The Arora Rao Vazirani algorithm for Sparsest Cut

<u>Recall:</u> $G(V,E)$: undirected graph on n vertices, m edges.

$\forall S \subseteq V$, $\Phi(S) = \dfrac{|E(S,\bar{S})|}{|S|\cdot|\bar{S}|}$

$$\Phi(G) = \min_{S \subseteq V} \Phi(S)$$

↳ Sparsity of the cut

<u>Quadratic program</u>

$$\min \frac{1}{4} \sum_{\{i,j\} \in E} (x_i - x_j)^2$$

guess.

$$\text{s.t. } \frac{1}{4} \sum_{1 \le i,j \le n} (x_i - x_j)^2 = \beta.$$

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

For illustration today, we will assume that

$$\beta = 2|S|(n-|S|) = 0.1 n^2$$

This is equivalent to assuming that the optimal set is roughly balanced, i.e. $|S| = n - |S| = \Omega(n)$.

This also happens to be the hardest case & the general case can be reduced to it.

Vector Program:

$$\min \frac{1}{4} \sum_{\{i,j\} \in E} \|V_i - V_j\|_2^2$$

$$\|V_i\|_2^2 = 1 \quad \forall i$$

$$\frac{1}{4} \sum_{1 \le i,j \le n} \|V_i - V_j\|_2^2 = \cancel{\beta} \cdot 0.1 \, n^2$$

$\forall i,j,k$: $\|V_i - V_k\|_2^2 \le \|V_i - V_j\|_2^2 + \|V_j - V_k\|_2^2$

①

SQUARED TRIANGLE INEQUALITY.

CAUTION: Arbitrary vectors V_i, V_j, V_k satisfy the triangle inequality:

$$\|V_i - V_k\|_2 \le \|V_i - V_j\|_2 + \|V_j - V_k\|_2$$

This does not imply the squared triangle inequality.

Why is this SDP feasible?

Let $S_* \subseteq V$ be the sparsest cut such that $2|S_*|\cdot(n-|S_*|) = \beta$.

Let $$v_i = \begin{cases} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \text{if } i \in S_* \\ \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \text{if } i \notin S_*. \end{cases}$$

Then $$\frac{1}{4}\sum_{1\leq i,j\leq n} \|v_i - v_j\|_2^2 = 2\cdot|S_*|(n-|S_*|) = \beta.$$

and $\frac{1}{4}\sum_{\{i,j\}\in E} \|V_i - V_j\|_2^2$

$$= |E(S_*, \bar{S}_*)|.$$

$$= \Phi(G)\cdot\beta.$$

$\|V_i\|_2^2 = 1 \ \forall i$ ✓

What about squared triangle inequality?

Obs: for any $a, b, c \in \{\pm 1\}$,

$$(a-b)^2 \leq (a-c)^2 + (c-b)^2$$

trivial if $a = b$.

if not one of $a \neq c$ or $b \neq c$. $\square$.

So the vectors also satisfy the squared triangle inequality.

MAIN THEOREM

<u>Thm:</u> Let G be a graph that admits a sparsest cut of size $2|S_*|(n-|S_*|)=\beta$. Let $v_1 \ldots v_n$ be unit vectors such that

$$\frac{1}{4}\sum_{\{i,j\}\in E} \|v_i - v_j\|_2^2 = C \quad \text{and}$$

satisfying the constraint system ①

Then, there's a randomized rounding algorithm that outputs a cut S s.t.

$$\frac{|E(S,\bar{S})|}{|S|\cdot(n-|S|)} \lesssim O(\sqrt{\log n})\cdot\left(\frac{C}{\beta}\right).$$

HIGH LEVEL OVERVIEW

Def: For each $i, j \leq n$, let

$$d(i,j) = \frac{1}{4}\|v_i - v_j\|_2^2$$

Then, $d(i,j) \leq d(i,k) + d(k,j) \ \forall i,j,k$.

Thus, d is a "metric".

In fact, it's a "squared Euclidean" metric ~ a special kind of metric often also called "metric of negative type".

So, one can view SDP a refinement of Leighton Rao LP that corresponds to d being an arbitrary metric.

How might you want to round?

Suppose you try to do "GW-style" rounding.

Let $g \sim N(0,1)^n$: std. gaussian vector.

