Summer school on semidefinite optimization

Approximation & Complexity

David Steurer

Cornell University

Part 1

September 6, 2012

Overview

Part 1	Unique Games Conjecture & Basic SDP
Part 2	SDP Hierarchies: Algorithms
Part 3	SDP Hierarchies: Limits

Constraint Satisfaction Problems

variables $x_1, \ldots, x_n$ over finite alphabet Σ

list of predicates/constraints

$P_1(x)$

.

.

.

$P_m(x)$

local: only depend on $O(1)$ variables

only predicates of certain *types*

Goal: satisfy as many predicates as possible

Constraint Satisfaction Problems

MAX 3SAT

variables $x_1, \ldots, x_n$ over finite alphabet Σ $= \{\text{true}, \text{false}\}$

list of predicates/constraints

$P_1(x)$ $= x_1 \vee x_2 \vee \overline{x_4}$

.

.

.

$P_m(x)$ $= \overline{x_9} \vee x_{42} \vee \overline{x_7}$

Goal: satisfy as many predicates as possible

Constraint Satisfaction Problems

MAX CUT

variables $x_1, \ldots, x_n$ over finite alphabet Σ $= \mathbb{F}_2$

list of predicates/constraints

$P_1(x)$ $= \{x_1 + x_2 = 1\}$

.

.

.

$P_m(x)$ $= \{x_{13} + x_5 = 1\}$

Goal: satisfy as many predicates as possible

Constraint Satisfaction Problems

UNIQUE GAMES(k)

variables $x_1, \dots, x_n$ over finite alphabet $\Sigma \;\; = \mathbb{F}_k$

list of predicates/constraints

$$P_1(x) \;\; = \{x_1 + x_2 = 4\}$$

.
.
.

$$P_m(x) \;\; = \{x_{13} + x_5 = 9\}$$

value of one variable *uniquely* determines value of other variable

Goal: satisfy as many predicates as possible

Optimization & Complexity

inherent difficulty, required computational resources

Goal: understand complexity of optimization problems

upper bounds

lower bounds

What are good algorithms?

What are hard instances?

Optimization & Complexity

Goal: understand complexity of optimization problems

require prohibitive resources (assuming P≠NP)

1970s Most discrete optimization problems are NP-hard [Cook, Karp, Levin] (including MAX 3SAT, MAX CUT, and UNIQUE GAMES)

So we can't hope to prove anything and have to resort to heuristics?

No!

Do not (blindly) trust impossibility results!

Optimization is not all or nothing!
What about approximate *solutions?*

(Many classical algorithms for convex optimization are fundamentally approximation algorithms)

Goal

understand trade-off between complexity and approximation

Approximation

Goal

understand trade-off between
complexity and approximation

How to measure approximation?

α-approximating

easy to state, but sometimes too coarse

$$\mathrm{ALG} \geq \alpha \cdot \mathrm{OPT}$$

(c,s)-approximating

finest measure

$$\text{if } \mathrm{OPT} \geq c, \text{ then } \mathrm{ALG} \geq s$$

Approximation

Goal

understand trade-off between
complexity and approximation

poly-time approximation algorithms:

non-trivial approximations for many problems,
e.g., 0.878-approx for MAX CUT [Goemans-Williamson]

NP-hardness of approximation

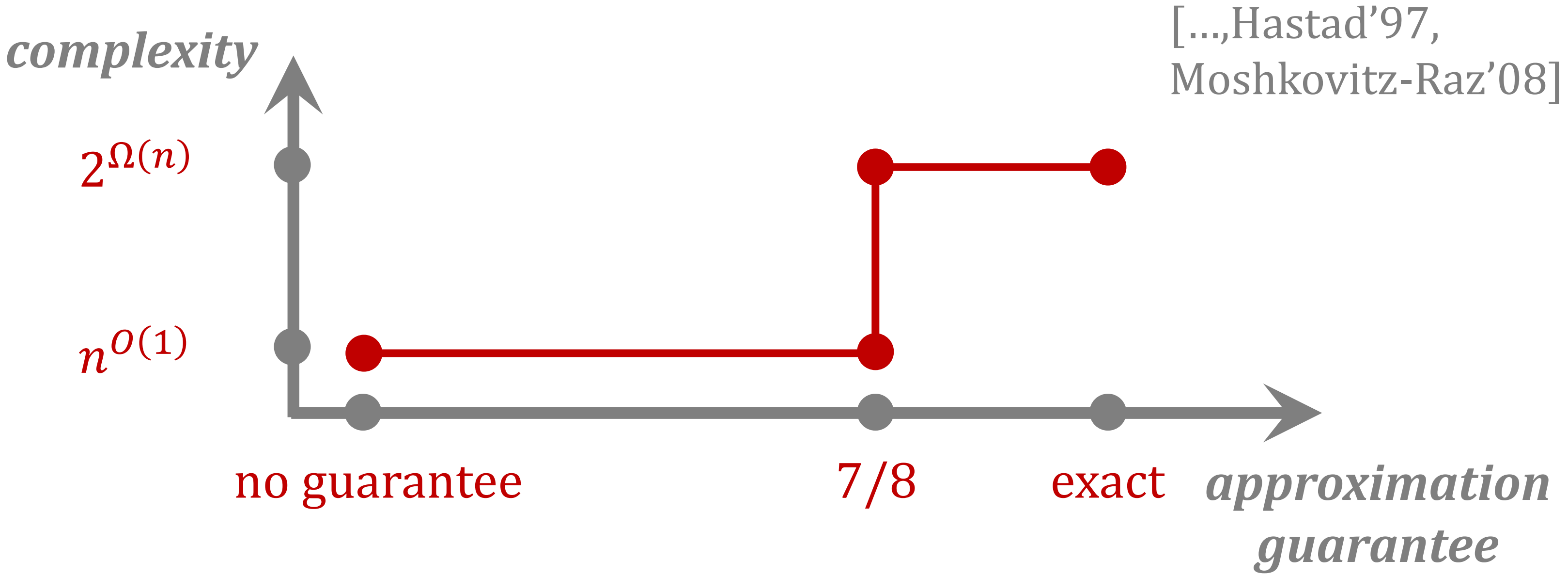
as hard as solving it exactly!

