Reading Group (Spring 2022)

Week 1 – ARV (Part 1)

Alex Tung

17 May 2022

Agenda

- Sparsest cut background
- ARV outline
- Region growing argument
- Structure theorem

Agenda

- Sparsest cut background
- ARV outline
- Region growing argument
- Structure theorem

Sparsest cut background

Task: Given graph $G = (V, E)$, find a cut $S \subseteq V$ such that

$$\phi(S) := \frac{|E(S, S^c)|}{|S| \cdot |S^c|}$$

$\Theta(n)$

is minimized.

$\min(|S|, |S^c|)$

- Spectral: $O(\frac{1}{\varphi})$-approximation* of conductance
- Leighton-Rao: $O(\log n)$-approximation of expansion
- Hardness of sparsest cut: $O(1)$-approximation is UGC-hard

* φ is conductance here

ARV

- Based on SDP $O(\sqrt{\log n})$
- Best of both worlds:

Spectral
(Geometric embedding)

LP
(Metric)

- Key ingredient: l_2^2-triangle inequality

Agenda

- Sparsest cut background
- ARV outline
- Region growing argument
- Structure theorem

Outline of ARV cut-finding procedure

- Step 1: solve the following SDP relaxation of sparsest cut

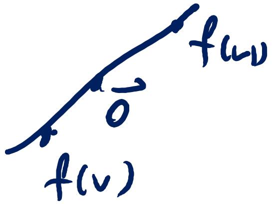
$$SDP(G) := \min_{f:V\to\mathbb{R}^n} \sum_{(u,v)\in E} \|f(u)-f(v)\|^2$$

$$\text{s.t.} \quad \forall u,v,w \quad \|f(u)-f(v)\|^2 + \|f(v)-f(w)\|^2 \geq \|f(u)-f(w)\|^2$$

$$\sum_{u,v\in V} \|f(u)-f(v)\|^2 = n^2$$

$d(u,v) := \|f(u)-f(v)\|^2$

$S \subseteq V$, $\quad f(u) := c_1 \vec{1}$ for $u \in S$, $\quad f(v) := -c_2 \cdot \vec{1}$ for $v \in S^c$



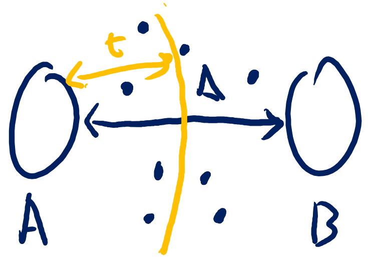
- $\|f(u)-f(v)\|^2 + \|f(v)-f(w)\|^2 \geq \|f(u)-f(w)\|^2$ is the (l_2^2-)triangle inequality
- $\sum_{u,v\in V} \|f(u)-f(v)\|^2$ is normalization term

- Step 2: find two large, well-separated subsets $A, B \subseteq V$

Theorem 1 (ARV Structure Theorem). Given solution $\{f(u)\}_{u \in V}$ to $SDP(G)$. Let $d(u,v) \coloneqq \|f(u) - f(v)\|^2$. For some $\Delta = \Theta\left(\frac{1}{\sqrt{\log n}}\right)$, we can find subsets $A, B \subseteq V$, such that:

- $|A|, |B| \geq \Omega(n)$
- $d(A,B) \coloneqq \min_{u \in A, v \in B} d(u,v) \geq \Delta$

- Step 3: find a sparse cut from threshold cuts based on $d(u, A)$



<u>Theorem 2</u> ("Region growing" argument). Given two sets $A, B \subseteq V$ such that $|A|, |B| \geq \Omega(n)$, $d(A, B) \geq \Delta$ and $W \coloneqq \sum_{(u,v)\in E} \|f(u) - f(v)\|^2$, there exists some $t \in (0, \Delta]$ such that the set $S_t \coloneqq \{u \in V: d(u, A) < t\}$ has expansion $O(W/\Delta)$.