Let $x_i = \text{sign}(\langle g, v_i \rangle)$. $\forall i$

<u>Lemma</u> [Gaussian Rounding]

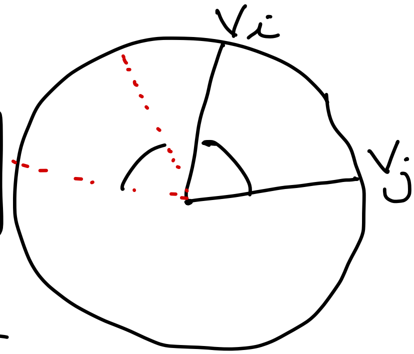
$$\Pr[\,\text{sign}(\langle v_i, g\rangle) \neq \text{sign}(\langle v_j, g\rangle)\,] = \Theta\left(\sqrt{d(i,j)}\right).$$

<u>Proof:</u> familiar 2D analysis.

$\frac{1}{4}\|v_i - v_j\|_2^2 = \delta$ (say).

Then $\frac{1}{2} - \frac{1}{2}\langle v_i, v_j\rangle = \delta$

or $\langle v_i, v_j\rangle = 1 - 2\cdot\delta$.



Thus

$$\Pr[\mathrm{sign}(\langle g, v_i\rangle) \neq \mathrm{sign}(\langle g, v_j\rangle)]$$

$$= \frac{\text{arc-cos}(1-2\delta)}{\pi}$$

$$\sim \frac{\sqrt{2\delta}}{\pi} \sim \Theta(\|v_i - v_j\|_2).$$

$$= \Theta\left(\sqrt{d(i,j)}\right).$$

$d(i,j)$ is a # betⁿ 0 & 1

and can be, say $\frac{1}{\log^{10} n}$.

then $\sqrt{d(i,j)}^{-1} = \frac{1}{\log^5 n} \gg (\sqrt{\log n})\, \frac{1}{\log^{10} n}$

So, GW rounding separates short edges (i.e. $d(i,j) \sim \frac{1}{\text{polylog } n}$) with too large a probability.

Idea 2 : (Region Growing)

Def : For any $A \subseteq V$, & $i \in V$

$$d(i,A) = \min_{j \in A} d(i,j)$$

"distance of i from nearest point in A".

Then, triangle inequality for d implies that

$$d(i,A) \leq d(i,j) + d(j,A).$$

<u>Lemma (Region Growing)</u>

Let $A \subseteq V$ be any set of vertices.

Choose $r \in [0,1]$ uniformly at random and set

$$S = \{k \mid d(k,A) \leq r\}$$

Then, for every i,j,

$$\Pr[X_i \neq X_j] \leq d(i,j).$$

Proof: Suppose $d(i,A) \geq d(j,A)$.

$$d(i,A) - d(j,A) \leq d(i,j)$$

Triangle Inequality

So $\Pr[X_i \neq X_j]$

$$= \Pr[d(i,A) \geq r \geq d(j,A)]$$

$$= |d(i,A) - d(j,A)|$$

$$\leq d(i,j).$$

So
$$\mathbb{E} \sum_{\{i,j\} \in E} \mathbb{1}(X_i \neq X_j)$$

$$\leq \sum_{\{i,j\} \in E} d(i,j).$$

$$\leq C \longrightarrow \text{SDP opt!}$$

Why aren't we done?

For all we know, $S = V$!

But we need $2|S|(n-|S|) > \frac{\beta}{O(\sqrt{\log n})}$

Why would S be too large (i.e. the cut be too imbalanced?) because somehow most vertices i ended up near the set A (i.e. $d(i,A)$ was small) and thus were put in S.

In order to avoid this; we will need that a significant fraction of vertices be "separated" or far from the set A we choose in region growing.

Key ARV Lemma: Such a set A exists.

Let's formalize this.

Theorem [ARV Structure Theorem] [Arora-Rao-Vazirani '04, Lee '05]

Let $v_1, \ldots, v_n$ be unit vectors. Let $d(i,j) = \frac{1}{4}\|v_i - v_j\|_2^2$ satisfy triangle inequality. Further

Suppose $\sum_{\{i,j\} \in [n]} d(i,j) \geq 0.1 n^2$. "balanced case".

Then, there exist sets $A, B \subseteq V$

s.t. ① $|A|, |B| = \Omega(n)$.

② $\min_{i \in A, j \in B} d(i,j) \geq \Delta$ for $\Delta = \Theta\left(\frac{1}{\sqrt{\log n}}\right)$

Further, can find A, B as above in polynomial time.

<u>Observe</u>: Immediately done if structure theorem is true!