for many problems, some approximation is NP-hard
e.g., 0.999-approx for MAX CUT [PCP Theorem]

For very few problems, upper and lower bounds match!

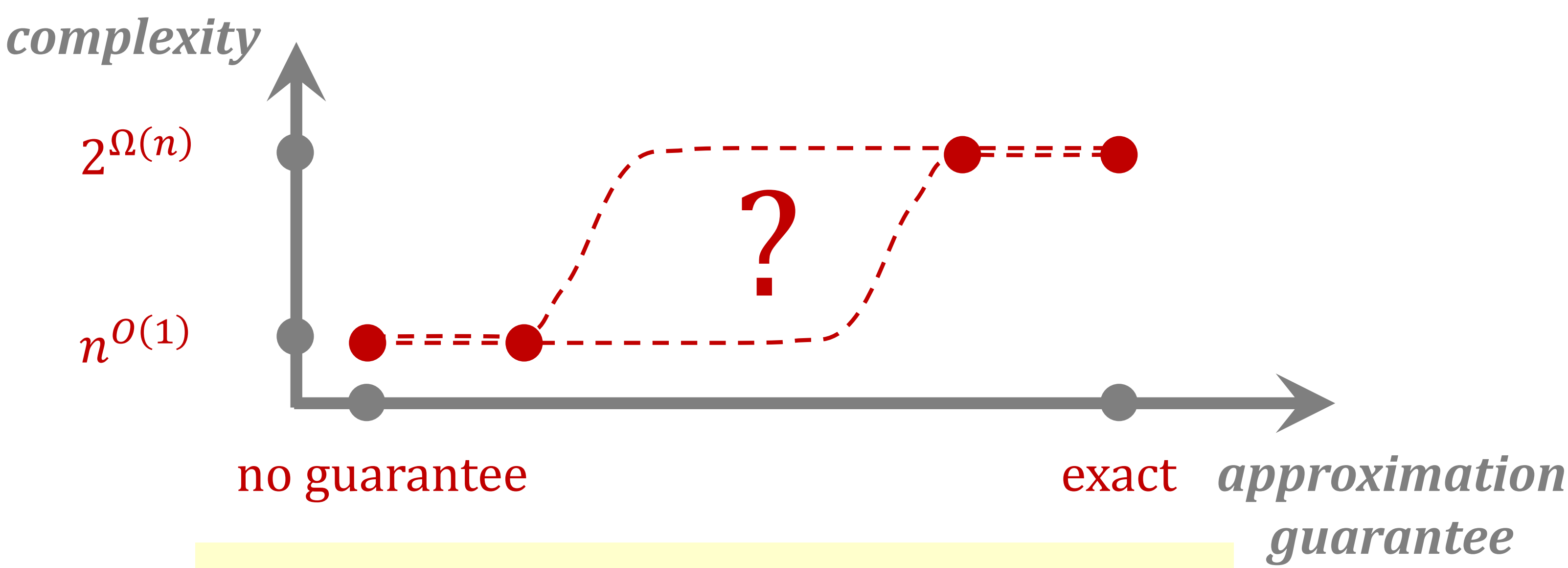
Complexity vs Approximation Trade-off

Max 3Sat



Complexity vs Approximation Trade-off

Most other problems



What algorithms are we missing?

What hard instances do we not know of?