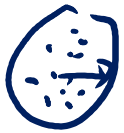
$\Delta = \Theta(1/\sqrt{\log n})$

approx. ratio is $O(1/\Delta)$

$= O(\sqrt{\log n})$

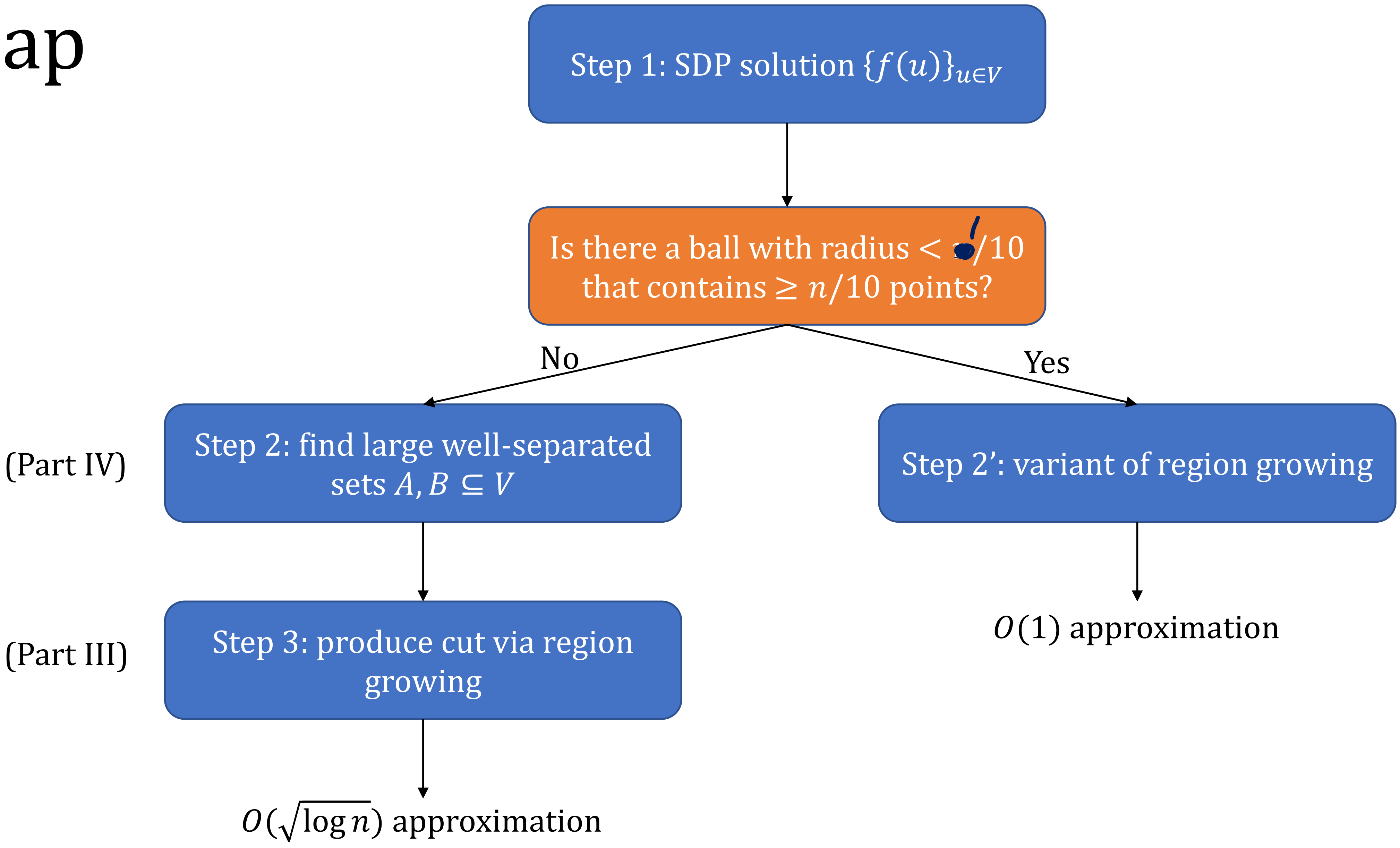
Special case…

- We can only find well-separated $A, B \subseteq V$ if $\{f(u)\}_{u \in V} \subseteq \mathbb{R}^n$ is "well-spread", i.e. no large cluster



- If there is a large cluster, use a variant of region growing
- If there is no large cluster, proceed as planned

