Lecture 7: Arora Rao Vazirani

Lecture Outline

- Part I: Semidefinite Programming Relaxation for Sparsest Cut
- Part II: Combining Approaches
- Part III: Arora-Rao-Vazirani Analysis Overview
- Part IV: Analyzing Matchings of Close Points
- Part V: Reduction to the Well-Separated Case
- Part VI: Open Problems

Part I: Semidefinite Programming Relaxation for Sparsest Cut

Problem Reformulation

- Reformulation: Want to minimize $\sum_{i,j:i<j,(i,j)\in E(G)} \left(x_j - x_i\right)^2$ over all cut pseudo-metrics normalized so that $\sum_{i,j:i<j} \left(x_j - x_i\right)^2 = 1$
- More precisely, take $d^2(i,j) = \left(x_j - x_i\right)^2$ and minimize $\sum_{i,j:i<j,(i,j)\in E(G)} d^2(i,j)$ subject to:
  1. $\exists c: \forall i, \mathrm{x_i} \in \{-c, +c\}$
  2. $\sum_{i,j:i<j} d^2(i,j) = 1$

Problem Relaxation

- Reformulation: Minimize

  $\sum_{i,j:i<j,(i,j)\in E(G)} (x_i^2 - 2x_i x_j + x_j^2)$ subject to:

  1. $\exists c: \forall i, \mathrm{x_i} \in \{-c, +c\}$
  2. $\sum_{i,j:i<j} (x_i^2 - 2x_i x_j + x_j^2) = 1$

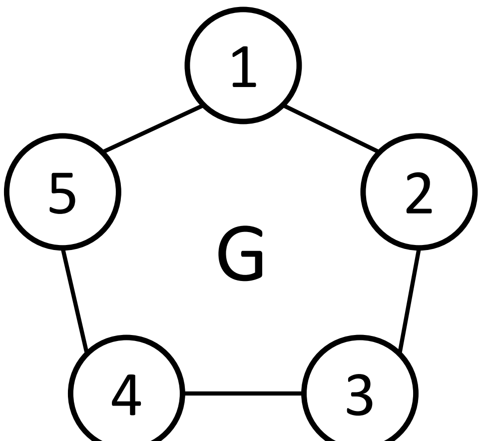
- Relaxation: Minimize

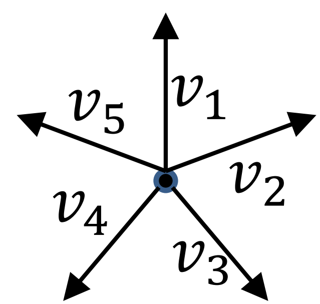
  $\sum_{i,j:i<j,(i,j)\in E(G)} (M_{ii} - 2M_{ij} + M_{jj})$ subject to:

  1. $\forall i, j, M_{ii} = M_{jj}$
  2. $\sum_{i,j:i<j} (M_{ii} - 2M_{ij} + M_{jj}) = 1$
  3. $M \succcurlyeq 0$

Bad Example: The Cycle

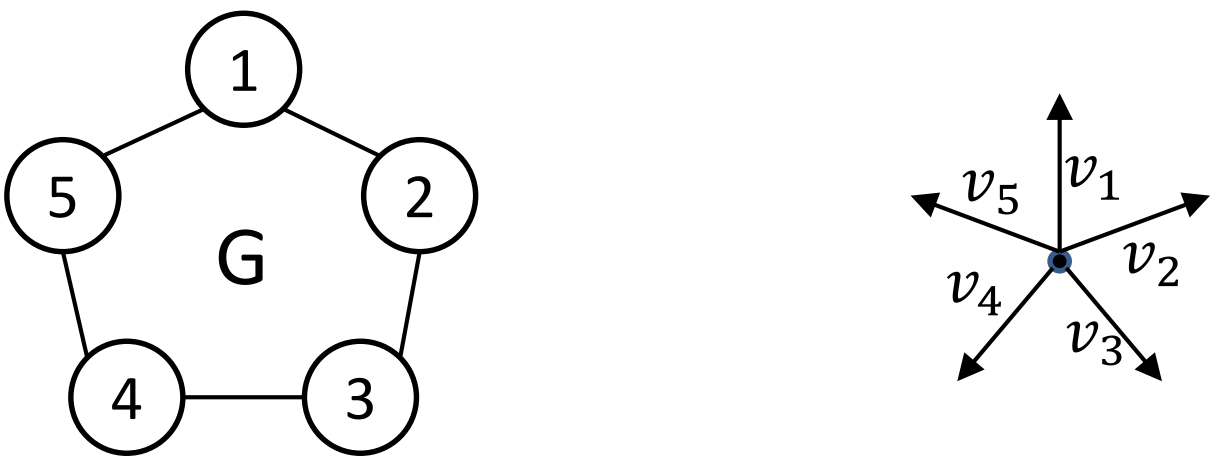
- Consider the cycle of length n. The semidefinite program can place the cycle on the unit circle and assign each x_i the corresponding vector v_i.