Unique Games Conjecture (UGC)

[Khot'02]

For every $\varepsilon > 0$, there exists k,

constraints: $x_i - x_j = c \bmod \mathrm{k}$

$(1-\varepsilon, \varepsilon)$-approximation for UNIQUE GAMES(k) is NP-hard

Implications of UGC

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

For every CSP, the *Basic SDP relaxation* has optimal integrality gap
(→ higher-degree sum-of-squares relaxation have same gap)

Is the conjecture true?

Is the conjecture true?

subexponential-time algorithm

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

$(1-\varepsilon,\varepsilon)$-approximation for UG in time $\exp\left(n^{\varepsilon^{1/3}}\right)$

contrast: all known hardness results for CSPs imply $2^{\Omega(n)}$-hardness

part of framework for rounding SDP hierarchies

lower bounds for *certain* SDP hierarchies

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

subexp.-time essentially optimal within the rounding framework

hard instances based on new kind graphs (with extremal spectral properties)

sum-of-squares relaxations

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

"all known" instances of UG are solved in $O(1)$-degree sos relaxation

(including instances that are hard for other SDP hierarchies)

Generic Approximation Algorithm for CSPs [Raghavendra-S.'09]

For any CSP X,

OPT vs SDP

approximation for X = *integrality gap* of Basic SDP for X

ALG vs OPT

based on rounding optimal solutions to Basic SDP relaxation

new perspective on previous rounding algorithms, like GW

no explicit approximation guarantee

polynomial-time but huge constants (depending on desired accuracy)

Basic SDP Relaxation for Constraint Satisfaction Problems

variables $x_1, \ldots, x_n$ over finite alphabet Σ

list of predicates/constraints

$P_1(x) = x_1 \vee x_2 \vee \overline{x_4}$

.
.
.

