Approximate Constraint Satisfaction requires Large LP Relaxations

David Steurer
Cornell

Siu On Chan
MSR

James R. Lee
Washington

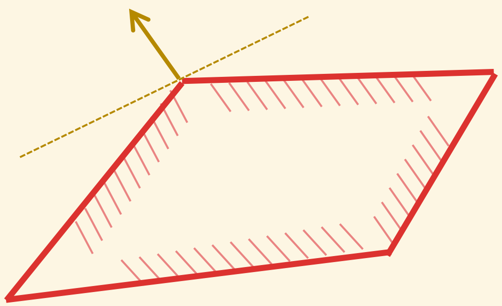
Prasad Raghavendra
Berkeley

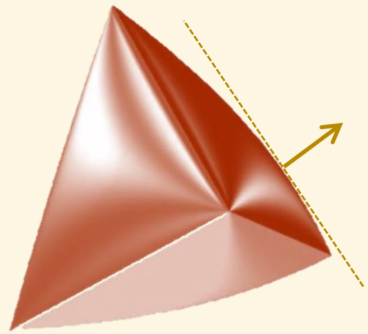
TCS+ Seminar, December 2013

best-known (approximation) algorithms for many combinatorial optimization problems:

Max Cut, Traveling Salesman, Sparsest Cut, Steiner Tree, ...

common core = linear / semidefinite programming (LP/SDP)





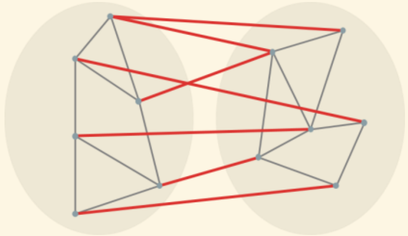
$$\max \quad \frac{1}{|E|}\sum_{ij\in E}\mu_{ij}$$
$$\text{s.t.} \quad \mu_{ij}-\mu_{ik}-\mu_{kj}\le 0$$
$$\mu_{ij}+\mu_{ik}+\mu_{kj}\le 2$$
$$0\le \mu_{ij}\le 1$$

LP / SDP relaxations

particular kind of *reduction* from hard problem to LP/SDP
running time: polynomial in ***size*** of relaxation

what guarantees are possible for approximation and running time?

example: basic LP relaxation for Max Cut



Max Cut: Given a graph, find bipartition $x \in \{\pm 1\}^n$ that cuts as many edges as possible

maximize $\frac{1}{|E|} \sum_{ij \in E} \mu_{ij}$

subject to $\mu_{ij} - \mu_{ik} - \mu_{kj} \leq 0$

$\mu_{ij} + \mu_{ik} + \mu_{kj} \leq 2$

~~$\mu_{ij} \in \{0,1\}$~~ $\mu_{ij} \in [0,1]$

intended solution

$$(\mu_x)_{ij} = \begin{cases} 1, & \text{if } x_i \neq x_j, \\ 0, & \text{otherwise.} \end{cases}$$

~~integer~~ linear program

(relax integrality constraint)

$O(n^3)$ inequalities

depend only **on instances size** (but not instance itself)

approximation guarantee

optimal value of instance
vs.
optimal value of LP relaxation

challenges

many possible relaxations for same problem
small difference syntactically → big difference for guarantees
goal: identify "right" polynomial-size relaxation

hierarchies = systematic ways to generate relaxations
best-known: Sherali-Adams (LP), sum-of-squares/Lasserre (SDP); best possible?
goal: compare hierarchies and general LP relaxations

often: more complicated/larger relaxations → better approximation
P ≠ NP predicts limits of this approach; can we confirm them?
goal: understand computational power of relaxations

Rule out that poly-size LP relaxations show* P = NP*?

hierarchies

[Lovász–Schrijver, Sherali–Adams, Parrilo / Lasserre]

great variety (sometimes different ways to apply same hierarchy)
current champions: Sherali–Adams (LP) & sum-of-squares / Lasserre (SDP)

connections to *proof complexity* (Nullstellensatz and Positivstellensatz refutations)

lower bounds

[Mathieu–Fernandez de la Vega
Charikar–Makarychev–Makarychev]

Sherali-Adams requires size $2^{n^{\Omega(1)}}$ to beat ratio ½ for Max Cut

[Grigoriev, Schoenebeck]

sum-of-squares requires size $2^{\Omega(n)}$ to beat ratio $^7/_8$ for Max 3-Sat

upper bounds

implicit: many algorithms (e.g., Max Cut and Sparsest Cut) [Goemans-Williamson, Arora-Rao-Vazirani]

explicit: Coloring, Unique Games, Max Bisection [Chlamtac, Arora-Barak-S., Barak-Raghavendra-S., Raghavendra-Tan]

lower bounds for general LP formulations (extended formulations)

