Semidefinite Programming Hierarchies and the Unique Games Conjecture

David Steurer
MSR New England

based on joint works with

Sanjeev Arora
Princeton University

Boaz Barak
MSR New England

Prasad Raghavendra
Georgia Tech

Approximability of CSPs, Toronto, August 2011

Semidefinite Programming Hierarchies

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

powerful algorithmic technique

Better approximations for combinatorial optimization problems?

Example: **MAX CUT** [Goemans-Williamson'94]

Unique Games Conjecture

[Khot'02,...]

strong implications for *hardness of approximation*

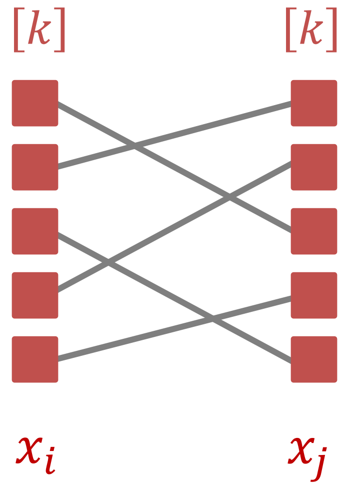
Is the conjecture true or false?

Constraint Satisfaction Problems

[e.g., works on *dense* or *pseudo-dense* instances]

important class of optimization problems

What instances are hard / easy?



UNIQUE GAMES

Input: list of constraints of form $x_i - x_j = c \bmod k$

Goal: satisfy as many constraints as possible

UNIQUE GAMES

Input: list of constraints of form $x_i - x_j = c \bmod k$

Goal: satisfy as many constraints as possible

Unique Games Conjecture (UGC) [Khot'02]

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

$\mathrm{UG}(\varepsilon)$

Input: **UNIQUE GAMES** instance with $k = \log n$ (say)

Goal: Distinguish two cases

YES: more than $1 - \varepsilon$ of constraints satisfiable

NO: less than ε of constraints satisfiable

Implications of UGC

For many basic optimization problems,
it is NP-hard to beat current algorithms
(based on simple LP or SDP relaxations)

Examples:

VERTEX COVER [Khot-Regev'03],
MAX CUT [Khot-Kindler-Mossel-O'Donnell'04,
Mossel-O'Donnell-Oleszkiewicz'05],
every **MAX CSP** [Raghavendra'08], ...

Implications of UGC

For many basic optimization problems,
it is NP-hard to beat current algorithms
(based on simple LP or SDP relaxations)

Unique Games Barrier

Example: $(\alpha_{\mathrm{GW}}+\varepsilon)$-approximation for **MAX CUT**
at least as hard as $\mathrm{UG}(\varepsilon')$

$\alpha_{\mathrm{GW}} = 0.878\ldots$
Goemans–Williamson
bound for **MAX CUT**

UNIQUE GAMES is common barrier for improving current algorithms of many basic problems

Subexponential Algorithm for Unique Games

$\mathrm{UG}(\varepsilon)$ in time $\exp\left(n^{\varepsilon^{1/3}}\right)$

Time vs Approximation Trade-off

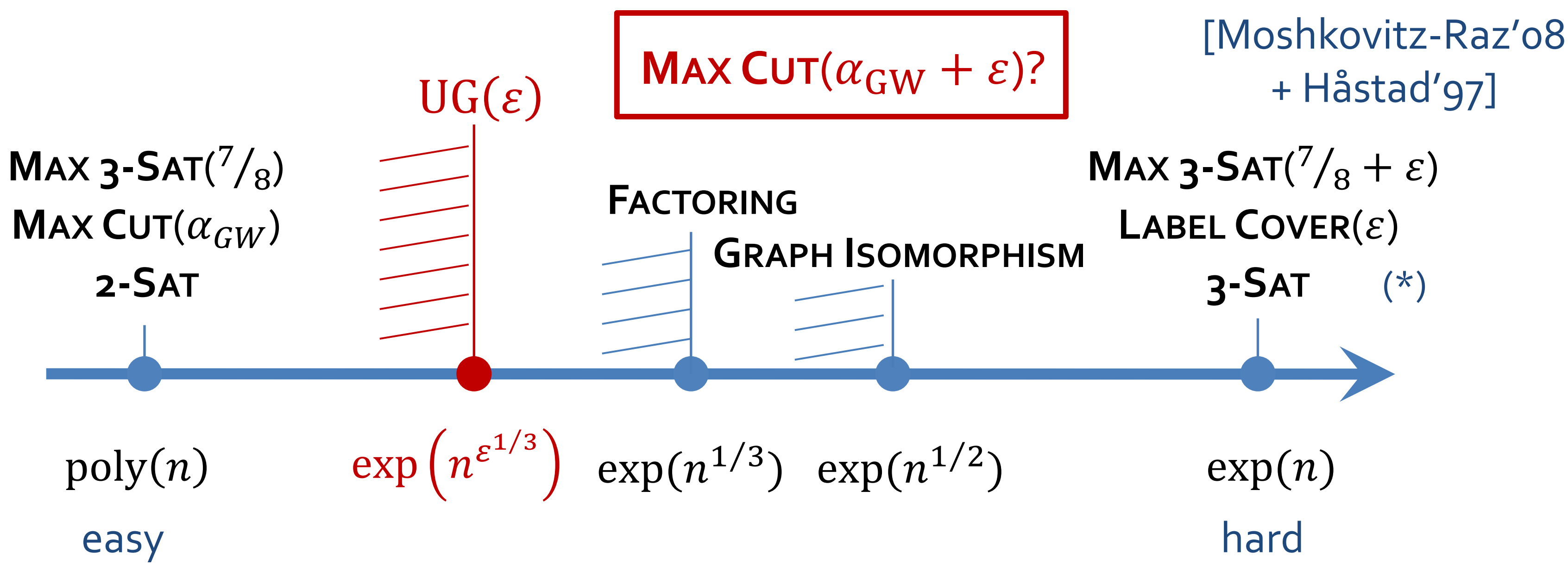
Input: **UNIQUE GAMES** instance with alphabet size k such that $1-\varepsilon$ of constraints are satisfiable,

