Steurer's Conjecture and Sparsest Cut (II)

Reading Group Spring '22

Alex Tung

16 August 2022

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

Before start…

- I couldn't yet figure out how Steurer's conjecture implies subexponential sparsest cut, after a month
- There's some technical difficulties that we will discuss in the end

- Today we will follow [BRS] and deal with low-rank case

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

SDP hierarchy: setup

- Combinatorial optimization
- Objective function $g: \{0, 1\}^n \to \mathbb{R}$, extensible to $g: [0, 1]^n \to \mathbb{R}$
  - We are minimizing $g(x)$ over $x \in \{0,1\}^n$, g convex
- Subject to linear constraints: $x \in K \coloneqq \{x' : Ax' \geq b\}$

Lasserre hierarchy

Definition 1. Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$. We define the *$t$-th level of the Lasserre hierarchy* $\text{Las}_t(K)$ as the set of vectors $y \in \mathbb{R}^{2^{[n]}}$ that satisfy

$$M_t(y) := (y_{I \cup J})_{|I|,|J| \leq t} \succeq 0; \quad M_t^{\ell}(y) := \Big(\sum_{i=1}^{n} A_{\ell i} y_{I \cup J \cup \{i\}} - b_\ell y_{I \cup J} \Big)_{|I|,|J| \leq t} \succeq 0 \quad \forall \ell \in [m]; \quad y_{\emptyset} = 1.$$

The matrix $M_t(y)$ is called the *moment matrix of* y and the second type $M_\ell^t(y)$ is called *moment matrix of slacks.* Furthermore, let $\text{Las}_t^{\text{proj}}(K) := \{(y_{\{1\}}, \ldots, y_{\{n\}}) \mid y \in \text{Las}_t(K)\}$ be the projection on the original variables.

- Increasingly tight relaxations of the integer program
- y_I ($I \subseteq [n]$) as joint probabilities

$y_I = \Pr[X_I = 1] = \Pr[\bigwedge_{i \in I} X_i = 1]$

Geometric interpretation of $M_t(y)$

- We have $M_t(y) \coloneqq \left(y_{I\cup J}\right)_{|I|,|J|\leq t} \succcurlyeq 0$
- Meaning: there exists vectors $\{v_I\}_{|I|\leq t} \subseteq \mathbb{R}^{2^n}$ such that $\langle v_I, v_J\rangle = y_{I\cup J}$
- These vector solutions are useful in (i) interpreting rounding algorithm and (ii) relating Steurer's conjecture and sparsest cut

Projecting Lasserre solutions

Lemma 2. *For $t \geq 1$, let $y \in \mathrm{LAS}_t(K)$ and $i \in [n]$ be a variable with $0 < y_i < 1$. If we define*

$$z_I^{(1)} := \frac{y_{I \cup \{i\}}}{y_i} \quad \text{and} \quad z_I^{(0)} := \frac{y_I - y_{I \cup \{i\}}}{1 - y_i}$$

then we have $y = y_i \cdot z^{(1)} + (1 - y_i) \cdot z^{(0)}$ with $z^{(0)}, z^{(1)} \in \mathrm{LAS}_{t-1}(K)$ and $z_i^{(0)} = 0, z_i^{(1)} = 1$.

Corresponding vector solution:

$$v_I^{(1)} := \frac{v_{I \cup \{i\}}}{\|v_i\|}, v_I^{(0)} := \frac{v_I - v_{I \cup \{i\}}}{\|v_\emptyset - v_i\|}$$

Goals of today

- Study local distribution in more detail
- State and prove key results relating global and local correlations
- Discuss how to apply Lasserre hierarchy to sparsest cut

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

Local assignments

- We mentioned that Lasserre variables $y_I = \Pr[\wedge_{i\in I} X_i = 1]$
- They are enough to give probabilities on different assignments (to variables in I)

$\{j : f(j) = 1\}$

- Given level-t Lasserre solution (y_I), for $|I| \le t$ and $f: I \to \{0,1\}$ define $y_I(f) := \sum_{I' : f^{-1}(1) \subseteq I' \subseteq I} (-1)^{|I' \setminus f^{-1}(1)|} \cdot y_{I'}$

$y_I(f) = \Pr[\bigwedge_{i\in I} X_i = f(i)]$

- This is the language used in [BRS] and [GS]

Explanation

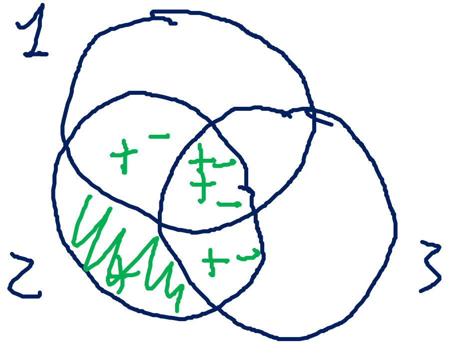
- Definition: $y_I(f) \coloneqq \sum_{I' : f^{-1}(1) \subseteq I' \subseteq I} (-1)^{|\underline{I' \setminus f^{-1}(1)}|} \cdot y_{I'}$
- Interpret as $y_I(f) = \Pr[\wedge_{i \in I} X_i = f(i)]$
- Based on inclusion-exclusion. Example: $y_{\{1,2,3\}}(010)$

$\{2\} \subseteq I' \subseteq \{1,2,3\}$

$$y_{123}(010) = (-1)^{1-1} \cdot y_2 + (-1)^{2-1} \cdot y_{12} + (-1)^{2-1} \cdot y_{23} + (-1)^{3-1} \cdot y_{123}$$