Recap

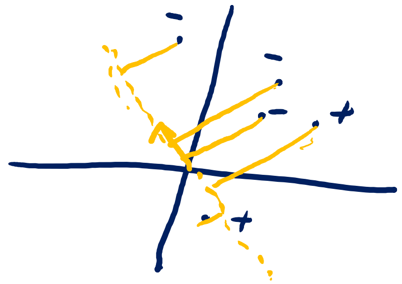

Step 1: SDP solution $\{f(u)\}_{u\in V}$
Is there a ball with radius $< \cancel{n}/10$ that contains $\geq n/10$ points?
No
Yes
(Part IV)
Step 2: find large well-separated sets $A, B \subseteq V$
Step 2': variant of region growing
(Part III)
Step 3: produce cut via region growing
$O(1)$ approximation
$O(\sqrt{\log n})$ approximation

Agenda

- Sparsest cut background
- ARV outline
- Region growing argument
- Structure theorem

Motivation

$\{f(u)\}$

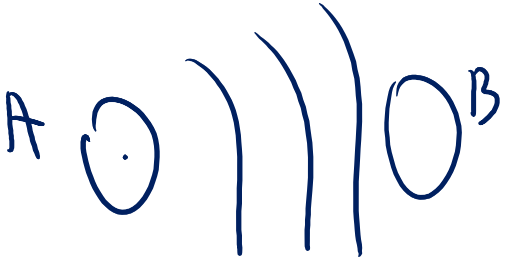


$\underbrace{\|f(u) - f(v)\|}$

$\|f(u) - f(v)\|^2$

- "Embedding" cut finding:
  - $\{f(u)\}_{u \in V} \subseteq \mathbb{R}^n$, project along random direction, then divide at 0
  - + ensures size of cut set
  - - edge cut probability suffers Cauchy-Schwarz loss
- "Metric" cut finding:
  - Find $u \in V$ and $r \sim [0,1]$, return $\{v \in V : d(u,v) \leq r\}$
  - + edge cut probability bounded by $d(u,v) = \|f(u) - f(v)\|^2$
  - - no control of cut set size

- ARV solution: "Metric" with control of cut set size



$|S_t| \geq \Omega(n)$, $|S_t^c| \geq \Omega(n)$

Proof of Theorem 2

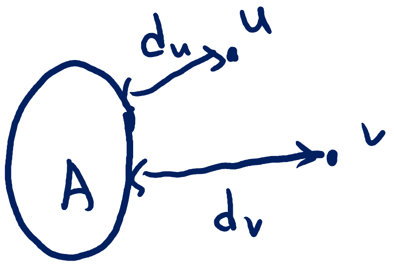
<u>Theorem 2</u> ("Region growing" argument). Given two sets $A, B \subseteq V$ such that $|A|, |B| \geq \Omega(n)$, $d(A, B) \geq \Delta$ and $W := \sum_{(u,v)\in E} \|f(u) - f(v)\|^2$, there exists some $t \in (0, \Delta]$ such that the set $S_t := \{u \in V : d(u, A) < t\}$ has expansion $O(W/\Delta)$.

- $\phi(S_t) = \dfrac{|E(S_t, S_t^c)|}{|S_t| \cdot |S_t^c|}$
- $|S_t| \geq \Omega(n)$, $|S_t^c| \geq \Omega(n) \Rightarrow$ Denominator $\geq \Omega(n^2)$.
- Want: upper bound $\mathbb{E}_t |E(S_t, S_t^c)|$.

Blank Page

$$\mathbb{E}_t\, |E(S_t, S_t^c)| = \sum_{(u,v)\in E} \Pr_t\left[(u,v) \text{ is cut}\right]$$

$$= \sum_{(u,v)\in E} \Pr_{t\sim[0,\Delta]}\left[d_u < t \le d_v\right]$$



$$= \sum_{(u,v)\in E} \frac{|d_v - d_u|}{\Delta}$$

(assume $d_u < d_v$)

$$\le \sum_{(u,v)\in E} \frac{d(u,v)}{\Delta}$$

(u,v) cut $\Leftrightarrow d_u < t \le d_v$.

$$= W/\Delta.$$

Agenda

- Sparsest cut background
- ARV outline
- Region growing argument
- Structure theorem

Theorem 1 restated

Theorem 1 (ARV Structure Theorem). Given well-spread vectors $\{f(u)\}_{u \in V} \subseteq \mathbb{R}^m$. Let $d(u,v) \coloneqq \|f(u) - f(v)\|^2$. Suppose $\{f(u)\}_{u \in V}$ satisfies the l_2^2-triangle inequality. For some $\Delta = \Theta\left(1/\sqrt{\log n}\right)$, we can find subsets $A, B \subseteq V$, such that:

- $|A|, |B| \geq \Omega(n)$
- $d(A,B) \coloneqq \min\limits_{u \in A, v \in B} d(u,v) \geq \Delta$

Succeeds with probability $\geqslant c$.