$P_m(x) = \overline{x_9} \vee x_{42} \vee \overline{x_7}$

local distributions

first two moments are consistent and positive semidefinite

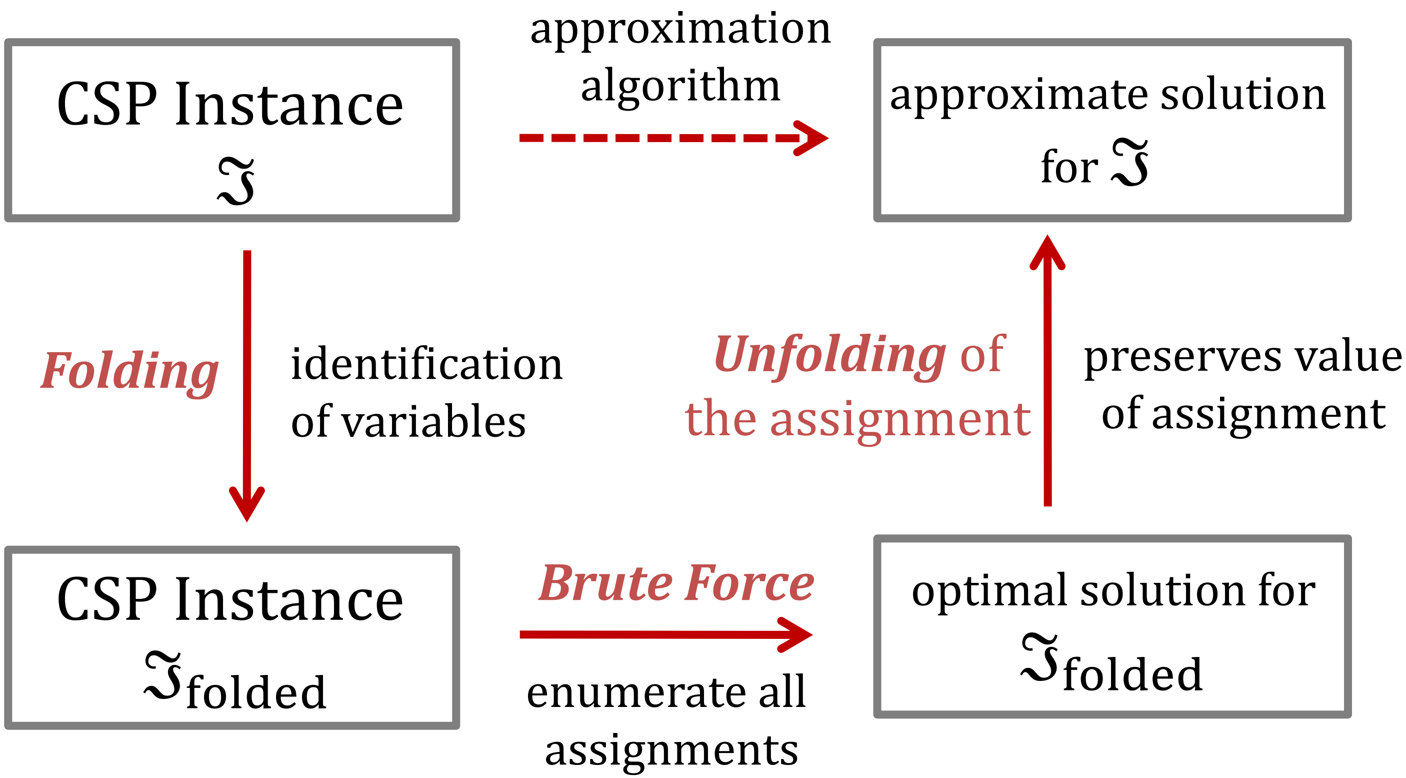
D_1

.
.
.