$(I' = \{2\})$ $(I' = \{1,2\})$

$$= y_2 - y_{12} - y_{23} + y_{123}$$



Vector solutions

- For each subset $|I| \leq t$ and assignment $f: I \to \{0, 1\}$, define a vector $v_I(f)$ accordingly:

$$v_I(f) \coloneqq \sum_{I': f^{-1}(1) \subseteq I' \subseteq I} (-1)^{|I' \setminus \mathrm{f}^{-1}(1)|} \cdot v_{I'}$$

Properties

- We can derive properties of $y_I(f)$ and $v_I(f)$ from properties of y_I and v_I:
  - (a) $\langle v_I(f), v_J(g)\rangle = 0$ if f and g are inconsistent $\langle V_I, V_J\rangle = y_{I\cup J}$
  - (b) $\langle v_I(f), v_J(g)\rangle = \langle v_{I'}(f'), v_{J'}(g')\rangle$ if $I \cup J = I' \cup J'$ and $f \cup g = f' \cup g'$
  - (c) (Marginals) $\|v_i(0)\|^2 + \|v_i(1)\|^2 = 1$ $\langle V_i(0), V_i(0)\rangle = y_i(0) = \Pr[X_i = 0]$
  - (d) (Marginals) $v_I\big(f \cup (i \mapsto 0)\big) + v_I\big(f \cup (i \mapsto 1)\big) = v_{I\setminus\{i\}}(f)$ for $f : I \setminus \{i\} \to \{0,1\}$

Properties

- We can derive properties of $y_I(f)$ and $v_I(f)$ from properties of y_I and v_I:
  - (e) $y_{I\cup J}(f \cup g) = \langle v_I(f), v_J(g)\rangle$ if $f: I \to \{0,1\}$ and $g: J \to \{0,1\}$ are consistent
  - (f) $0 \leq y_I(f) \leq 1$
  - (g) $y_I(f) \geq y_{I'}(f')$ if $I' \supseteq I$ and $f'|_I = f$
  - (h) (Marginals) $y_I\big(f \cup (i \mapsto 0)\big) + y_I\big(f \cup (i \mapsto 1)\big) = y_{I\setminus\{i\}}(f)$ for $f: I \setminus \{i\} \to \{0,1\}$

    $\Pr[X_{I\setminus\{i\}} = f \wedge X_i = 0] + \Pr[X_{I\setminus\{i\}} = f \wedge X_i = 1]$

    $= \Pr[X_{I\setminus\{i\}} = f]$
  - (i) (Total probability) $\sum_{f:I\to\{0,1\}} y_I(f) = 1$

Selected proofs

We shall prove:

- (a) $\langle v_I(f), v_I(g)\rangle = 0$ if f and g are inconsistent
- (b) $\left\langle v_I(f), v_J(g)\right\rangle = \left\langle v_{I'}(f'), v_{J'}(g')\right\rangle$ if $I \cup J = I' \cup J'$ and $f \cup g = f' \cup g'$
- (d) (Marginals) $v_I\big(f \cup (i \mapsto 0)\big) + v_I\big(f \cup (i \mapsto 1)\big) = v_{I\setminus\{i\}}(f)$ for $f: I \setminus \{i\} \to \{0, 1\}$

(a) $\langle v_I(f), v_J(g)\rangle = 0$ if f and g are inconsistent

$i \in I \cap J$ s.t. $f(i) = 1$, $g(i) = 0$.

$$\sum_{f^{-1}(1) \subseteq I' \subseteq I} \; \sum_{g^{-1}(1) \subseteq J' \subseteq J} (-1)^{|I' \setminus f^{-1}(1)|} \cdot (-1)^{|J' \setminus g^{-1}(1)|} \cdot \langle V_{I'}, V_{J'} \rangle$$

Observe: $i \in I'$ always, and for each J' s.t. $i \notin J'$, there corresponds $J' \cup \{i\}$

$$(-1)^{|I' \setminus f^{-1}(1)|} \cdot [1 + (-1)] \, \cancel{\langle V_{I'}, V_{J'} \rangle} \, y_{I' \cup J'}.$$

$\langle V_{I'}, V_{J'} \rangle = \langle V_{I'}, V_{J' \cup \{i\}} \rangle$

$= y_{I' \cup J'}$ ($\because i \in I'$)

$\therefore$ Sum to 0.

(b) $\langle v_I(f), v_J(g)\rangle = \langle v_{I'}(f'), v_{J'}(g')\rangle$ if $I \cup J = I' \cup J'$ and $f \cup g = f' \cup g'$

$$\langle V_I(f), V_J(g)\rangle = \langle V_{I\cup J}(f\cup g), V_\emptyset\rangle.$$

$$\text{LHS} = \sum_{\substack{f^{-1}(1)\subseteq I'\subseteq I \\ g^{-1}(1)\subseteq J'\subseteq J}} \underbrace{(-1)^{|I'\setminus f^{-1}(1)|}\cdot(-1)^{|J'\setminus g^{-1}(1)|}}\cdot\langle V_{I'}, V_{J'}\rangle$$

$$= \sum_{(f\cup g)^{-1}(1)\subseteq K\subseteq I\cup J} \quad \sum_{\substack{I', J': \\ I'\cup J' = K}} \Big(\quad \text{''} \quad \Big)\cdot y_K.$$

To do: to show that $\displaystyle\sum_{I'\cup J'=K}(\;\text{''}\;) = (-1)^{|K\setminus(f\cup g)^{-1}(1)|}$.