Output: assignment satisfying $1-C\sqrt{\varepsilon}$ of constraints

Time: $\exp\left(k\, n^{1/C^{2/3}}\right)$

Subexponential Algorithm for Unique Games

UG(ε) in time $\exp\left(n^{\varepsilon^{1/3}}\right)$



(*) assuming *Exponential Time Hypothesis* [Impagliazzo-Paturi-Zane'01]
(3-Sat has no $2^{o(n)}$ algorithm)

Subexponential Algorithm for Unique Games

$\mathrm{UG}(\varepsilon)$ in time $\exp\left(n^{\varepsilon^{1/3}}\right)$

Consequences

NP-hardness reduction for $\mathrm{UG}(\varepsilon)$ must have blow-up $n^{1/\varepsilon^{1/3}}$ (*)
→ rules out certain classes of reductions for proving UGC

Analog of UGC with subconstant ε (say $\varepsilon = 1/\log\log n$) is false (*)
(*contrast*: subconstant hardness for **LABEL COVER** [Moshkovitz-Raz'08])

(*) assuming 3-SAT does not have subexponential algorithms, $\exp(n^{o(1)})$

Subexponential Algorithm for Unique Games

$\mathrm{UG}(\varepsilon)$ in time $\exp\left(n^{\varepsilon^{1/3}}\right)$

here: via semidefinite programming hierarchies [Sherali-Adams'90, Lovász-Schrijver'91, Lasserre'01]

general framework for rounding SDP hierarchies

(especially for "*pseudo-random*" instances of 2-CSPs)

constraint graph has few significant eigenvalues

Unique Games

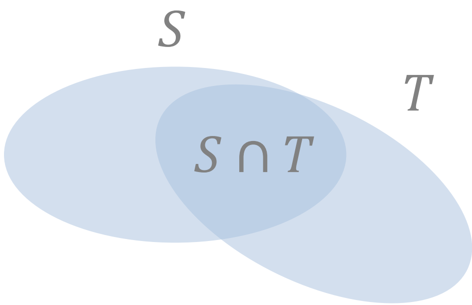
Input: list of constraints of form $x_i - x_j = c \bmod k$

Goal: satisfy as many constraints as possible

What we want:

$X_1, \dots, X_n$ *jointly distributed* random variables over $[k]$

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



time $n^{O(r)}$

distributions X_S over $[k]^S$ for all S with $|S| \le r$

consistency

same marginal for $S \cap T$ in distributions X_S and X_T

positive semidefiniteness

X_{ia} = indicator of $\{X_i = a\}$

$\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = \alpha)$ p.s.d. for all S with $|S| \le r$-2

r-level SDP hierarchy:

~~*What we want:*~~

$X_1, \dots, X_n$ r-local ~~*jointly distributed*~~ random variables over $[k]$

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

Goal: produce *global* random variables $X'_1, \dots, X'_n$

$\{X'_i, X'_j\} \approx \{X_i, X_j\}$ for most constraints $x_i - x_j \equiv c$

Here: iterative procedure for $r = n^{O(\varepsilon^{1/3})}$ [Arora-Barak-S.'10 + Barak-Raghavendra-S.'11]