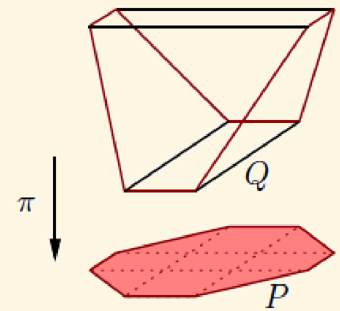
characterization; symmetric formulations for TSP & matching [Yannakakis'88]

general, exact formulations for TSP & Clique [Fiorini–Massar–Pokutta–Tiwary–de Wolf'12]

approximate formulations for Clique [Braun–Fiorini–Pokutta–S.'12 Braverman–Moitra'13]

general, exact formulation for maximum matching [Rothvoß'13]

geometric idea: complicated polytopes can be ***projections*** of simple polytopes



universality result for LP relaxations of Max CSPs [this talk]

general polynomial-size LP relaxations are no more powerful than polynomial-size Sherali-Adams relaxations

also holds for almost quasi-polynomial size

concrete consequences

unconditional lower bound in powerful computational model

confirm non-trivial prediction of P≠NP:

poly-size LP relaxations cannot achieve 0.99 approximation for Max Cut, Max 3-Sat, or Max 2-Sat (NP-hard approximations)

approximability and UGC:

poly-size LP relaxation cannot refute Unique Games Conjecture (cannot improve current Max CSP approximations)

separation of LP relaxation and SDP relaxation:

poly-size LP relaxations are strictly weaker than SDP relaxations for Max Cut and Max 2Sat

universality result for LP relaxations of Max CSPs [this talk]

general polynomial-size LP relaxations are no more powerful than polynomial-size Sherali-Adams relaxations

also holds for almost quasi-polynomial size

for concreteness: focus on Max Cut

notation: $\text{cut}_G(x)$ = fraction of edges that bipartition x cuts in G
Max Cut_n = Max Cut instances / graphs on n vertices

compare: general $n^{(1-\varepsilon)d}$-size LP relaxation for Max Cut_n
vs. n^d-size Sherali-Adams relaxations for Max Cut_n

general LP relaxation for Max Cut_n

example linearization

$L_G(\mu) = \frac{1}{|E|}\sum_{ij\in E}\mu_{ij}$

$(\mu_x)_{ij} = \begin{cases} 1, & \text{if } x_i \neq x_j, \\ 0, & \text{otherwise.} \end{cases}$

linearization

$G \mapsto L_G\colon \mathbb{R}^m \to \mathbb{R}$ linear

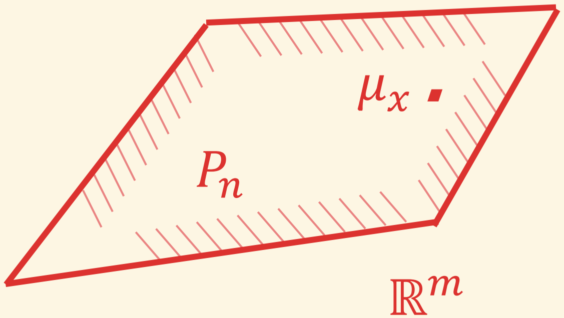
$x \mapsto \mu_x \in \mathbb{R}^m$

such that $L_G(\mu_x) = \mathrm{cut}_G(x)$

polytope of size R

$P_n \subseteq \mathbb{R}^m$, at most R facets,

$\{\mu_x\}_{x\in\{\pm1\}^n} \subseteq P_n$



same polytope for all instances of size n
makes sense because solution space
for Max Cut depends only on n

computing with size-R LP relaxation $\mathcal{L}$

input	computation	output
graph G on n vertices	maximize $L_G(\mu)$ subject to $\mu \in P_n$	value $\mathcal{L}(G) = \max_{\mu \in P} L_G(\mu)$

poly(R)-time computation

always upper-bounds Opt G
how far in the worst-case?

approximation ratio α

$\mathcal{L}(G) \leq \alpha \cdot \mathrm{Opt}(G)$

for all $G \in \text{Max Cut}_n$

(c, s)-approximation

$\mathrm{Opt}(G) \leq s \Rightarrow \mathcal{L}(G) \leq c$

for all $G \in \text{Max Cut}_n$

general computational model—how to prove lower bounds?

