i - x_j)^2\right)$

Multivariate moments

$x^\alpha = \prod x_i^{\alpha_i}$

$M_\alpha = \mathbb{E}_{x\sim\mu} x^\alpha$ is moment of monomial x^α Example: $M_{ijjk} = \mathbb{E}_{x\sim\mu} x_i x_j^2 x_k$

$n^{O(\ell)}$ monomial-moments of degree ℓ → easy to represent

For polynomial $p = \sum p_\alpha x^\alpha$, moment $M(p) = \mathbb{E}_{x\sim\mu} p(x) = \sum p_\alpha M_\alpha$

degree-ℓ moments form convex set

→ *try to describe it by linear equalities and inequalities*

Which linear equalities? all of them! e.g., $M_{ijjk} = M_{ik}$

Which linear inequalities? just one (class)! $M(p^2) \geq 0$ for all p

Level-ℓ Sum-of-Squares (SoS) relaxation (for Max Cut)

degree-ℓ *pseudo-moments* M of distributions over $\{\pm 1\}^n$

$$\max_M M\left(\sum_{i\sim j} \tfrac{1}{4}(x_i - x_j)^2\right)$$

degree-ℓ *pseudo-moments* $M = (M_\alpha; |\alpha| \le \ell)$

one variable M_α per degree-ℓ monomial x^α

Notation: $M(p) = \sum p_\alpha M_\alpha$ for degree-ℓ polynomial $p = \sum p_\alpha x^\alpha$

Constraints on pseudo-moments M

depends on $\{\pm 1\}^n$

all valid linear equalities — e.g., $M_{ijjk} = M_{ik}$, $M_{ii} = M_\emptyset = 1$

non-negativity of squares — $M(p^2) \ge 0$ for all degree-$\ell/2$ polynomials p

Separation Problem

independent of $\{\pm 1\}^n$

$\min_p M(p^2)$ is smallest eigenvalue of quadratic form $p \mapsto M(p^2)$

$\sum_{\alpha,\beta} M_{\alpha,\beta} p_\alpha p_\beta$

→ running time $n^{O(\ell)}$

Plan

Overview

Sum-of-Squares (SoS) SDP Hierarchy

Rounding SDP Hierarchies via Global Correlation

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

Power of Sum-of-Squares Proofs

[Barak-Brandão-Harrow-Kelner-S.-Zhou'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

[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 loss due to conditioning

Sampling ~~*Rounding*~~ *problem*

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

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

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

Given

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

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 → later in 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 $X_i' := \{X_i | X_j\}$

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

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*

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

typical pair of variables
almost pairwise independent

How can low global correlation help?

How can low global correlation help?

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

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?

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

For Unique Games

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

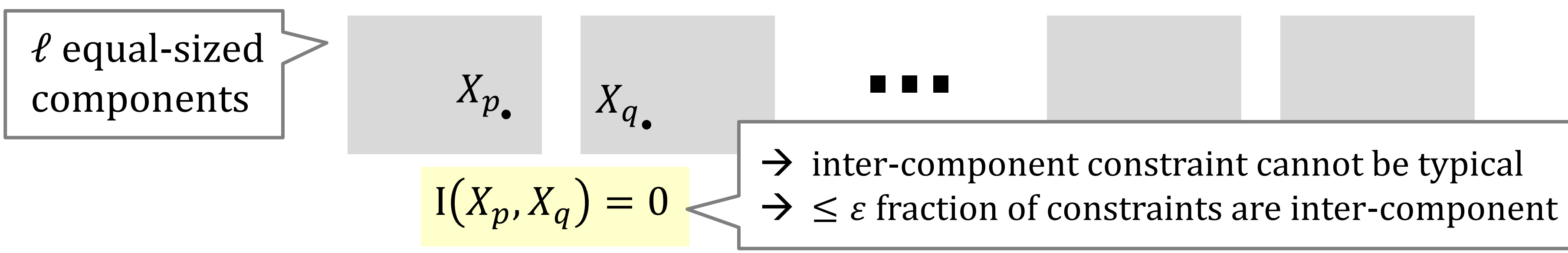
$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Extreme cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

all variables have uniform marginals & ∃ 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

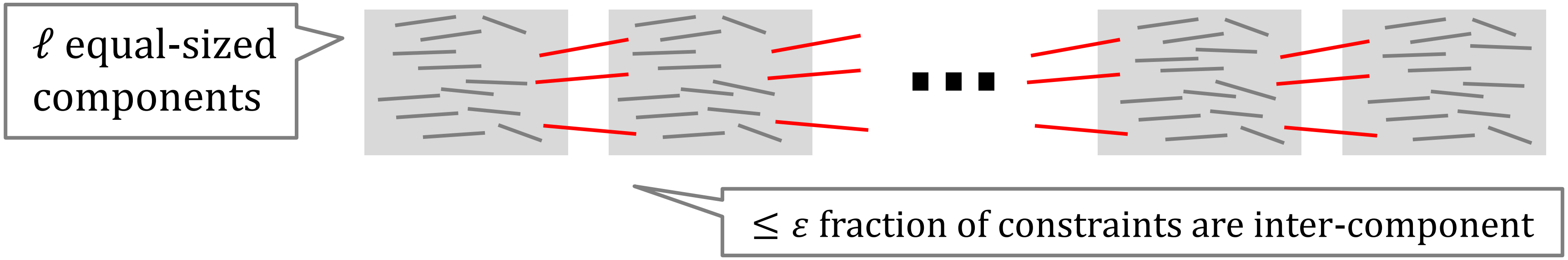
~~Extreme~~ cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