- The result does not depend on dimension of the $f(u)$'s
- It has nothing to do with cut-finding, nor with the graph G!

ARV set-finding algorithm

Input: SDP solution f(u)

Step 1: project $f(u)$ along "random" direction w: $h(u) \coloneqq \langle f(u), w \rangle$

Step 2: for certain thresholds $a < b$, set $A \coloneqq \{u \in V : h(u) \leq a\}$ and $B \coloneqq \{u \in V : h(u) \geq b\}$

FAIL 1 if $|A| < cn$ or $|B| < cn$

Step 3: find any M: maximal matching on edge set $\{(u, v) \in A \times B : d(u, v) < \Delta\}$

FAIL 2 if $|M| > c'n$

Output: $A' = A \setminus V(M), B' = B \setminus \mathrm{V}(M)$

The plan

- The goal is to show that the algorithm succeeds with $\Omega(1)$ probability

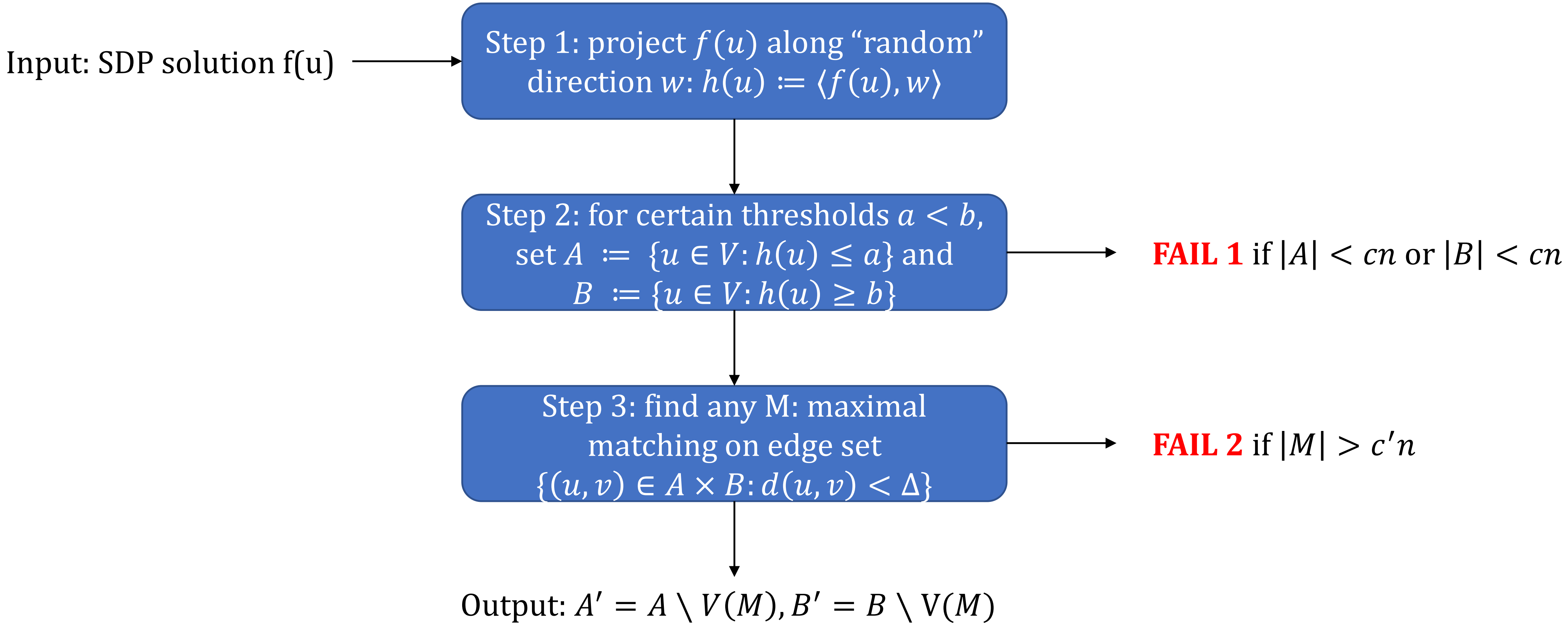
- First, we show that FAIL 1 happens with probability $\leq 1 - c_1$
  - need well-spread assumption here
  - the constant in "$|A|, |B| \geq \Omega(n)$" depends on c_1
- Then, we show that FAIL 2 happens with probability $\leq c_2$
  - $c_2 > 0$ can be made arbitrarily small
  - the constant in "$\Delta = \Theta(1/\sqrt{\log n})$" depends on c_2
- Therefore, success probability is $\geq c_1 - c_2$