geometric characterization (à la Yannakakis'88)

every **size-R LP relaxation $\mathcal{L}$ for Max Cut$_n$**

corresponds to

nonnegative functions $q_1, \dots, q_R$ $: \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ such that

$\mathcal{L}(G) \leq c$ iff $c - \mathrm{cut}_G = \sum_r \lambda_r q_r$ and $\lambda_1, \dots, \lambda_R \geq 0$

for all $G \in$ Max Cut$_n$

certifies $\mathrm{cut}_G \leq c$ over $\{\pm 1\}^n$
canonical linear program
of size R

example

2^n standard basis functions correspond to
exact 2^n-size LP relaxation for Max Cut$_n$

geometric characterization (à la Yannakakis'88)

every **size-R LP relaxation $\mathcal{L}$ for Max Cut$_n$**

corresponds to

nonnegative functions $q_1, \dots, q_R$: $\{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ such that

$\mathcal{L}(G) \leq c$ iff $c - \mathrm{cut}_G = \sum_r \lambda_r q_r$ and $\lambda_1, \dots, \lambda_R \geq 0$

for all $G \in$ Max Cut$_n$

intuition: all inequalities for functions on $\{\pm 1\}^n$ with local proofs

connection to Sherali-Adams hierarchy

n^d-size Sherali-Adams relaxation for Max Cut$_n$

exactly corresponds to

generated by n^d "base juntas"

d-*junta* = function on $\{\pm 1\}^n$ depends on $\leq d$ coordinates

nonnegative combinations of nonnegative d-*juntas* on $\{\pm 1\}^n$

geometric characterization (à la Yannakakis'88)

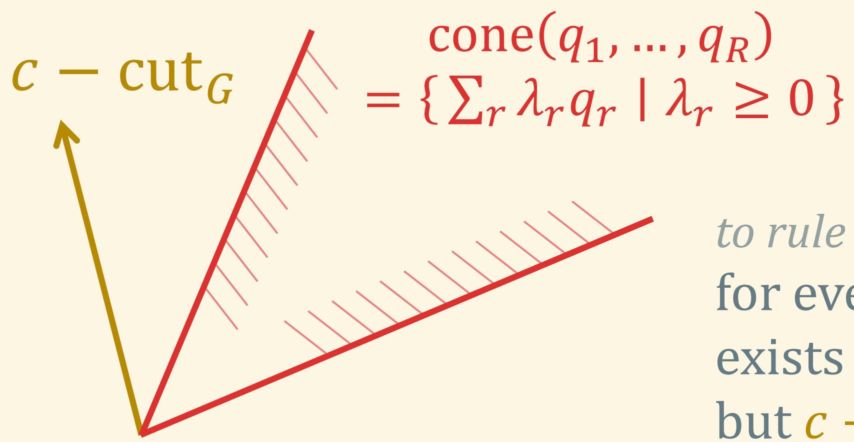
every **size-R LP relaxation $\mathcal{L}$ for Max Cut$_n$**

corresponds to

nonnegative functions $q_1, \ldots, q_R$: $\{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ such that

$\mathcal{L}(G) \leq c$ iff $c - \mathrm{cut}_G = \sum_r \lambda_r q_r$ and $\lambda_1, \ldots, \lambda_R \geq 0$

for all $G \in$ Max Cut$_n$



to rule out (c,s)-approx. by size-R LP relaxation, show:

for every size-R nonnegative cone,
exists $G \in$ Max Cut$_n$ with Opt$(G) \leq s$
but $c - \mathrm{cut}_G$ outside of cone

lower-bound for Sherali–Adams relaxations of size n^d

↓

lower-bounds for size-n^d nonneg. cones with *restricted functions*

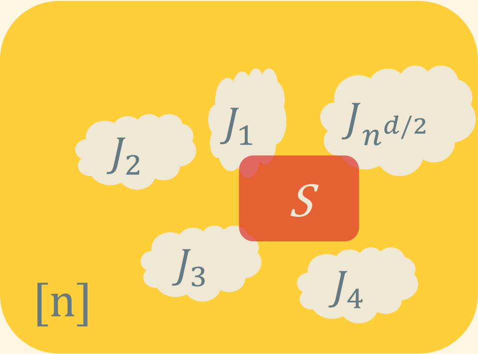
d-juntas → n^{ε}-juntas → non-spiky → general