Use A to "region grow"

Then, expected $|E(S,\bar{S})| \leq C$.

while $\mathbb{E}\,|S|(n-|S|) = \Omega\left(\frac{n^2}{\sqrt{\log n}}\right)$.

$$= \frac{\beta}{\alpha(\sqrt{\log n})}$$

<u>So</u>

$$\frac{\mathbb{E}\,|E(S,\bar{S})|}{\mathbb{E}\,|S|\cdot(n-|S|)} \lesssim O(\sqrt{\log n})\ \text{SDP-OPT}$$

$$\lesssim O(\sqrt{\log n})\ \Phi(G).$$

$\square$

Will now focus on proving structure theorem.

OBSERVE : Structure theorem has NOTHING to do with graph. It's a statement about existence of "well separated" sets in a "ℓ_2-squared" metric space where avg. distances are large.

Let's now proceed to proving struct thm.

Let's now define a <u>directed</u> graph H in terms of y_i's.

H has vertices $i \in [n]$.

$E(H) = \{(i,j) \in [n]^2 \mid d(i,j) \leq \Delta\}$

<u>Def (Vertex Separator)</u>

$A, B \subseteq [n]$, (A,B) is a vertex separator in H if no edge of H goes from A to B

That is, $E(H) \cap A \times B = \emptyset$.

We say that a vertex separator (A, B) is good if $|A| \cdot |B| = \Omega(n^2)$.

Then struct theorem is same as saying that H has a good vertex separator for $\Delta = \theta\left(\frac{1}{\sqrt{\log n}}\right)$

Randomized Algorithm to find vertex separators in H.

Let $y_i = \langle g, v_i \rangle \quad \forall i$

Then, $\mathbb{E} y_i^2 = 1$, $d(i,j) = \frac{1}{4} \mathbb{E}(y_i - y_j)^2$

1. Choose subsets A^0, B^0 s.t.

   $A^0 = \{ i \mid y_i \leq -1 \} \quad B^0 = \{ j \mid y_j \geq 1 \}$.

2. Find maximal matching of $E(H) \cap A^0 \times B^0$ by greedily processing edges $[n] \times [n]$ in alphabetical order.

3. Output $A = A^0 \setminus V(M)$, $B = B^0 \setminus V(M)$

Note: 1) A^0, B^0, A, B are all random variables as they are functions of the gaussian vector g.

2) edges of matching M are directed.

In particular,

$$(i,j) \in E(M) \Rightarrow y_j - y_i \geq 2.$$

Prop: (A, B) form a vertex sep in H.

Proof: Consider any edge in $E(H) \cap A^0 \times B^0$. If it exists in $E(H) \cap A \times B$, can remove it, add it to M and thus M is not maximal.

Obs: $M(y)$ = reverse of $M(-y)$.

processing order of edges in greedy algo is independent of y.

& $\Pr[M] = \Pr[\mathrm{rev}(M)]$

Lemma (Vertex Sep or Large Matching)

There's a const. c' s.t. if

$\mathbb{E}|A||B| < c' \cdot n^2$ then $\mathbb{E}|M| > c'n$.

Proof: Fix any i, j, then, absolute constant

$\Pr[y_i \le -1, y_j \ge 1] \geqslant c \cdot d(i,j)$.

(Similar proof as GW analysis. Exercise!)

Thus, $\mathbb{E}\,|A^0|\cdot|B^0| \geq 0.1\cdot c\cdot n^2$

$$|A|\cdot|B| \geq |A^0|\cdot|B^0| - n\cdot|M|$$

$$\mathbb{E}\,|A|\,|B| \geq \underset{\overset{\parallel}{0.1\cdot c\cdot n^2}}{\mathbb{E}\,|A^0|\,|B^0|} - n\cdot\mathbb{E}\,|M|$$

So if $\mathbb{E}\,|A|\cdot|B| < \overbrace{0.05c}^{c'}\cdot n^2$

then $\mathbb{E}\,|M| \geq 0.05\,c\cdot n.$

$\square$

We will analyze $\mathbb{E}\,|M|$ and prove it is small.

Lemma: (Main Tech Lemma)

For some abs. const. $c > 0$;

$$\frac{c}{\Delta}\left(\frac{\mathbb{E}|M|}{n}\right)^3 \lesssim \mathbb{E} \max_{i,j} y_j - y_i$$

$$\lesssim \sqrt{2\log n}.$$

In words, we will analyze the Expectation of maximum of

$$\{ y_j - y_i \mid 1 \le i, j \le n \}$$

By standard results this is atmost $\sqrt{2\log n}$

OTOH, we'll show that if $\exists$ a large matching then expected maximum of $y_j - y_i$ is large.