Components of iterative procedure

Rounding

sample variables independently according to their marginals

Conditioning

pick a vertex j and sample X_j

condition $X_1, \ldots, X_n$ on sample for X_j

Partitioning

find vertex subset S

break dependence between X_S and $X_{V \setminus S}$

general framework for rounding SDP hierarchies

(Pairwise) Correlation

$\mathrm{Corr}(X_i, X_j)$ measures how much the distribution of X_i changes when conditioned on X_j

Examples:

$\mathrm{Corr}(X_i, X_j) = 0 \quad \Leftrightarrow \quad X_i$ and X_j independent

$\mathrm{Corr}(X_i, X_j) = 1 \quad \Leftrightarrow \quad X_i$ fixed after fixing X_j

(Pairwise) Correlation

$\mathrm{Corr}(X_i, X_j)$

statistical distance between $\{X_i, X_j\}$ and $\{X_i\}\{X_j\}$

$$\max_c \sum_a \left|\mathrm{Cov}\left(X_{ia}, X_{j(a+c)}\right)\right|$$

decrease in variance when conditioning on X_j

$$\sum_a \left(Var(X_{ia}) - Var(X_{ia}|X_j)\right)$$

similar to *mutual information*
$I(X;Y) = H(X) - H(X \mid Y)$

Important fact:

can approximate $\mathrm{Corr}(X_i, X_j)$ by Gram matrix of unit vectors (tensoring trick)

Rounding	*Conditioning*	*Partitioning*

sample variables independently according to their marginals

$\mathrm{Corr}(X_i, X_j) \approx$ statistical distance between independent and correlated sampling

If $\mathrm{Corr}(X_i, X_j) < 1 - O(\varepsilon)$ then *independent sampling* satisfies constraint with probability $\Omega(\varepsilon)$

Rounding fails $\Rightarrow$ $\mathbf{E}_{i \sim j}\mathrm{Corr}(X_i, X_j) > 1 - O(\varepsilon)$

Local Correlation
(over edges of constraint graph)

pick a vertex j and sample X_j

condition $X_1, \dots, X_n$ on sample for X_j

Issue:

computationally expensive

(level r $\rightarrow$ level $r-1$)

$\mathrm{Corr}(X_i, X_j)$ measures decrease in variance when conditioning on X_j

Idea:

condition on vertex j only if $\mathbf{E}_i \mathrm{Corr}(X_i, X_j) > n^{-\beta}$

$\Rightarrow$ can condition at most n^{β} times on such vertices

Conditioning fails $\Rightarrow$ $\mathbf{E}_{i,j} \mathrm{Corr}(X_i, X_j) < n^{-\beta}$

Global Correlation
(over random vertex pairs)

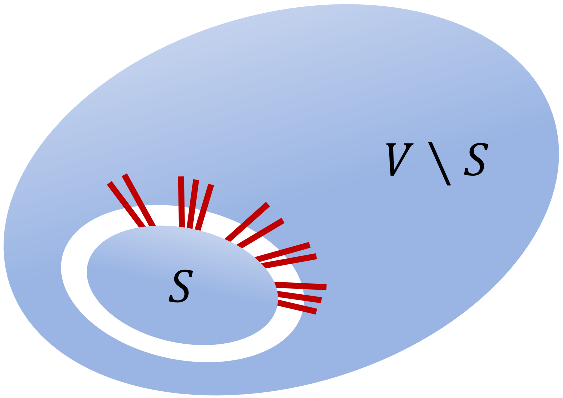
find vertex subset S

break dependence between X_S and $X_{V \setminus S}$

Issue:

destroys correlation for constraints between S and $V \setminus S$

in total: should break at most $^1/_2$ of the constraints



***Wanted**:*

set S with *small expansion*

& *small cardinality*

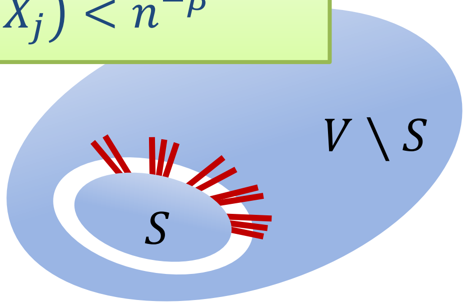
Rounding *Conditioning* ***Partitioning***

fails only if *local correlation high*
$\mathbf{E}_{i\sim j}\mathrm{Corr}(X_i, X_j) > 1 - O(\varepsilon)$

fails only if *global correlation low*
$\mathbf{E}_{i,j}\mathrm{Corr}(X_i, X_j) < n^{-\beta}$

find vertex subset S

break dependence between X_S and $X_{V\setminus S}$



Correlation Propagation

For endpoints of random path of length $t = \Omega(\beta/\varepsilon \cdot \log n)$