Bad Example: The Cycle

- $\sum_{i,j:i<j}(d^2(i,j)) = \Theta(n^2)$
- $\sum_{i,j:i<j,(i,j)\in E(G)}(d^2(i,j)) = \Theta(n \cdot 1/n^2)$
- Gives sparsity $\Theta(1/n^3)$, true value is $\Theta(1/\mathrm{n}^2)$
- Gap is $\Omega(n)$, which is horrible!



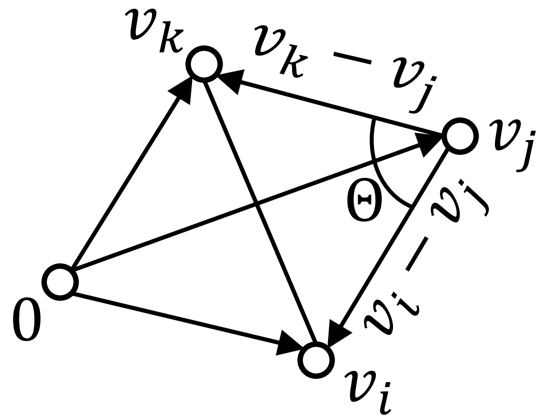
Part II: Combining Approaches

Adding the Triangle Inequalities

- Why did the semidefinite program do so much worse than the linear program?
- Missing: Triangle inequalities
$$d^2(i,k) \leq d^2(i,j) + d^2(j,k)$$
- What happens if we add the triangle inequalities to the semidefinite program?

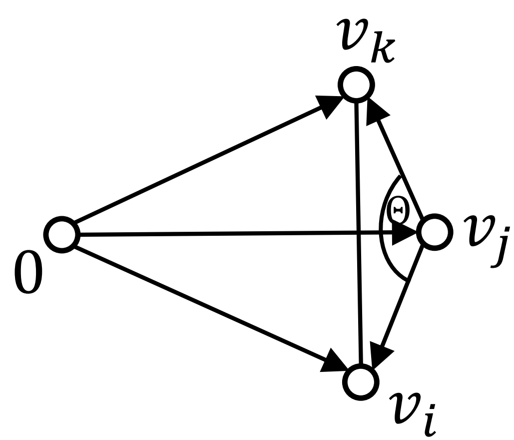
Geometric Picture

- Let Θ be the angle between $v_i - v_j$ and $v_k - v_j$
- $\|v_k - v_i\|^2 = \|v_j - v_i\|^2 + \|v_k - v_j\|^2$ if $\Theta = \frac{\pi}{2}$
- $\|v_k - v_i\|^2 > \|v_j - v_i\|^2 + \|v_k - v_j\|^2$ if $\Theta > \frac{\pi}{2}$
- $\|v_k - v_i\|^2 < \|v_j - v_i\|^2 + \|v_k - v_j\|^2$ if $\Theta < \frac{\pi}{2}$
- Triangle inequalities ⇔ no obtuse angles



Fixing Cycle Example

- Putting $n > 4$ vectors in a circle violates triangle inequality, so the semidefinite program no longer behaves badly on the cycle. In fact, it gets very close to the right answer.