(Almost) blank page

(d) (Marginals) $v_I\big(f \cup (i \mapsto 0)\big) + v_I\big(f \cup (i \mapsto 1)\big) = v_{I\setminus\{i\}}(f)$ for $f: I \setminus \{i\} \to \{0,1\}$

$$\mathrm{LHS} = \Bigg[\underbrace{\sum_{f^{-1}(1) \subseteq I' \subseteq I} (-1)^{|I' \setminus f^{-1}(1)|} \cdot V_{I'}}_{\cancel{i \in I'} \qquad \cancel{i \notin I'}} + \sum_{\substack{f^{-1}(1) \subseteq I' \subseteq I \\ i \in I'}} (-1)^{|I' \setminus f^{-1}(1)|-1} \cdot V_{I'} \Bigg]$$

$$= \sum_{\substack{f^{-1}(1) \subseteq I' \subseteq I \\ i \notin I'}} (-1)^{|I' \setminus f^{-1}(1)|} \cdot V_{I'}$$

$$= \sum_{f^{-1}(1) \subseteq I' \subseteq I \setminus \{i\}} \quad " \quad = V_{I \setminus \{i\}}(f). = \mathrm{RHS}.$$

Dependency for the other parts

- ✓ (a) $\langle v_I(f), v_I(g)\rangle = 0$ if f and g are inconsistent
- ✓ (b) $\big\langle v_I(f), v_J(g)\big\rangle = \big\langle v_{I'}(f'), v_{J'}(g')\big\rangle$ if $I \cup J = I' \cup J'$ and $f \cup g = f' \cup g'$
- (direct) (c) (Marginals) $\|v_i(0)\|^2 + \|v_i(1)\|^2 = 1$
- ✓ (d) (Marginals) $v_I\big(f \cup (i \mapsto 0)\big) + v_I\big(f \cup (i \mapsto 1)\big) = v_{I\setminus\{i\}}(f)$ for $f: I \setminus \{i\} \to \{0,1\}$
- (b) (e) $y_{I\cup J}(f \cup g) = \langle v_I(f), v_J(g)\rangle$ if $f: I \to \{0,1\}$ and $g: J \to \{0,1\}$ are consistent
- (e), (h) (f) $0 \le y_I(f) \le 1$
- (h) (g) $y_I(f) \ge y_{I'}(f')$ if $I' \supseteq I$ and $f'|_I = f$
- (d) (h) (Marginals) $y_I\big(f \cup (i \mapsto 0)\big) + y_I\big(f \cup (i \mapsto 1)\big) = y_{I\setminus\{i\}}(f)$ for $f: I \setminus \{i\} \to \{0,1\}$
- (h) (i) (Total probability) $\sum_{f:I\to\{0,1\}} y_I(f) = 1$

How projection affects $y_I(f)$ and $v_I(f)$

- Similar formula holds for $y_I^{(k)}(f)$ and $v_I^{(k)}(f)$ when conditioning on variable i taking value $k \in \{0, 1\}$. For example,

$$y_I^{(1)}(f) = y_{I\cup\{i\}}(f \cup (i \mapsto 1))/y_i$$

$$\Pr[X_I = f \mid X_i = 1] = \frac{\Pr[X_I = f \wedge X_i = 1]}{\Pr[X_i = 1]}$$

- More generally, we can condition on a partial assignment $h: J \to \{0,1\}$. For example,

$$y_I^{(h)}(f) = y_{I\cup J}(f \cup h)/y_J(h)$$

Summary

- For each partial assignment $f : I \to \{0, 1\}$, there corresponds ~~marginal~~ probability $y_I(f)$ and a vector $v_I(f)$
- Definition requires only y_I and v_I and uses inclusion-exclusion
- Has nice properties that justify the probability interpretation that $y_I(f) = \Pr[X_I = f]$

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

Propagation Sampling

Recall that we round level-t Lasserre solutions as follows:

- (1) Pick variables $i_t, i_{t-1}, \dots, i_1$ one by one and condition on their values. Conditioning according to local distribution $y_I(f)$
- (2) For other variables i', assign value independently, according to marginal $y_{i'}$ (after conditioning)

Graphs

- Assume there is a given graph $G = (V, E)$, and the objective is a linear combination of y_i for $i \in V$ and $y_{i,j}$ for $(i,j) \in E$
- This captures all 2-CSP's (with two labels)

- ϵ-threshold rank of G ($rank_{\geq \epsilon}(G)$) is defined as the number of eigenvalues of $A(G)$ that are $\geq \epsilon$

normalized

Main Theorem

$r = \text{rank}_{\geqslant \Omega(\varepsilon^2)}(G)/\varepsilon^4$.

Theorem 5.7. *Let* $\varepsilon > 0$ *and* $r = O(k) \cdot \text{rank}_{\geqslant \Omega(\varepsilon/k)^2}(G)/\varepsilon^4$. *Suppose that the r-round Lasserre value of the* MAX 2-CSP *instance* $\mathfrak{I}$ *is* σ. *Then, given an optimal r-round Lasserre solution, Algorithm 5.5 (Propagation Sampling) outputs an assignment with expected value at least* $\sigma - \varepsilon$ *for* $\mathfrak{I}$.