D_m

Goal: maximize expected number of satisfied predicates

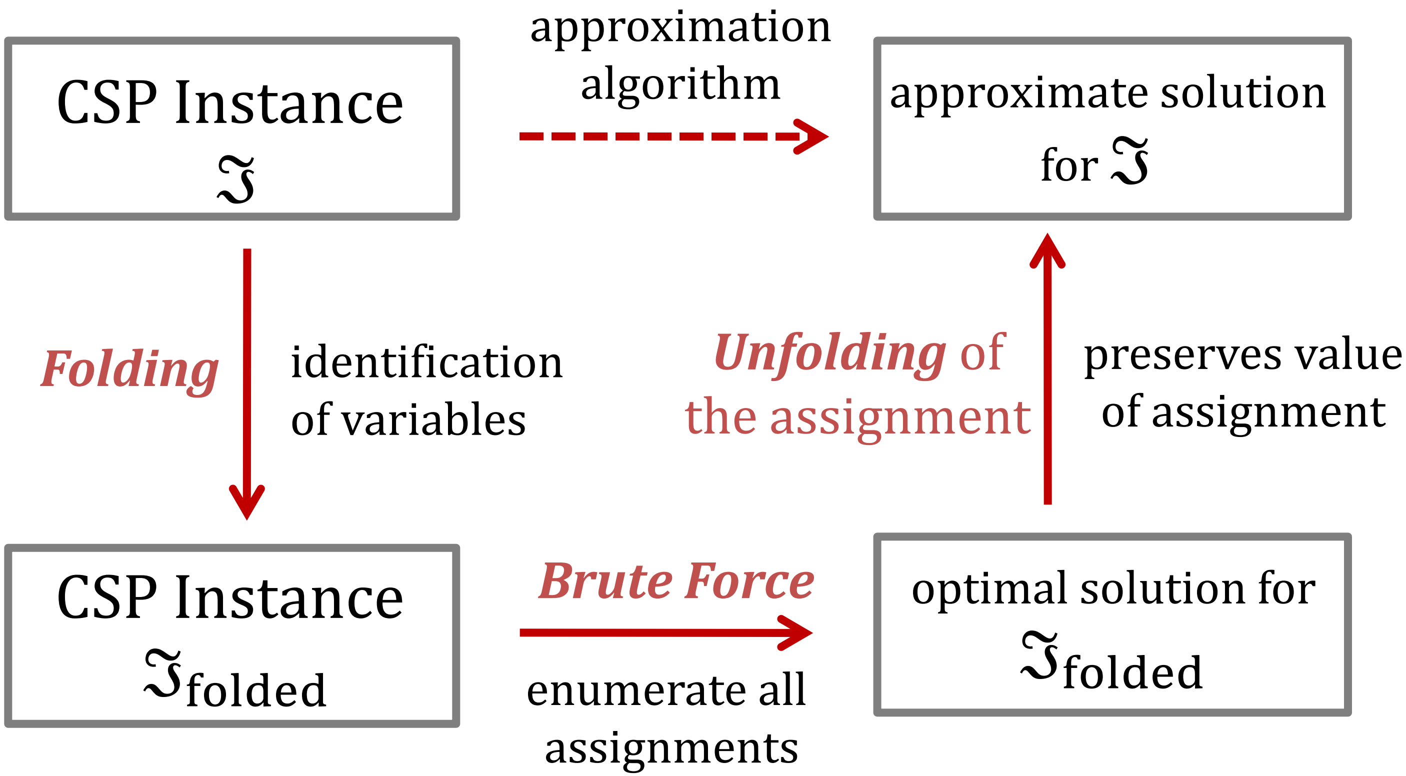
Approximating CSPs using Folding



"*Efficient*" whenever folding leaves only O(1) distinct variables

Challenge: ensure $\mathfrak{I}_{\text{folded}}$ has a good solution

Approximating CSPs using Folding

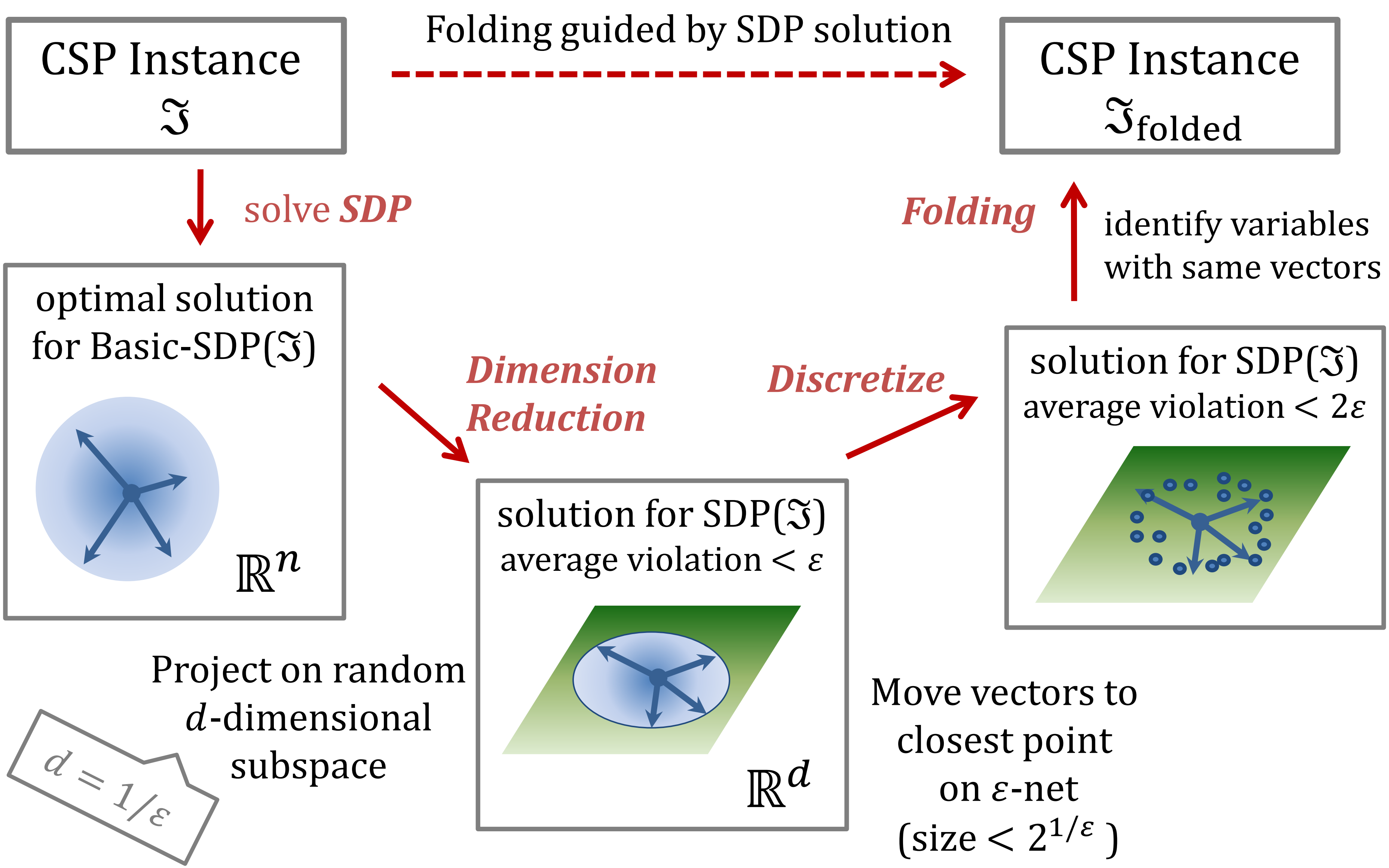


Theorem can fold *every* CSP instance efficiently to $2^{\text{poly}(1/\varepsilon)}$ variables

$$\text{sdp}(\mathfrak{I}_{\text{folded}}) \geq \text{sdp}(\mathfrak{I}) - \varepsilon$$

→ optimal rounding scheme

How to fold using SDP solutions



How to fold using SDP solutions

CSP Instance $\mathfrak{I}$ → Folding guided by SDP solution → CSP Instance $\mathfrak{I}_{\text{folded}}$

found solution for $\text{SDP}(\mathfrak{I}_{\text{folded}})$ with value $\geq \text{sdp}(\mathfrak{I}) - 2\varepsilon$

But: some constraints violated, on average by $\leq 2\varepsilon$

Robustness property of Basic SDP relaxation

can repair violations at proportional cost for objective value

→ $\text{sdp}(\mathfrak{I}_{folded}) \geq \text{sdp}(\mathfrak{I}) - 4\varepsilon$ ■

Summer school on semidefinite optimization

Approximation & Complexity

David Steurer

Cornell University

Part 2

September 7, 2012

Overview

Part 1	Unique Games Conjecture & Basic SDP
Part 2	SDP Hierarchies: Algorithms
Part 3	SDP Hierarchies: Limits

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

$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ 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)

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*

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

typical pair of variables almost 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 $X_i' \coloneqq \{X_i | X_j\}$

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

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*

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

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?