Goemans-Linial Relaxation

- Semidefinite program (proposed by Goemans and Lineal): Minimize $\sum_{i,j:i<j:(i,j)\in E(G)}(M_{ii} - 2M_{ij} + M_{jj})$ subject to:
  1. $\forall i, j, \mathrm{M}_{ii} = M_{jj}$
  2. $\forall i, j, k, d^2(i,k) \leq d^2(i,j) + d^2(j,k)$ where $d^2(i,j) = M_{ii} - 2M_{ij} + M_{jj}$
  3. $\sum_{i,j:i<j}\left(M_{ii} - 2M_{ij} + M_{jj}\right) = 1$
  4. $M \succcurlyeq 0$

Arora-Rao-Vazirani Theorem

- Theorem [ARV]: The Goemans-Linial relaxation for sparsest cut gives an $O\left(\sqrt{logn}\right)$-approximation and has a polynomial time rounding algorithm.

L_2^2 Metric Spaces

- Also called metrics of negative type
- Definition: A metric is an L_2^2 metric if it is possible to assign a vector v_x to every point x such that $d(x,y) = \left\|v_y - v_x\right\|^2$.
- Last time: General metrics can be embedded into L^1 with $O(\log n)$ distortion.
- Theorem [ALN08]: Any L_2^2 metric embeds into L^1 with $O\left(\sqrt{logn}(loglogn)\right)$ distortion.
- [ARV] analyzes the algorithm more directly

Goemans-Linial Relaxation and SOS

- Degree 4 SOS captures the triangle inequality: if $x_i^2 = x_j^2 = x_k^2$ then
$$x_i^2(x_k - x_i)^2 \leq x_i^2\left(x_j - x_i\right)^2 + x_i^2\left(x_k - x_j\right)^2$$
$$\Leftrightarrow 2x_i^2\left(x_i^2 - x_i x_k\right) \leq 2x_i^2(2x_i^2 - x_i x_j - x_i x_j)$$
- Proof:
$$\left(x_i - x_j\right)^2\left(x_j - x_k\right)^2 = 4(x_i^2 - x_i x_j)(x_i^2 - x_j x_k)$$
$$= 4x_i^2\left(x_i^2 - x_i x_j - x_j x_k + x_i x_k\right) \geq 0$$
- Thus, degree 4 SOS captures the Goemans-Linial relaxation

Part III: Arora-Rao-Vazirani Analysis Overview

Well-Spread Case

- Semidefinite program gives us one vector v_i for each vertex i.
- We first consider the case when these vectors are spread out.
- Definition: We say that a set of n vectors $\{v_i\}$ is well-spread if it can be scaled so that:
  1. $\forall i, \|v_i\| \leq 1$
  2. $\frac{1}{n^2}\sum_{i<j} d_{ij}^2$ is $\Omega(1)$ (the average squared distance between vectors is constant)
- We will assume we are using this scaling.

Structure Theorem

- Theorem: Given a set of n vectors $\{v_i\}$ which are well-spread and obey the triangle inequality, there exist well-separated subsets X and Y of these vectors of linear size. In other words, there exist X, Y such that:
  1. X and Y are Δ far apart (i.e. $\forall v_i \in X, v_j \in Y, d_{ij}^2 \geq \Delta$) where Δ is $\Omega\left(\frac{1}{\sqrt{logn}}\right)$
  2. $|X|$ and $|Y|$ are both $\Omega(n)$

Finding a Sparse Cut

- Idea: If we have well-separated subsets X, Y, take a random cut of the form $(S_r, \bar{S}_r)$ where

$$S_r = \{i : d^2(v_i, X) = \min_{j:v_j \in Y} d_{ij}^2 \leq r\} \text{ and } r \in [0, \Delta]$$

- All $(i, j) \in E(G)$ contribute at most $\frac{d_{ij}^2}{\Delta}$ to the expected number of edges cut and d_{ij}^2 to $\sum_{i,j:i<j,(i,j)\in E(G)} d_{ij}^2$ (the number of edges the SDP "thinks" are cut)

Finding a Sparse Cut Continued

- Since X, Y have size $\Omega(n)$ and are always on opposite sides of the cut, we always have that $|S_r| \cdot |\bar{S}_r|$ is $\Theta(n^2)$. This matches $\sum_{i,j:i<j} d_{ij}^2$ up to a constant factor. (this is why we need X and Y to have linear size!)
- Thus, the expected ratio of the sparsity to the SDP value is at most $\frac{1}{\Delta} = O\left(\sqrt{logn}\right)$, as needed.