- Section 5 of [BRS] works more generally for 2-CSP's with k labels
- You may assume $k = 2$ without loss

Intuition

- When threshold rank is low, the graph is not too badly connected
- Higher-order Cheeger says that the graph doesn’t have a lot of disjoint sparse cuts
- Therefore, conditioning on enough vertices, the other vertices should be almost determined

Notations

- X_i denotes the i-th variable (distributed according to $\{y_I\}$)
- NOTE: X_i's are not jointly distributed

- X_{ia} denotes the indicator variable $[X_i = a]$

- $\{X_I\}$ denotes the distribution over possible outcomes f: $I \rightarrow \{0,1\}$. $\Pr[X_I = f(I)] = y_I(f)$.

Broad outline of proof

- Use variance $\frac{1}{\mathrm{n}}\sum_i y_i(1-y_i)$ as potential function

- To show: discrepancy $\mathbb{E}_{(i,j)\in E}\left\|\{X_iX_j\}-\{X_i\}\{X_j\}\right\|_1$ small in the end
- Prove that large discrepancy -> large drop in potential

$y_I = y_i \cdot z_I^{(1)} + (1-y_i)\cdot z_I^{(0)}$ OBJ is linear combination of the y_i's and y_{ij}'s.

- Conditioning on S doesn't change the expected objective value
- If $\mathbb{E}_{(i,j)\in E}\left\|\{X_iX_j|X_S\}-\{X_i|X_S\}\{X_j|X_S\}\right\|_1 < \eta$, by independently sampling X_{V-S} we lose at most η in objective value

Variance drop and covariance

Lemma 5.2. *For any two vertices $i, j \in V$,*

$$\operatorname{Var} X_i - \mathop{\mathbb{E}}_{\{X_j\}} \operatorname{Var}\left[X_i \mid X_j\right] \geqslant \frac{1}{k} \sum_{a,b\in[k]} \mathop{\mathbb{E}}_{\{X_{ia}, X_{jb}\}} \operatorname{Cov}(X_{ia}, X_{jb})^2 / \operatorname{Var} X_{jb}$$

Proof: Z_1, Z_2 : $\{0,1\}$-variables, then

$$\to \operatorname{Var}[Z_1] - \cancel{\operatorname{Var}} \mathop{\mathbb{E}}_{\{Z_2\}} \operatorname{Var}[Z_1 | Z_2] = \frac{\operatorname{Cov}(Z_1, Z_2)^2}{\operatorname{Var}(Z_2)}.$$

Take $Z_1 = X_{ia}$, $Z_2 = X_{jb}$. Sum over $a \in [k]$,

$$\operatorname{Var}[X_i] - \mathop{\mathbb{E}}_{\{X_{jb}\}} [X_i | X_{jb}] = \sum_{a \in [k]} \frac{\operatorname{Cov}(X_{ia}, X_{jb})^2}{\operatorname{Var} X_{jb}}.$$

Take the best index b', $\operatorname{Var}[X_i] - \mathop{\mathbb{E}}_{\{X_{jb'}\}} \operatorname{Var}[X_i | X_{jb'}] \geq \frac{1}{k} \sum_{a,b} \frac{\operatorname{Cov}(X_{ia}, X_{jb})^2}{\operatorname{Var} X_{jb}}.$

Covariance and "covariance vector"

Proposition. The matrix $\left(Cov(X_{ia}, X_{jb})\right)_{i,j\in V, a,b\in[k]}$ is PSD.

Proof: In fact we will show $Cov(X_I(f), X_J(g))$ is PSD

$$\langle v_I(f), v_J(g)\rangle = y_{I\cup J}(f\cup g) = \mathbb{E}[X_I(f) X_J(g)]$$

$$u_I(f) := v_I(f) - \|v_I(f)\|^2 \cdot v_\phi .$$

$$\langle u_I(f), u_J(g)\rangle = \langle v_I(f) - \|v_I(f)\|^2 \cdot v_\phi , v_J(g) - \|v_J(g)\|^2 \cdot v_\phi\rangle$$

$$= y_{I\cup J}(f\cup g) - y_I(f)\cdot y_J(g)$$

$$= \langle v_I(f), v_J(g)\rangle - \|v_I(f)\|^2 \cdot \underbrace{\langle v_\phi, v_J(g)\rangle}_{\|v_J(g)\|^2} = \langle v_J(g), v_J(g)\rangle = y_J(g)$$

$$- \|v_J(g)\|^2 \cdot \underbrace{\langle v_I(f), v_\phi\rangle}_{\|v_I(f)\|^2} + \|v_I(f)\|^2 \cdot \|v_J(g)\|^2 \cdot \underbrace{\langle v_\phi, v_\phi\rangle}_{1}$$

Covariance and "covariance vector"

Lemma 5.3. *Suppose that the matrix* $\big(\mathrm{Cov}(X_{ia}, X_{jb})\big)_{i\in V, a\in[k]}$ *is positive semidefinite. Then, there exists vectors* $\tilde{u}_1, \ldots, \tilde{u}_n$ *in the unit ball such that for all vertices* $i, j \in V$,

$$\frac{1}{k^2}\Big(\sum_{(a,b)\in[k]^2} \big|\mathrm{Cov}(X_{ia}, X_{jb})\big|\Big)^2 \leqslant \langle \tilde{u}_i, \tilde{u}_j\rangle \leqslant \frac{1}{k}\sum_{(a,b)\in[k]^2} \tfrac{1}{2}\Big(\frac{1}{\mathrm{Var}\,X_{ia}} + \frac{1}{\mathrm{Var}\,X_{jb}}\Big)\mathrm{Cov}(X_{ia}, X_{jb})^2\,.$$

What is the significance of this lemma?