Point of divergence…

- There are (at least) three known proofs of ARV:
  - ARV original
  - Rothvoss
  - Barak & Steurer

- We will follow Rothvoss's proof

Rothvoss version of set-finding



Rothvoss version of set-finding

Input: SDP solution f(u) →

Step 1: project $f(u)$ along random Gaussian $g \sim N(0,1)^m$. Let $h(u) := \langle f(u), g \rangle$

↓

Step 2: set $A := \{u \in V: h(u) \leq -1\}$ and $B := \{u \in V: h(u) \geq +1\}$ → **FAIL 1** if $|A| < cn$ or $|B| < cn$

↓

Step 3: find any M: maximal matching on edge set $\{(u,v) \in A \times B: d(u,v) < \Delta\}$ → **FAIL 2** if $|M| > c'n$

↓

Output: $A' = A \setminus V(M), B' = B \setminus \mathrm{V}(M)$

Analysis of Pr(FAIL 1)

- Pre-processing:

Recall $\sum_{u,v} \|f(u) - f(v)\|^2 = n^2$

• Dispose of u s.t. $\underbrace{\|f(u)\|^2 < 1/10}_{\text{well-spread.}}$ or $\|f(u)\|^2 > 10$.

$\rightarrow$ $n/2$ points remaining, now $\|f(u)\|^2 \in [1/10, 10]$.

FAIL1 if $|A| < cn$, or $|B| < cn$.

$\mathbb{E}[|A| \cdot |B|] \gtrsim \Omega(n^2)$. $\Rightarrow \Pr[|A| \cdot |B| < cn^2] \leq 1 - c_1$

for suitably chosen c, c_1.

Blank Page

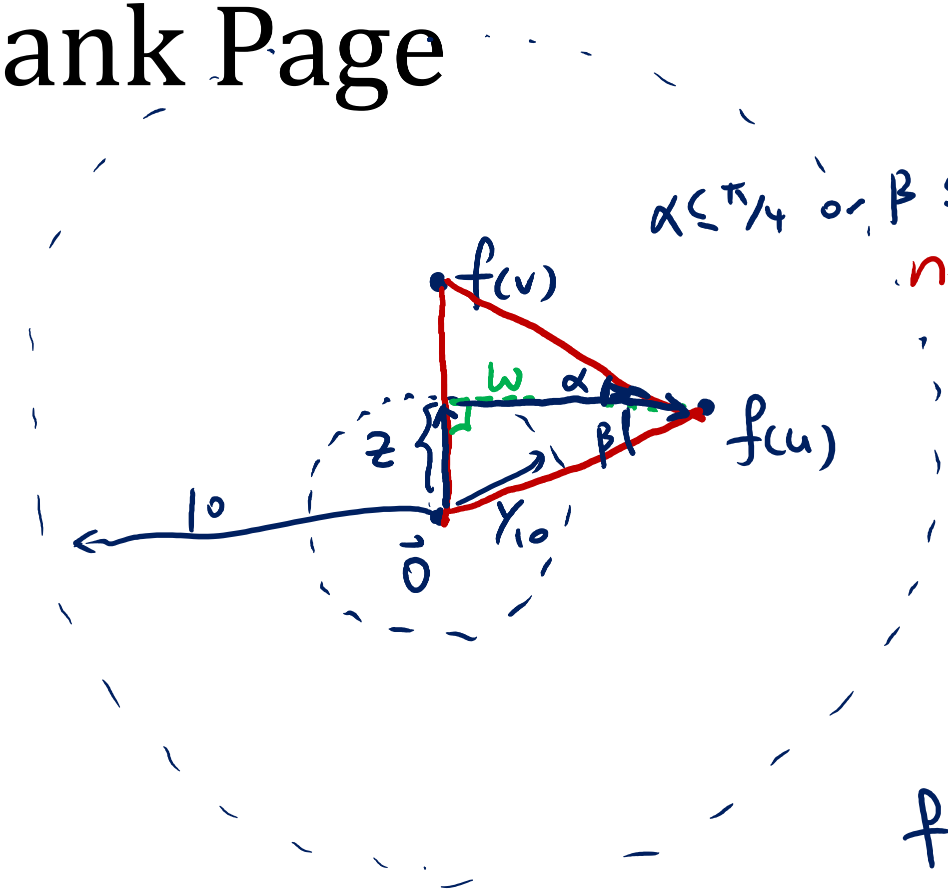
$$\mathbb{E}[|A|\cdot|B|] = \sum_{u,v \in V} \Pr[u \in A, v \in B]$$