$\mathbf{E}_{i\sim_t j}\mathrm{Corr}(X_i, X_j) > (1 - O(\varepsilon))^t \gg n^{-\beta/2}$

$\mathrm{Corr}(X_i, X_j)$ is Gram matrix of unit vectors

$\Rightarrow$ random walk stuck for t steps on $n^{-\beta/2}$ fraction of vertices

$\Rightarrow \exists$ vertex set S with $|S| < |V|/n^{\beta/2}$ and expansion $O\left(\sqrt{\varepsilon/\beta}\right)$

A vertex is cut in $\leq 2/\beta$ partitioning steps $\Rightarrow$ break only $O\left(\sqrt{\varepsilon/\beta^3}\right)$ edges

■

Can we avoid partitioning?

Yes, for a large family of instances!

Local vs Global Correlation and Higher Eigenvalues

$$\mathbf{E}_{i\sim j}\mathrm{Corr}(X_i, X_j) > \lambda$$
$$\mathbf{E}_{i,j}\mathrm{Corr}(X_i, X_j) < 1/r$$

$\Rightarrow$ constraint graph has at least $\varepsilon \cdot r$ eigenvalues larger than $\lambda - \varepsilon$

Rounding — *Conditioning* — *Partitioning*

fails only if *local correlation high*
$\mathbf{E}_{i\sim j}\mathrm{Corr}\big(X_i, X_j\big) > 1 - O(\varepsilon)$

fails only if *global correlation low*
$\mathbf{E}_{i,j}\mathrm{Corr}\big(X_i, X_j\big) < 1/r$

Local vs Global Correlation and Higher Eigenvalues

$\mathbf{E}_{i\sim j}\mathrm{Corr}\big(X_i, X_j\big) > \lambda$

$\mathbf{E}_{i,j}\mathrm{Corr}\big(X_i, X_j\big) < 1/r$

$\Rightarrow$ constraint graph has at least $\varepsilon \cdot r$ eigenvalues larger than $\lambda - \varepsilon$

Corollary

r^{th} levels of hierarchy solves $\mathrm{UG}(\varepsilon)$ if constraint graph has at most $\varepsilon \cdot r$ eigenvalues larger than $1 - O(\varepsilon)$

[improves Kolla'10]

Rounding — *Conditioning* — *Partitioning*

fails only if *local correlation high*
$\mathbf{E}_{i\sim j}\mathrm{Corr}(X_i, X_j) > O(\eta)$

fails only if *global correlation low*
$\mathbf{E}_{i,j}\mathrm{Corr}(X_i, X_j) < 1/r$

Local vs Global Correlation and Higher Eigenvalues

$\mathbf{E}_{i\sim j}\mathrm{Corr}(X_i, X_j) > \lambda$

$\mathbf{E}_{i,j}\mathrm{Corr}(X_i, X_j) < 1/r$

$\Rightarrow$ constraint graph has at least $\varepsilon \cdot r$ eigenvalues larger than $\lambda - \varepsilon$

Approximation Scheme for ***Max Cut***

[improves works on (pseudo-)dense problems: Arora-Karger-Karpinski'95, Fernandez de la Vega'96, ...]

r^{th} levels of hierarchy finds optimal cut up to $\mathrm{O}(\eta)$ fraction of edges if graph has at most $\eta \cdot r$ eigenvalues larger than η

canonical SDP integrality gap instances have rapidly decaying eigenvalues ($r = \mathrm{polylog}(n)$)

More SDP-hierarchy algorithms

For general **2-CSP**: [Barak-Raghavendra-S.'11]

PTAS if constraint graph is random (degree $\gg$ alphabet)

Subsequent works:

better **3-COLORING** approximation on some graph families [Arora-Ge'11]

better approximations for **MAX-BISECTION** [Raghavendra-Tan'11]

Independent work: [Guruswami-Sinop'11]

approximation schemes for quadratic integer programming
with p.s.d. objective & few relevant eigenvalues

Open Questions

What else can be done in subexponential time?

Better approximations for **Max Cut**, **Vertex Cover** on general instances?

Example: $f(\varepsilon)$-approximation for **Sparsest Cut** in time $\exp(n^{\varepsilon})$?

Towards settling the Unique Games Conjecture

SDP integral gap instances with $\mathrm{poly}(n)$ large eigenvalues?

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

∃ gap instances with $\mathrm{qpoly(n)} = 2^{(\log n)^{\Omega(1)}}$ large eigenvalues

Also: gap remains for $\mathrm{qpoly}(n)$ levels of an SDP hierarchy

Thank you! Questions?