Lemma 5.3. *Suppose that the matrix $\left(\mathrm{Cov}(X_{ia}, X_{jb})\right)_{i\in V, a\in[k]}$ is positive semidefinite. Then, there exists vectors $v_1, \ldots, v_n$ in the unit ball such that for all vertices $i, j \in V$,*

$$\frac{1}{k^2}\Big(\sum_{(a,b)\in[k]^2} \left|\mathrm{Cov}(X_{ia}, X_{jb})\right|\Big)^2 \leqslant \langle v_i, v_j\rangle \leqslant \frac{1}{k}\sum_{(a,b)\in[k]^2} \frac{1}{2}\Big(\frac{1}{\mathrm{Var}\, X_{ia}} + \frac{1}{\mathrm{Var}\, X_{jb}}\Big)\mathrm{Cov}(X_{ia}, X_{jb})^2 \,.$$

$w \in \mathbb{R}^n$

Proof: $w^{\otimes 2} = \begin{pmatrix} w_{11} \\ \vdots \\ w_{1n} \\ w_{21} \\ \vdots \\ w_{2n} \\ \vdots \\ w_{nn} \end{pmatrix}$ $w_{ij} = w_i w_j$.

Define $\tilde{u}_i := k^{-\frac{1}{2}} \sum_a u_{ia}^{\otimes 2} / \|u_{ia}\|$. Then: $u_{ia} = u_i(a)$, s.t. $\langle u_{ia}, u_{jb}\rangle = \mathrm{Cov}(X_{ia}, X_{jb})$.

$$\langle \tilde u_i, \tilde u_j\rangle = \frac{1}{k}\sum_{a,b} \frac{\langle u_{ia}^{\otimes 2}, u_{jb}^{\otimes 2}\rangle}{\|u_{ia}\|\cdot\|u_{jb}\|} = \frac{1}{k}\sum_{a,b}\frac{\mathrm{Cov}(X_{ia}, X_{jb})^2}{\sqrt{\mathrm{Var}\,X_{ia}\cdot \mathrm{Var}\,X_{jb}}}.$$

Upper bound:

Apply AM-GM to $\frac{1}{\mathrm{Var}\,X_{ia}}$ and $\frac{1}{\mathrm{Var}\,X_{jb}}$

Lower bound:

$$\Big(\sum_{a,b}\left|\mathrm{Cov}(X_{ia}, X_{jb})\right|\Big)^2 \le \sum_{a,b}\sqrt{\mathrm{Var}\,X_{ia}\,\mathrm{Var}\,X_{jb}}\cdot\sum_{a,b}\frac{\mathrm{Cov}(X_{ia}, X_{jb})^2}{\sqrt{\mathrm{Var}\,X_{ia}\cdot\mathrm{Var}\,X_{jb}}}$$

$$= k\cdot\sum_{a,b}\sqrt{\mathrm{Var}\,X_{ia}\,\mathrm{Var}\,X_{jb}}\cdot\langle\tilde u_i, \tilde u_j\rangle.$$

$\bigstar$: $\sum_a \mathrm{Var}\,X_{ia} \le \sum_a \mathbb{E} X_{ia}^2 = 1$

Unit ball:

$$(\bigstar \Rightarrow) \le k\cdot\sqrt{k}\cdot\sqrt{k}\cdot\langle\tilde u_i, \tilde u_j\rangle.$$

Local and global correlation

So far, we have:

- (Lemma 5.3) $\langle \tilde{u}_i, \tilde{u}_j \rangle \leq \frac{1}{2k} \sum_{a,b} \left(\frac{1}{Var\ X_{ia}} + \frac{1}{Var\ X_{jb}} \right) Cov\left(X_{ia}, X_{jb}\right)^2$
- (Lemma 5.2) $\frac{1}{k} \sum_{a,b} \frac{Cov\left(X_{ia}, X_{jb}\right)^2}{Var\ X_{jb}} \leq Var[X_i] - \mathbb{E}_{\{X_j\}} Var[X_i | X_j]$

- We get $\mathbb{E}_{i,j} \langle \tilde{u}_i, \tilde{u}_j \rangle \leq \mathbb{E}_{i,j} [Var[X_i] - \mathbb{E}_{\{X_j\}} Var[X_i | X_j]]$

Global correlation of $\tilde{u}_i \leq$ Expected variance drop

Local and global correlation

So far, we have:

- (Lemma 5.3) $\frac{1}{k^2}\left(\sum_{a,b}\left|Cov\left(X_{ia}, X_{jb}\right)\right|\right)^2 \leq \langle \tilde{u}_i, \tilde{u}_j \rangle$
- (Easy) $\left\|\left\{X_i X_j\right\} - \{X_i\}\left\{X_j\right\}\right\|_1 = \sum_{a,b} \left|Cov\left(X_{ia}, X_{jb}\right)\right|$

- We get $\mathbb{E}_{(i,j)\in E}\left\langle \tilde{u}_i, \tilde{u}_j \right\rangle \geq \frac{1}{k^2}\left(\mathbb{E}_{(i,j)\in E}\left\|\left\{X_i X_j\right\} - \{X_i\}\left\{X_j\right\}\right\|_1\right)^2$

Local correlation of $\tilde{u}_i \geq$ Discrepancy across edges

Local and global correlation

$$\text{Global correlation of } \tilde{u}_i \leq \text{Expected variance drop}$$
$$\text{Local correlation of } \tilde{u}_i \geq \text{Discrepancy across edges}$$

- If it is true that local correlation ≤ global correlation, then:

$$\text{Discrepancy across edges} \leq \text{Expected variance drop}$$

Local and global correlation

The following lemma shows that a violation of the local vs global correlation condition implies that the graph has high threshold rank.