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

For Unique Games

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

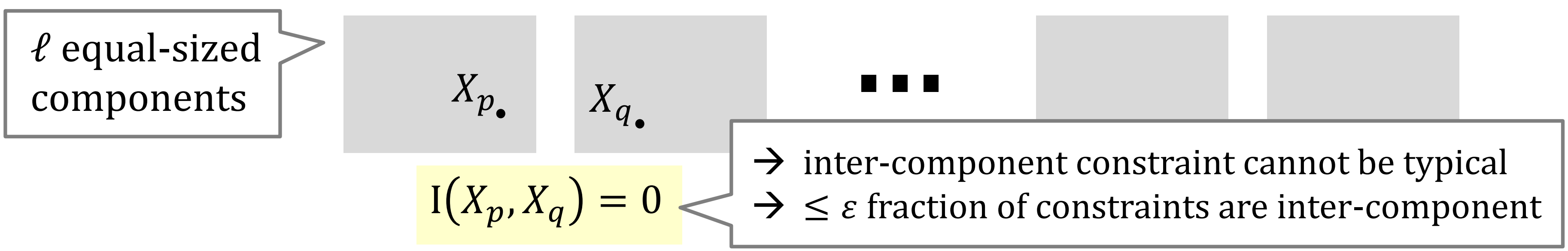
$\Pr(X_i - X_j = c) \geq 1 - \varepsilon$ for typical constraint $x_i - x_j = c$

Extreme cases with low global correlation

1) no entropy: all variables are fixed

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

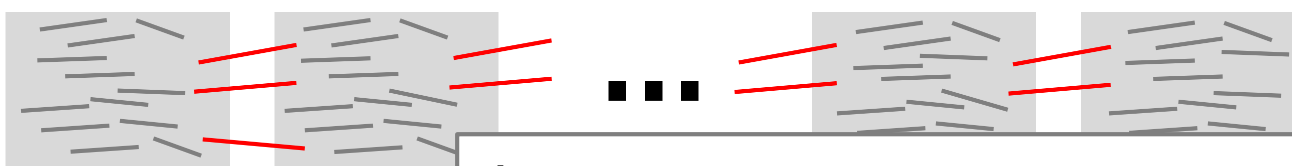
1) no entropy: all variables are fixed

2) many small independent components:

Show: no other cases are possible! (informal)

all variables have uniform marginal distribution & ∃ partition:

ℓ equal-sized components



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

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

Idea: round components independently & recurse on them

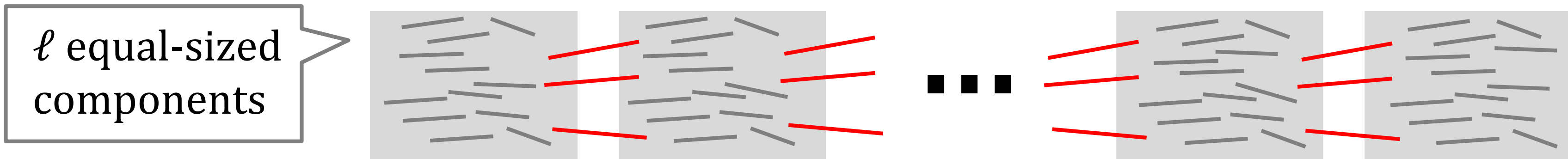
How many edges ignored in total? (between different components)

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^β}-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

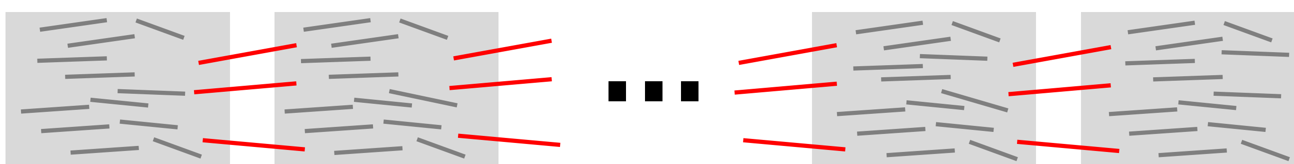
~~Extreme~~ cases with low global correlation

1) no entropy: all variables are fixed

2) many small independent components:

all variables have uniform marginal distribution & ∃ partition:

ℓ equal-sized components



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

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

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

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

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

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

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 $X_1, \dots, X_n$ 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 $\leq$ ~~$1/n^{2\beta}$~~ $1/\ell$

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

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

(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}

Thanks!