$$\geq \sum_{u \in V} \sum_{v : d(u,v) \geq 1/10} \Pr[u \in A, v \in B]$$

\# pairs (u,v) s.t. $d(u,v) \geq 1/10$ is $\Omega(n^2)$
by well spread.

Goal: $\Pr[u \in A, v \in B] \geq \Omega(1)$ for $d(u,v) \geq 1/10$.

Blank Page



$\alpha \le \pi/4$ or $\beta \le \pi/4$.

non-obtuse.

$\alpha \Rightarrow \|w\| \geq \|f(v) - f(u)\| \cdot \cos\alpha$
$\geq \Omega(1)$

$\beta \Rightarrow \|w\| \geq \|f(u)\| \cdot \cos\alpha$
$\geq \Omega(1)$

Plan is to decompose g.

$v \in A \iff \underline{\langle f(v), g \rangle \leq -1}$

$u \in B \iff \langle f(u), g \rangle \geq +1$.

$f(u) = z + w$

Given direction g. if $\langle g, w \rangle$ large

and $\langle g, z \rangle$ not too small,

$(\langle g, f(u) \rangle = \langle g, z+w \rangle \geq 3 + (-2) = 1.)$

$v \in A$ and $u \in B$

$\Leftarrow \underline{\langle g, w \rangle \geq +3 \text{ and}}$

$\underline{\langle g, f(v) \rangle \in [-2, -1]}$

Blank Page

- $\Pr[\langle g, f(w)\rangle \in [-2,-1]] \geq \Omega(1)$

  since $\|f(w)\|^2 \in [1/10, 10]$

- If $\|w\|$ is lower bounded, then

  $\Pr[\langle g, w\rangle \geq +3] \geq \Omega(1)$. ✓

Proof idea that Pr(FAIL 2) is small

- Consider the graph of Δ-short edges: $E = \{(u, v) \in V \times V : d(u, v) < \Delta\}$
- Start from a vertex u_0, travel along k short edges $u_0 \to u_1 \to \cdots \to u_k$, so that each $\langle u_{i+1} - u_i, g\rangle$ is at least 2
- For at least one vertex u_0, for "many" directions g, can find path $u_0 \to u_1 \to \cdots \to u_k$ s.t. $\langle u_k - u_0, g\rangle \geq \Omega(k)$
- On the other hand, since $\|u_k - u_0\| \leq \sqrt{\sum_i \|u_{i+1} - u_i\|^2} \leq \sqrt{k}\,\Delta$, for **any** u_0' $\Pr_g(|\langle u_k' - u_0', g\rangle| \geq C\sqrt{\log n} \cdot \sqrt{k\Delta})$ is very small
- Plug suitable $k = \Theta(\sqrt{\log n})$ and $\Delta = \Theta(1/\sqrt{\log n})$ to get contradiction

Next time

- A proof that FAIL 2 probability is small
- ARV through the lens of sos (à la Barak & Steurer)
- A comparison of ARV original, Rothvoss, and Barak & Steurer

- I will also write up some supplementary notes