Lemma 6.1. *Suppose there exist vectors* $v_1, \ldots, v_n \in \mathbb{R}^n$ *such that*

$$\mathop{\mathbb{E}}_{ij\sim G} \langle v_i, v_j \rangle \geqslant 1 - \varepsilon\,, \quad \mathop{\mathbb{E}}_{i,j\in V} \langle v_i, v_j \rangle^2 \leqslant \tfrac{1}{m}\,, \quad \mathop{\mathbb{E}}_{i\in V} \|v_i\|^2 = 1\,.$$

Then for all $C > 1$, $\lambda_{(1-1/C)m} \geqslant 1 - C \cdot \varepsilon$. *In particular,* $\lambda_{m/2} > 1 - 2\varepsilon$.

Lemma 6.1. *Suppose there exist vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ such that*

Proof

$1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1.$

$$\mathop{\mathbb{E}}_{ij\sim G} \langle v_i, v_j \rangle \geqslant 1-\varepsilon\,, \quad \mathop{\mathbb{E}}_{i,j\in V} \langle v_i, v_j\rangle^2 \leqslant \tfrac{1}{m}\,, \quad \mathop{\mathbb{E}}_{i\in V} \|v_i\|^2 = 1\,.$$

Then for all $C > 1$, $\lambda_{(1-1/C)m} \geqslant 1 - C\cdot\varepsilon$. In particular, $\lambda_{m/2} > 1-2\varepsilon$.

$X = (X_{rs})$ Gram matrix corresponding to vectors v_i, w.r.t. the eigenbasis of $A(G)$. Then,

$$\mathop{\mathbb{E}}_{(i,j)\in E} \langle v_i, v_j \rangle = \sum_r \lambda_r X_{r,r}\,, \quad \mathop{\mathbb{E}}_{i,j} \langle v_i, v_j\rangle^2 = \sum_{r,s} X_{r,s}^2\,, \quad \mathop{\mathbb{E}}_i \|v_i\|^2 = \sum_r X_{r,r}.$$

$m' :=$ largest index in $[n]$ s.t. $\lambda_{m'} \geq 1 - C\cdot\varepsilon$. Goal: $m' \geq (1-1/C)m$.

$p_r := X_{r,r}$. $\quad q := \sum_{r=1}^{m'} p_r \leq \underbrace{m' \sum_{r=1}^{m'} p_r^2} \leq m' \sum_{r=1}^{n} p_r^2 \leq m' \sum_{r,s} X_{r,s}^2 \leq \frac{m'}{m}.$

?

$$1-\varepsilon \leq \sum_{r=1}^{n} \lambda_r p_r \leq \sum_{r=1}^{m'} p_r + (1-C\cdot\varepsilon)\cdot\sum_{r=m'+1}^{n} p_r$$

$$= q + (1-C\cdot\varepsilon)(1-q) = (C\cdot\varepsilon)q + (1-C\cdot\varepsilon) \leq (C\cdot\varepsilon)\frac{m'}{m} + (1-C\varepsilon)$$

Applying the Lemma

Lemma 6.1. *Suppose there exist vectors* $v_1, \dots, v_n \in \mathbb{R}^n$ *such that*

$$\mathbb{E}_{ij \sim G} \langle v_i, v_j \rangle \geqslant 1 - \varepsilon\,, \qquad \mathbb{E}_{i,j \in V} \langle v_i, v_j \rangle^2 \leqslant \tfrac{1}{m}\,, \qquad \mathbb{E}_{i \in V} \|v_i\|^2 = 1\,.$$

Then for all $C > 1$, $\lambda_{(1-1/C)m} \geqslant 1 - C \cdot \varepsilon$. *In particular,* $\lambda_{m/2} > 1 - 2\varepsilon$.

Lemma 4.1. *Let* $v_1, \dots, v_n$ *be vectors in the unit ball. Suppose that the vectors are correlated across the edges of a regular* n*-vertex graph* G,

$$\mathbb{E}_{ij \sim G} \langle v_i, v_j \rangle \geqslant \rho\,.$$

Then, the global correlation of the vectors is lower bounded by

$$\mathbb{E}_{i,j \in V} |\langle v_i, v_j \rangle| \geqslant \Omega(\rho)/\mathrm{rank}_{\geqslant \Omega(\rho)}(G)\,.$$

where $\mathrm{rank}_{\geqslant \rho}(G)$ *is the number of eigenvalues of adjacency matrix of* G *that are larger than* ρ.