Tight Example: Hypercube

- Take the hypercube $\left\{-\frac{1}{\sqrt{\log_2 n}}, \frac{1}{\sqrt{\log_2 n}}\right\}^{\log_2 n}$
- $\mathrm{X} = \left\{x: \sum_i x_i \leq -1\right\}$ and $\mathrm{Y} = \left\{y: \sum_i x_i \geq 1\right\}$ have the following properties:
  1. X and Y have linear size
  2. $\forall x \in X, y \in Y$, x, y differ in $\geq 2\sqrt{\log_2 n}$ coordinates. Thus, $d^2(x,y) \geq \frac{2\sqrt{\log_2 n}}{\log_2 n} = \frac{2}{\sqrt{\log_2 n}}$

Finding Well-Separated Sets

- Let d be the dimension such that $\forall i, v_i \in \mathbb{R}^d$.
- Algorithm (Parameters $\sigma > 0, \Delta, d$)
  1. Choose a random $u \in \mathbb{R}^d$.
  2. Find a value a such that there are $\Omega(n)$ vectors v_i with $v_i \cdot u \leq a$ and $\Omega(n)$ vectors v_j with $v_j \cdot u \geq a + \frac{\sigma}{\sqrt{d}}$. Let X' and Y' be these two sets of vectors
  3. As long as there is a pair $x \in X', y \in Y'$ such that $d(x, y) < \Delta$, delete x from X' and y from Y'. The resulting sets will be the desired X, Y.
- Need to show: $\mathrm{P}[X, Y \text{ have size } \Omega(n)]$ is $\Omega(1)$

Finding Well-Separated Sets

- Will first explain why step 1,2 succeed with probability $2\delta > 0$.
- Will then show that the probability step 3 deletes a linear number of points is $\leq \delta$
- Together, this implies that the entire algorithm succeeds with probability at least $\delta > 0$.

Behavior of Gaussian Projections

- What happens if we project a vector v of length l in a random direction in $\mathbb{R}^d$?
- Without loss of generality, assume $v = e_1$
- To pick a random unit vector in $\mathbb{R}^d$, choose each coordinate according to $N\left(0, \frac{1}{d}\right)$ (the normal distribution with mean 0 and standard deviation $\frac{1}{\sqrt{d}}$), then rescale.
- If d is not too small, w.h.p. very little rescaling will be needed.

Behavior of Gaussian Projections

- What happens if we project a vector of length l in a random direction in $\mathbb{R}^d$?
- Resulting value has a distribution which is $\approx$ normal distribution of mean 0, standard deviation $\frac{1}{\sqrt{d}}$ (difference comes from the rescaling step)

Success of Steps 1,2

- If we take a random $u \in \mathbb{R}^d$, with probability $\Omega(1)$, $\sum_{i<j}\left|(v_j - v_i) \cdot u\right|$ is $\Omega\left(\frac{n^2}{\sqrt{d}}\right)$
- Note: this can fail with non-negligible probability, consider the case when $\forall i, v_i = \pm v$. If u is orthogonal to v then everything is projected to 0.
- For arbitrarily small $\epsilon > 0$, with very high probability, $|v_i \cdot u|$ is $O\left(\frac{1}{\sqrt{d}}\right)$ for $(1-\epsilon)n$ of the $i \in [1, n]$

Success of Steps 1,2

- Together, these facts imply that if we choose a random unit vector u, with probability $\Omega(1)$, there exist X', Y', a_1, a_2 such that
  1. X', Y' have size $\Omega(n)$
  2. $\forall x \in X', u \cdot x \leq a_1$
  3. $\forall y \in Y', u \cdot y \geq a_2$
  4. $a_2 - a_1$ is $\Omega(1)$

Remaining Steps

- We need to show that the probability step 3 eliminates $\frac{min\{|X|,|Y|\}}{2}$ pairs of points is at most δ
- We also need to show how the general case can be reduced to the well-spread case.