Show: no other cases are possible! (informal)

all variables have uniform marginals & ∃ 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 $X_1, \dots, X_n$ over $\mathbb{Z}_k$

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

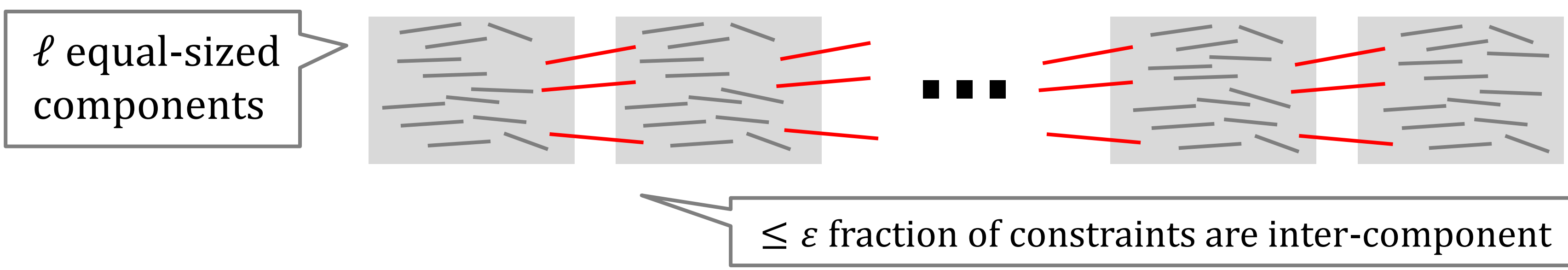
Only

~~Extreme~~ cases with low global correlation

1) no entropy: all variables are fixed → easy to "sample"

2) many small independent components: *?*

all variables have uniform marginals & ∃ partition:



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 $\geq n^\beta$

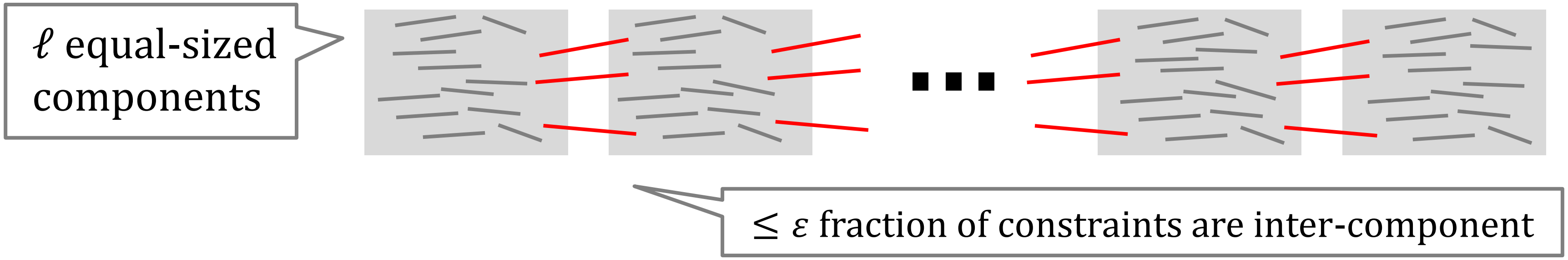
→ at most $1/\beta$ levels of recursion

→ total fraction of ignored edges $\leq \varepsilon/\beta \ll 1$

→ 2^{n^β}-time algorithm for UG(ε)

2) many small independent components: ***?***

all variables have uniform marginals & ∃ partition:



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

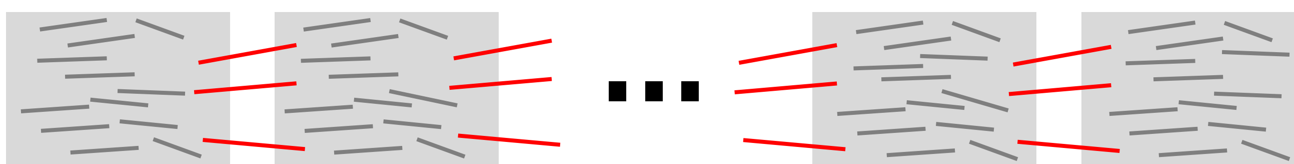
~~Extreme~~ cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

all variables have uniform marginals & $\exists$ partition:

ℓ equal-sized components



$\le \varepsilon$ fraction of constraints are inter-component

Proof: global correlation → mixing of random walks → small-set expansion

Plan

Overview

Sum-of-Squares (SoS) SDP Hierarchy

Rounding SDP Hierarchies via Global Correlation

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

Power of Sum-of-Squares Proofs

[Barak-Brandão-Harrow-Kelner-S.-Zhou'12]

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 ⇨ $Q(z) \leq c$

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

Rules

Polynomial operations

$R(z)^2 \geq 0$ for any polynomial R

} "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

use $\langle x, y \rangle \leq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2$
instead of $\langle x, y \rangle \leq \|x\|\|y\|$

Influence decoding

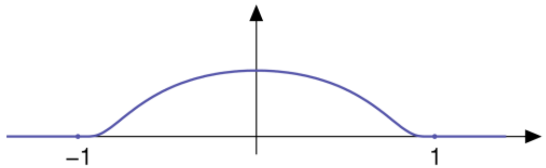
Serious issues:

Hypercontractivity

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

Invariance Principle

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



Open Questions

Unique Games Conjecture

Does level-8 of SoS hierarchy refute UGC?

Time vs Approximation Trade-offs

Better approximations for for MAX CUT, VERTEX COVER, …
in subexponential time?

Example:$1/\varepsilon$-approximation for SPARSEST CUT in time $\exp(n^{\varepsilon})$?

Thanks!