Rearrange:

$$\mathbb{E}|M| \le n \cdot \left(\frac{\sqrt{2 \log n} \cdot \Delta}{c} \right)^{1/3}$$

If $\Delta > O(\sqrt{\log n})$,

RHS $< c'n$. So done!

Fact (Proof on Piazza)

Suppose $z_1 \dots z_t$ are jointly gaussian with $\mathbb{E} z_i = 0$ & $\mathbb{E} z_i^2 \le 1$ for all i.

Then $\mathbb{E} \max_{i \le t} z_i \le 2\sqrt{\log 2t}$

Apply this to $\{y_j - y_i\}$ to get

RHS -

$\max_{i,j} y_j - y_i$: max of a bunch of jointly gaussian random variables

"gaussian process".

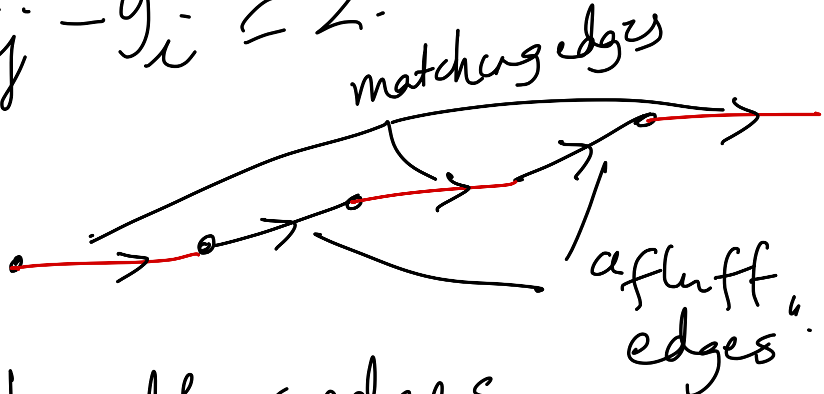
Proof of Main Tech Lemma

Idea of the proof:

"large matching implies that there's an j, i s.t. $y_j - y_i$" is large.

<u>why?</u> matching edge $i \to j$

$\Rightarrow \quad y_j - y_i \geq 2.$



Chain together edges and show that we can form long paths.

Then end points of the path must be very far!

we'll run this argument "in average".

Def

* $H^k(i)$ = Set of vertices reachable from i from at most k-steps in H.

* $Z_i^{(k)} = \max_{j \in H^k(i)} y_j - y_i$.

* $\Psi(k) = \sum_{i=1}^{n} \mathbb{E} Z_i^{(k)}$.

Then, notice that

$$\frac{\Psi(k)}{n} \leq \mathbb{E} \max_{i,j} y_j - y_i$$

$\boxed{\text{Goal: Prove } \Psi(k) \text{ is large if } \mathbb{E}|M| \text{ is large}}$

Lemma (Chaining lemma)

For some $C \geq 0$,

$$\Psi(k+1) \geq \Psi(k) + \overbrace{4 \cdot \mathbb{E}|M|}^{\text{large if } M \text{ is large on avg.}}$$

$$\underbrace{- \; C \cdot n \cdot \max_{\substack{i \in [n] \\ j \in H^{k+1}(i)}} \left(\mathbb{E}(y_j - y_i)^2 \right)^{1/2}}_{\text{"error/noise/variance term"}}$$

Proof of Main Tech Lemma assuming Chaining Lemma

$$\mathbb{E}(y_i - y_j)^2 = d(i,j)$$

$$\leq k \cdot \Delta \quad \text{Squared Triangle Inequality}$$

So $\Psi(k+1) \geq \Psi(k) + \mathbb{E}|M| - Cn\sqrt{k} \cdot \sqrt{\Delta}$

So for every $k \leq k_0 = \dfrac{\frac{1}{4C^2}\left(\frac{\mathbb{E}|M|}{n}\right)^2}{\Delta}$

$$\Psi(k+1) \geq \Psi(k) + \left(\frac{\mathbb{E}|M|}{2}\right).$$

<u>Thus</u>, $\dfrac{\Psi(k_0)}{n} \geq \dfrac{1}{2} \cdot k_0 \cdot \left(\dfrac{\mathbb{E}|M|}{n}\right)$

So $\dfrac{\Psi(k_0)}{n} \geq \dfrac{1}{8C^2} \cdot \left(\dfrac{\mathbb{E}|M|}{n}\right)^3 \dfrac{1}{\Delta}$

/

Fact [Borell]: $Z_1 \ldots Z_t$ mean 0, jointly gaussian.

$$\mathrm{Var}[\max Z_1 \ldots Z_t] \lesssim O(1) \cdot \max\{\mathrm{Var}(Z_1), \ldots, \mathrm{Var}(Z_t)\}.$$

Borell's Inequality, can be proved via "Lipschitz concentration" / Sudakov Tsirelson '74

Proof of Chaining Lemma [uses Borell]

$$Z_i^{k+1} \geq Z_j^k + X_j - X_i \quad (1)$$

for every $(i,j) \in E(H)$.