Note that Lemma 4.1 follows directly from the previous lemma by picking $C = \frac{(1-\rho/100)}{(1-\rho)}$ and observing that $\mathbb{E}_{i,j \in V} |\langle v_i, v_j \rangle| \geqslant \mathbb{E}_{i,j \in V} |\langle v_i, v_j \rangle|^2$ since $|\langle v_i, v_j \rangle| \leqslant 1$ for all $i, j \in V$

Putting things together

Theorem 5.6. *Let $X_1, \ldots, X_n$ be r-local random variables and let $X_1', \ldots, X_n'$ be the random variables produced by Algorithm 5.5 on input $X_1, \ldots, X_n$. Suppose that the matrices $(\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S))_{i\in V, a\in[k]}$ are positive semidefinite for every set $S \subseteq V$ with $|S| \leqslant r$ and local assignment $x_S \in [k]^S$. Then, if $r \gg O(k/\varepsilon^4) \cdot \mathrm{rank}_{\Omega(\varepsilon/k)^2}(G)$,*

$$\mathop{\mathbb{E}}_{ij\sim G} \left\| \{X_i X_j\} - \{X_i' X_j'\} \right\|_1 \leqslant \varepsilon\,.$$

- r-local means any r-subset can be jointly sampled
- This implies (additive) integrality gap $\leq \epsilon$ for level-r Lasserre SDP

Theorem 5.6. *Let $X_1, \ldots, X_n$ be r-local random variables and let $X'_1, \ldots, X'_n$ be the random variables produced by Algorithm 5.5 on input $X_1, \ldots, X_n$. Suppose that the matrices $(\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S))_{i \in V, a \in [k]}$ are positive semidefinite for every set $S \subseteq V$ with $|S| \leqslant r$ and local assignment $x_S \in [k]^S$. Then, if $r \gg O(k/\varepsilon^4) \cdot \mathrm{rank}_{\Omega(\varepsilon/k)^2}(G)$,*

$$\mathop{\mathbb{E}}_{ij \sim G} \left\| \{X_i X_j\} - \{X'_i X'_j\} \right\|_1 \leqslant \varepsilon .$$

Proof

- $\varepsilon_m := \mathop{\mathbb{E}}_{S \in \binom{V}{m}} \mathop{\mathbb{E}}_{\{X_S\}} \mathop{\mathbb{E}}_{(i,j) \in E} \| \{X_i X_j \mid X_S\} - \{X_i \mid X_S\}\{X_j \mid X_S\} \|_1 .$

  $\varepsilon_r \leq \varepsilon \quad \Leftarrow \quad \mathop{\mathbb{E}}_{m} \varepsilon_m \leq \varepsilon .$

- $\Phi_m := \mathop{\mathbb{E}}_{S \in \binom{V}{m}} \mathop{\mathbb{E}}_{\{X_S\}} \mathop{\mathbb{E}}_{i \in V} \mathrm{Var}(X_i \mid X_S) .$

- If $\varepsilon_m \geq \varepsilon/2$, Then $\Pr_{\substack{S, \{X_S\} \\ |S| = m}} \left\{ \mathop{\mathbb{E}}_{(i,j) \in E} \| \{X_i X_j \mid X_S\} - \{X_i \mid X_S\}\{X_j \mid X_S\} \|_1 \geq \varepsilon/4 \right\} \geq \varepsilon/4 .$

If discrepancy large, apply lemma 5.4,

$$\mathop{\mathbb{E}}_{S, \{X_S\}} \left[\mathop{\mathbb{E}}_{ij} \left[\mathrm{Var}[X_i \mid X_S] - \mathrm{Var}[X_i \mid X_S, X_j] \right] \right] \geq \frac{\varepsilon}{4} \cdot \frac{\Omega(\varepsilon^2/k)}{\mathrm{rank}_{\geq \Omega(\varepsilon/k)^2}(G)}$$

Proof

Theorem 5.6. *Let $X_1, \ldots, X_n$ be r-local random variables and let $X'_1, \ldots, X'_n$ be the random variables produced by Algorithm 5.5 on input $X_1, \ldots, X_n$. Suppose that the matrices $(\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S))_{i \in V, a \in [k]}$ are positive semidefinite for every set $S \subseteq V$ with $|S| \leqslant r$ and local assignment $x_S \in [k]^S$. Then, if $r \gg O(k/\varepsilon^4) \cdot \mathrm{rank}_{\Omega(\varepsilon/k)^2}(G)$,*

$$\mathop{\mathbb{E}}_{ij \sim G} \left\| \{X_i X_j\} - \{X'_i X'_j\} \right\|_1 \leqslant \varepsilon .$$

$$\Rightarrow \Phi_m - \Phi_{m+1} \geq \frac{\varepsilon^3/k}{\mathrm{rank}_{\geq \Omega(\varepsilon/k)^2}(G)} . \quad \left(\text{If } \varepsilon_m \geq \varepsilon/2\right)$$

$\Rightarrow$ At most $\cancel{\left[\frac{1}{(\varepsilon^3/k)/\mathrm{rank}_{\geq \Omega(\varepsilon/k)^2}(G)}\right]}$ indices m s.t. $\varepsilon_m \geq \varepsilon/2$.

$\Rightarrow$ If $r \geq \frac{1}{(\varepsilon/2)} \cdot [\quad]$, then $\mathop{\mathbb{E}}_{m} \varepsilon_m \leq \varepsilon$.

$\Rightarrow$ ~~need~~ suffices for $r \gtrsim \frac{\mathrm{rank}_{\geq \Omega(\varepsilon/k)^2}(G) \cdot k}{\varepsilon^4}$.

Recap

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

Sparsest Cut

- Given graph $G = (V, E)$, define sparsest cut (aka edge expansion) as

$$\phi(G) = \min_{S \subseteq V} \frac{|E(S, S^c)|}{|S| \cdot |S^c|}$$

Setup

Integer program:

$$\min_{x} \sum_{i\sim j} (x_i - x_j)^2$$

$$\text{s.t.} \quad \sum_i x_i = \mu$$

$$x_i \in \{0,1\}.$$

Look at $n/2$ programs for $\mu = 1, 2, \ldots, n/2$.