↓

lower-bound for general LP relaxations of size $n^{(1-\varepsilon)d}$

from d-juntas to n^ε-juntas

let $q_1, \dots, q_R$ be nonneg. n^ε-juntas on $\{\pm 1\}^n$ for $R = n^{(1-10\varepsilon)d}$



want: subset $S \subseteq [n]$ of size $m \approx n^\varepsilon$
where functions behave like d-juntas

let $J_1, \dots, J_R$ be junta-coordinates of $q_1, \dots, q_R$

claim: there exists subset $S \subseteq [n]$ of size $m = n^\varepsilon$ such that
$|J_r \cap S| \leq d$ for all $r \in [R]$

proof: choose S at random

$$\mathbb{P}\{|S \cap J_r| > d\} \leq \left(\frac{|S|}{n} \cdot |J_r|\right)^d = n^{-(1-2\varepsilon)d}$$

→ can afford **union bound** over R junta sets $J_1, \dots, J_R$

lower-bound for Sherali–Adams relaxations of size n^d

↓

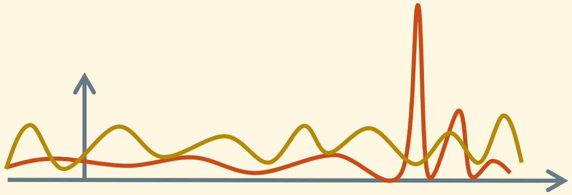
lower-bounds for size-n^d nonneg. cones with *restricted functions*

d-juntas → n^{ε}-juntas → non-spiky → general

↓

lower-bound for general LP relaxations of size $n^{(1-\varepsilon)d}$

from n^ε-juntas to non-spiky functions



let q be a nonnegative function on $\{\pm 1\}^n$ with $\mathbb{E}q = 1$

non-spiky: $\max q \leq 2^t$

junta structure lemma:

can approximate q by nonnegative n^ε-junta q',

error $\eta = q - q'$ satisfies $|\hat{\eta}_S|^2 \leq td/n^\varepsilon$ for $|S| < d$

small low-degree Fourier coefficients

proof:

nonnegative function q → probability distribution over $\{\pm 1\}^n$,

+1/-1 rand. variables $X_1, \dots, X_n$ (dependent)

non-spiky → entropy $H(X_1, \dots, X_n) \geq n - t$

want: $J \subseteq [n]$ of size n^ε such that $\forall S \subseteq [n] \setminus J$. $\{X_S \mid X_J\} \approx$ uniform, that is,

($|S| < d$) $|S| - H(X_S \mid X_J) \leq \beta$ for $\beta = \frac{td}{n^\varepsilon}$

construction: start with $J = \emptyset$; as long as bad S exists, update $J \leftarrow J \cup S$

analysis: total entropy defect $\leq t$ → stop after $\frac{t}{\beta}$ iterations → $|J| \leq \frac{dt}{\beta} = n^\varepsilon$

lower-bound for Sherali–Adams relaxations of size n^d

↓

lower-bounds for size-n^d nonneg. cones with *restricted functions*

d-juntas → n^{ε}-juntas → non-spiky → general

↓

lower-bound for general LP relaxations of size $n^{(1-\varepsilon)d}$

from non-spiky functions to general functions

let $q_1, \ldots, q_R$ be general nonneg. functions on $\{\pm 1\}^n$ for $R = n^d$

non-spiky

claim: exists nonneg. $q_1', \ldots, q_R'$ such that $q_i' \leq n^{2d}$, $\mathbb{E} q_i' = 1$ and

$$\mathrm{cone}(q_1, \ldots, q_R) \approx \mathrm{cone}(q_1', \ldots, q_R')$$

proof: truncate functions carefully

intuition: $c - \mathrm{cut}_G$ is non-spiky. Thus, spiky q_i don't help!

lower-bound for Sherali–Adams relaxations of size n^d

↓

lower-bounds for nonneg. cones of size n^d with *restricted functions*

d-juntas ⟶ n^ε-juntas ⟶ non-spiky ⟶ general

↓

lower-bound for general LP relaxations of size $n^{(1-\varepsilon)d}$

open problems

1. **LP size 2^{n^ε}** 2. **beyond CSPs (e.g., TSP)** 3. **SDPs**

lower-bound for Sherali–Adams relaxations of size n^d

↓

lower-bounds for nonneg. cones of size n^d with *restricted functions*

d-juntas ⟶ n^ε-juntas ⟶ non-spiky ⟶ general

Recent: for symmetric relaxations [Lee-Raghavendra-S.-Tan'13]

open problems

Thank you!

1. **LP size 2^{n^ε}**

2. **beyond CSPs (e.g., TSP)**

3. **SDPs**