Summer school on semidefinite optimization

Approximation & Complexity

David Steurer

Cornell University

Part 3

September 7, 2012

Overview

Part 1	Unique Games Conjecture & Basic SDP
Part 2	SDP Hierarchies: Algorithms
Part 3	SDP Hierarchies: Limits

Approximation limits of s.o.s. methods

random assignment has value ½ in expectation

predicates $x_i \oplus x_j \oplus \bar{x_j}$

[Grigoriev'01, Schoenebeck'08]

For a *random instance* $\mathfrak{I}$ of MAX 3XOR, with high probability

(1) value of $\mathfrak{I}$ is at most $1/2 + 0.01$

(2) value of degree-$0.01n$ s.o.s. relaxation for $\mathfrak{I}$ is at least 0.99

predicates ≫ # variables

Corresponding NP-hardness result is known!

Why is this result interesting?

independent of P vs NP question

suggests random instances are hard

evidence that NP-hard problem take exp. time

Approximation limits of s.o.s. methods

In terms of polynomials:

edges ≫ # vertices

random 3-uniform hypergraph H, random sign vector $\sigma \in \{\pm 1\}^H$

degree-3 polynomial $P = \sum_{e \in H} \sigma_e \cdot X^e$

Then, w.h.p.,

Chernoff bound over σ

(1) $P \leq 0.01$ over $\{\pm 1\}^n$

(2) all s.o.s. certificate for $P \leq 0.99$ over $\{\pm 1\}^n$ have degree $\Omega(n)$

(2') no degree-$o(n)$ s.o.s. refutation of the system

$$\{\, \sigma_e \cdot X^e = 1 \mid e \in H \,\} \cup \left\{ X_i^2 = 1 \mid i \in V \right\}$$

Interlude: ***Bounded-width Gaussian Elimination***

system of polynomials over $\{\pm 1\}^n$

system of affine linear forms over $\mathbb{F}_2^n$

$$X_1 X_2 X_3 = 1$$
$$\vdots$$
$$-X_2 X_6 X_8 = 1$$

$$\longleftrightarrow$$

$$x_1 + x_2 + x_3 = 0$$
$$\vdots$$
$$1 + x_2 + x_6 + x_8 = 0$$

width-d Gaussian refutation

derivation of $1 = 0$ by adding equations of *width* $\leq d$

variables in equation

Approximation limits of s.o.s. methods

Part 1

random 3-uniform signed hypergraph (H, σ)

→ corresponding system has *elimination width* $\Omega(n)$

Part 2

For systems we consider,

width-d Gaussian refutation

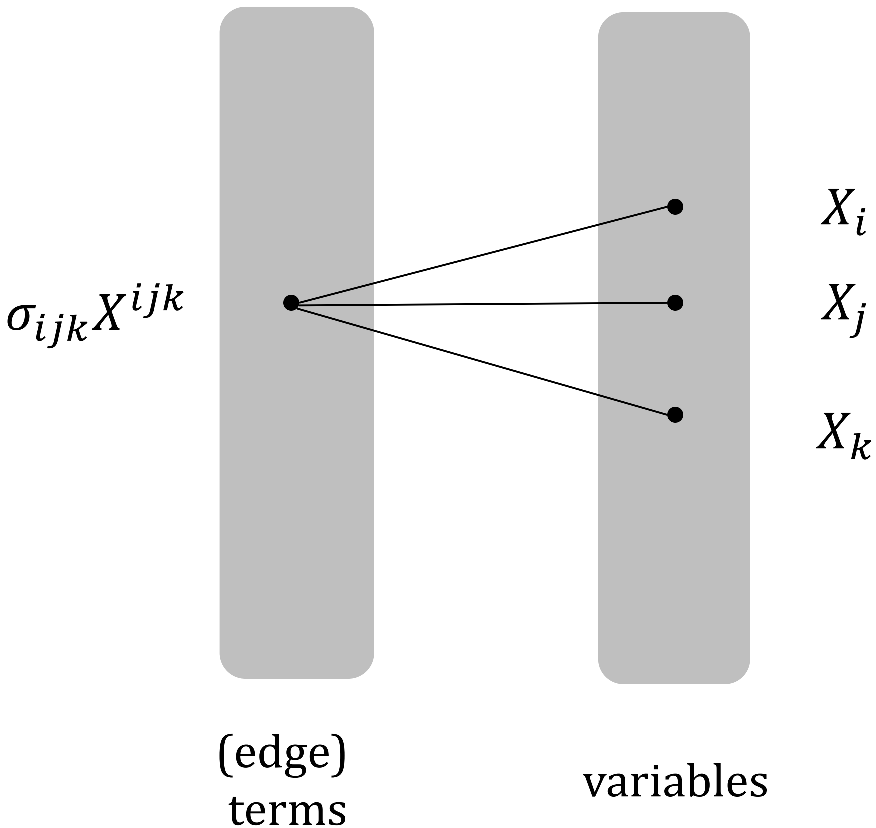
$\longleftrightarrow$ degree-d Nullstellensatz refutation