Dealing with cardinality constraint

- $\sum y_i = \mu$, but also because of $M^{(\ell)}(y) \succeq 0$, all projected solutions will satisfy the linear constraints.
- Independent sampling for the remaining variables.
- By Hoeffding, $\Pr[|\sum X_i - \mu| \text{ small}]$ is large.

  $\therefore$ With good probability, rounded soln will have good # of cut edges and cardinality close to μ.

Conclusion

Theorem 5.6. *Let $X_1, \ldots, X_n$ be r-local random variables and let $X'_1, \ldots, X'_n$ be the random variables produced by Algorithm 5.5 on input $X_1, \ldots, X_n$. Suppose that the matrices $(\mathrm{Cov}(X_{ia}, X_{jb} \mid X_S = x_S))_{i \in V, a \in [k]}$ are positive semidefinite for every set $S \subseteq V$ with $|S| \leqslant r$ and local assignment $x_S \in [k]^S$. Then, if $r \gg \underline{O(k/\varepsilon^4)} \cdot \underline{\mathrm{rank}_{\Omega(\varepsilon/k)^2}(G)}$,*

$$\mathop{\mathbb{E}}_{ij \sim G} \left\| \{X_i X_j\} - \{X'_i X'_j\} \right\|_1 \leqslant \varepsilon\,.$$

For sparsest cut, by taking $r \gtrsim \dfrac{\mathrm{rank}_{\Omega(\varepsilon^2)}(G)}{\varepsilon^4}$,

we can get constant factor approximation.

Agenda

- Recall
- More on local distribution
- Correlation rounding
- Sparsest cut *for low-rank graphs*
- Technical discussions

Steurer's conjecture

Conjecture 9.2. *For every $\varepsilon > 0$, there exists positive constants $\eta = \eta(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that the following holds: For every collection of unit vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ with $\mathbb{E}_{i,j\in[n]} \left|\langle v_i, v_j \rangle\right| \leqslant n^{-\varepsilon}$, there exists two sets $S, T \subseteq \{1, \ldots, n\}$ with $|S|, |T| \geqslant \delta n$ and $\|v_i - v_j\|^2 \geqslant \eta$ for all $i \in S$ and $j \in T$.*

In words: if a set of vectors have very low *global correlation*, there are two large cores of constant distance from each other

Sparsest cut for general graphs

- [ABS] proved that SSE can be solved in subexponential time
- High threshold rank graph is the "easy case" and dealt with using random walks

- We cannot use local-to-global correlation when $rank(G)$ is high
- If we follow [BRS], want to do something simple when global correlation (of what?) is low

- There are several possibilities how this might proceed…

Attempt 1: look at v_i

- Corresponds to good embedding with $\sum_{(i,j)\in E}\|v_i - v_j\|^2 \le O(OBJ)$ and $\sum_i \|v_i\|^2 = \mu$
- One way to map them to unit vectors, preserving objective:

$$w_i := \binom{v_i}{v_\emptyset - v_i} = \binom{v_i(1)}{v_i(0)}$$

- Not sure how to use $\mathbb{E}_{i,j}\langle w_i, w_j\rangle$ to lower bound variance drop
  - If $\mathbb{E}_{i,j}\langle w_i, w_j\rangle$ is large, then variance drop is large ← ?
  - If $\mathbb{E}_{i,j}\langle w_i, w_j\rangle$ is small, then can apply Steurer.

Attempt 2: look at $\tilde{u}_i$

- Can lower bound variance drop using $\mathbb{E}_{i,j}\langle \tilde{u}_i, \tilde{u}_j \rangle$

$$\sum_{i\sim j} \|\tilde{u}_i - \tilde{u}_j\|^2 \lesssim \sum_{i\sim j} \|v_i - v_j\|^2$$

- Not sure how to map to unit vectors while preserving objective
- (u_i is akin to projecting to $\mathrm{span}(v_\emptyset)^\perp$)

Comments

- I haven't figured out yet how Steurer's conjecture implies SUBEXP sparsest cut

- But we don't need to limit ourselves to the exact statement!
- Based on the proof flow today, can identify what needs to be true for Lasserre-based sparsest cut algorithm to work in SUBEXP

Summary

In this two-part talk, we have:

- Introduced the Lasserre hierarchy
  - Interplay between vector solutions v_I and probabilities y_I
  - Projecting Lasserre solutions via conditioning
  - Local distributions $y_I(f)$ from y_J
- Gone through [BRS] propagation rounding
  - Role of threshold rank of constraint graph
  - After conditioning, independent rounding $\approx$ correlated rounding
- Discussed application of Lasserre to sparsest cut

Discussions and idle thoughts

- Geometric picture of projection
- Additional properties from structure of vector solution?
- Using entropy instead of variance
- CSP’s on hypergraphs
- Can use SDP hierarchy to improve experimental design?

(Almost) blank page

References

- David Steurer, “On the Complexity of Unique Games and Graph Expansion”.
- Thomas Rothvoss, “The Lasserre hierarchy in Approximation algorithms”.
- Lap Chi Lau, “Lasserre Hierarchy”, Lecture 13 of CSCI5060.
- V. Guruswami and A.K. Sinop, “Lasserre Hierarchy, Higher Eigenvalues, and Approximation Schemes for Quadratic Integer Programming with PSD Objectives”.
- B. Barak, P. Raghavendra, and D. Steurer. “Rounding Semidefinite Programming Hierarchies via Global Correlation”.