This is because for any $r \in H_j^{(k+1)}$ there's a $\leq k+1$ length path from i.

"Fluff edges"

Let $N \subseteq [n] \times [n]$: arbitrary matching of vertices not in M. Thus,

$\underline{\forall (i,j) \in M}$: $\quad z_i^{k+1} \geq z_j + 2$

$\underline{\forall (i,j) \in N}$: $\quad \frac{1}{2} z_i^{k+1} + \frac{1}{2} z_j^{k+1}$

$$\geq \frac{1}{2} z_i^{k} + \frac{1}{2} z_j^{k}.$$

$$\underbrace{\hphantom{\forall (i,j) \in N : \frac{1}{2} z_i^{k+1} + \frac{1}{2} z_j^{k+1}}}$$

total of $\frac{n}{2}$ inequalities.

Add the inequalities, take exp.

$$\mathbb{E}\sum_{i=1}^{n} Z_i^{k+1}\cdot L_i \geq \mathbb{E}\sum_{j=1}^{n} Z_j^{k}\cdot R_j + 2|M|$$

where

would like $\mathbb{E} Z_i^{k+1}$ & $\mathbb{E} Z_j^{k}$. without L_i's.

if yes, then done.

$$L_i = \begin{cases} 1 & \text{if } i \text{ has an outgoing edge in } M \\ \frac{1}{2} & \text{if } i \text{ is not matched in } M \\ 0 & \text{if } i \text{ has an incoming edge in } M. \end{cases}$$

$$R_i = \begin{cases} 1 & \text{if } i \text{ has an incoming edge in } M \\ \frac{1}{2} & \text{if } i \text{ is not matched in } M \\ 0 & \text{if } i \text{ has an outgoing edge in } M. \end{cases}$$

Recall obs: $\Pr[i \text{ has an incoming edge in } M] = \Pr[i \text{ has an outgoing edge in } M]$

Thus $\mathbb{E} L_i = \frac{1}{2}$

$\mathbb{E} R_i = \frac{1}{2}$.

error incurred if we replace $\mathbb{E} Z_i^{k+1} L_i$ by $\mathbb{E} Z_i^{k+1} \cdot \mathbb{E} L_i$

$$\left| \mathbb{E} Z_i^{k+1} \cdot L_i - \mathbb{E} Z_i^{k+1} \cdot \mathbb{E} L_i \right|$$

$$= \left| \mathbb{E} \left| Z_i^{k+1} - \mathbb{E} Z_i^{k+1} \right| \cdot \left| L_i - \mathbb{E} L_i \right| \right|$$

$$\leq \sqrt{\mathbb{E} \left| \underbrace{Z_i^{k+1} - \mathbb{E} Z_i^{k+1}}_{} \right|^2} \cdot \sqrt{\mathbb{E} \left| L_i - \mathbb{E} L_i \right|^2}$$

Borell's inequality

$$O(1) \cdot \max_{j \in H^k(i)} \sqrt{\mathbb{E}(y_j - y_i)^2}$$

$\sqrt{\mathbb{E} |L_i - \mathbb{E} L_i|^2} \leq 1$. since L_i is in $[0,1]$

Similarly,

$$\left|\mathbb{E} Z_j^k \cdot R_j - \mathbb{E} Z_j^k \cdot \mathbb{E} R_j\right|$$

$$\lesssim O(1) \cdot \max_{j \in H^k(i)} \sqrt{\mathbb{E}(y_j - y_i)^2}$$

Thus,

$$\underbrace{\sum_{i=1}^n \mathbb{E} Z_i^{k+1}}_{\Psi(k+1)} \geq \overbrace{\sum_{j=1}^n \mathbb{E} Z_i^k}^{\Psi(k)}$$

$$+ 4 \cdot |M|$$

$$- O(n) \cdot \max_{j \in H^{k+1}(i)} \sqrt{\mathbb{E}(y_j - y_i)^2}$$

(error incurred)

$\square$