$\longleftrightarrow$ degree-d Positivstellensatz refutation

Want to show:

Random hypergraph system $\longrightarrow$ no width-$\Omega(n)$ Gaussian refutation

bipartite graph



vertex sets with $|S| < n/100$
→ $\Omega(|S|)$ unique neighbors

aX^{α} is product of edge terms S
→ aX^{α} has width $\geq \Gamma_{\text{unique}}(S)$

every refutation contains term aX^{α}
product of $\approx n/100$ edges terms

■

Want to show:

No width-$10d$ Gaussian refutation $\longrightarrow$ no degree-d Positivstellensatz refutation

How would degree-d s.o.s. refutation look like?

$\exists$ degree-d multipliers Q_e

$$1 + \mathrm{S.O.S.} = \sum_e Q_e \cdot (\sigma_e X^e - 1) \qquad \text{over } \{\pm 1\}^n$$

How to construct M?

→ *Gaussian elimination*

To rule out refutation:

exhibit linear form M on polynomials over $\{\pm 1\}^n$

$$M(1) = 1$$

$$M(\mathrm{S.O.S}) \geq 0 \qquad \forall\ \mathrm{S.O.S}$$

$$M\big(Q \cdot (\sigma_e X^e - 1)\big) = 0 \qquad \forall e, \text{ degree-}d\ \mathrm{Q}$$

Want to show:

No width-$10d$ Gaussian refutation $\longrightarrow$ no degree-d Positivstellensatz refutation

Let $\mathcal{E}$ be set of aX^{α} such that $aX^{\alpha} = 1$ derived by width-$10d$ elimination

Relation: $aX^{\alpha} \sim bX^{\beta}$ if $aX^{\alpha} = E \cdot bX^{\beta}$ over $\{\pm 1\}^n$ for some $E \in \mathcal{E}$

Claim: equivalence relation on degree-d terms

symmetry uses $X_i^2 = 1$

transitivity uses width $> 2d$

Define:

$$M(X^{\alpha}) = \begin{cases} 1 & \text{if } X^{\alpha} \sim 1 \\ -1 & \text{if } X^{\alpha} \sim -1 \\ 0 & \text{otherwise} \end{cases}$$

Want to show:

No width-$10d$ Gaussian refutation $\longrightarrow$ no degree-d Positivstellensatz refutation

Let $\mathcal{E}$ be set of aX^α such that $aX^\alpha = 1$ derived by width-$10d$ elimination

Relation: $aX^\alpha \sim bX^\beta$ if $aX^\alpha = E \cdot bX^\beta$ over $\{\pm 1\}^n$ for some $E \in \mathcal{E}$

$$M(X^\alpha) = \begin{cases} 1 & \text{if } X^\alpha \sim 1 \\ -1 & \text{if } X^\alpha \sim -1 \\ 0 & \text{otherwise} \end{cases}$$

We wanted:

$M(1) = 1$

$M(\text{S.O.S}) \geq 0$ $\quad \forall$ S.O.S ?

$M\big(Q \cdot (\sigma_e X^e - 1)\big) = 0$ $\quad \forall e$, degree-d Q

Want to show:

No width-$10d$ Gaussian refutation $\longrightarrow$ no degree-d Positivstellensatz refutation

Let $\mathcal{E}$ be set of aX^{α} such that $aX^{\alpha} = 1$ derived by width-$10d$ elimination

Relation: $aX^{\alpha} \sim bX^{\beta}$ if $aX^{\alpha} = E \cdot bX^{\beta}$ over $\{\pm 1\}^n$ for some $E \in \mathcal{E}$

$$M(X^{\alpha}) = \begin{cases} 1 & \text{if } X^{\alpha} \sim 1 \\ -1 & \text{if } X^{\alpha} \sim -1 \\ 0 & \text{otherwise} \end{cases}$$

$M(\mathrm{S.O.S}) \geq 0$

pair up equivalence classes

orthogonal unit vectors $v_1, \ldots, v_r$

v_1	v_2	v_r
$\mathcal{C}_1^+$	$\mathcal{C}_2^+$	$\mathcal{C}_r^+$
$\mathcal{C}_1^-$	$\mathcal{C}_2^-$	$\mathcal{C}_r^-$
$-v_1$	$-v_2$	$-v_r$

Check: $M\left(X^{\alpha} X^{\beta}\right) = \langle v_{\alpha}, v_{\beta} \rangle$ → $M \succcurlyeq 0$