Part IV: Analyzing Matchings of Close Points

Matching Covers

- If part 3 of the algorithm causes it to fail with probability δ, then for δ fraction of the directions u there is a matching M_u of points of size $c'n$ such that for each pair (v_i, v_j) in the matching:
  1. $\mathrm{d}^2(v_i, v_j) \leq \Delta$
  2. $|(v_j - v_i) \cdot u| \geq \frac{2\sigma}{\sqrt{d}}$

  where $\delta, c', \sigma > 0$ are constants
- Note: Corresponds to Definition 4 in [ARV]
- Define the matching graph M to be $M = \cup_u M_u$

Analyzing $\Delta = \Omega\left(\frac{1}{logn}\right)$

- Assume that $d(v_i, v_j) \leq \sqrt{\Delta}$ for some v_i, v_j
- $\mathrm{P}\left[\left|(v_j - v_i)\cdot u\right| \geq \frac{2\sigma}{\sqrt{d}}\right] \sim e^{-\frac{4\sigma^2}{d^2(v_i, v_j)}} \leq e^{-\frac{4\sigma^2}{\Delta}}$
- If Δ is a sufficiently small constant times $\frac{1}{logn}$, with high probability there are no pairs of close points at all between X' and Y'!

Key Idea for Larger Δ

- When the algorithm fails in step 3, this gives us pairs of points (v_i, v_j) which are edges of the matching graph M, implying that $d^2(v_i, v_j) \leq \Delta$ and $\left|(v_j - v_i) \cdot u\right| \geq \frac{2\sigma}{\sqrt{d}}$
- We will use this to find pairs of points (v_i, v_j) which are k steps apart in the matching graph where $\left|(v_j - v_i) \cdot u\right| \geq \frac{k\sigma}{\sqrt{d}}$

Key Idea for Larger Δ Continued

- We will find pairs of points (v_i, v_j) which are k steps apart in the matching graph where $\left|\left(v_j - v_i\right) \cdot u\right| \geq \frac{k\sigma}{\sqrt{d}}$
- Using triangle inequality, $d^2\left(v_i, v_j\right) \leq k\Delta$
- $\mathrm{P}\left[\left|\left(v_j - v_i\right) \cdot u\right| \geq \frac{k\sigma}{\sqrt{d}}\right] \sim e^{-\frac{k^2\sigma^2}{d^2(v_i, v_j)}} \leq e^{-\frac{k\sigma^2}{\Delta}}$
- For $\Delta = \Omega\left(\frac{1}{\sqrt{\log n}}\right)$, if we can apply this with $k = \Omega\left(\sqrt{\log n}\right)$, we again obtain a contradiction.

Average Degree to Minimal Degree

- Lemma: If a graph G has average degree d, we can find a non-empty subgraph of G which has minimal degree $\frac{d}{4}$.
- Proof: Iteratively delete vertices which have degree $\leq \frac{d}{4}$. The total number of edges deleted is at most $\frac{nd}{4}$. However, $2|E(G)| \geq nd$, so there must be $\geq \frac{nd}{4}$ edges remaining.

Minimal Probability Guarantee

- Average probability that a vertex is matched is at least $c'\delta$
- Can apply a similar idea and delete any vertex which is matched with probability $\leq \frac{c'\delta}{4}$
- By similar logic, at least half the edges are preserved.
- This implies that there are at least $c'n$ vertices remaining (otherwise more than half of every matching of $\geq c'n$ edges is deleted)
- Note: Corresponds to Lemma 4 of [ARV09]

Minimal Probability Guarantee

- Corollary: There is a set of vertices X of size $\geq c'n$ such that $\forall x \in X, P[x \text{ is matched with an } x' \in X] \geq \delta'$

  where $\delta' = \frac{c'\delta}{4}$

Building Up Projection Distances

- How can we find pairs of points whose projected distance is larger and larger by taking steps in the matching graph?
- Let’s assume we have a very convenient inductive setup.

Setup

- Have a set of points X of size $\geq c'n$

  $\forall x \in X, P[x \text{ is matched with an } x' \in X] \geq \delta'$
- Inductive setup: Assume we also have a subset $Z \subseteq X$ of points of size $\tau|X|$ such that

  $$\forall z \in Z, P\left[\exists z' \in X: d_M(z,z') \leq k, (z-z')\cdot u \geq \frac{k\sigma}{\sqrt{d}}\right] \geq 1 - \frac{\delta'}{4}$$

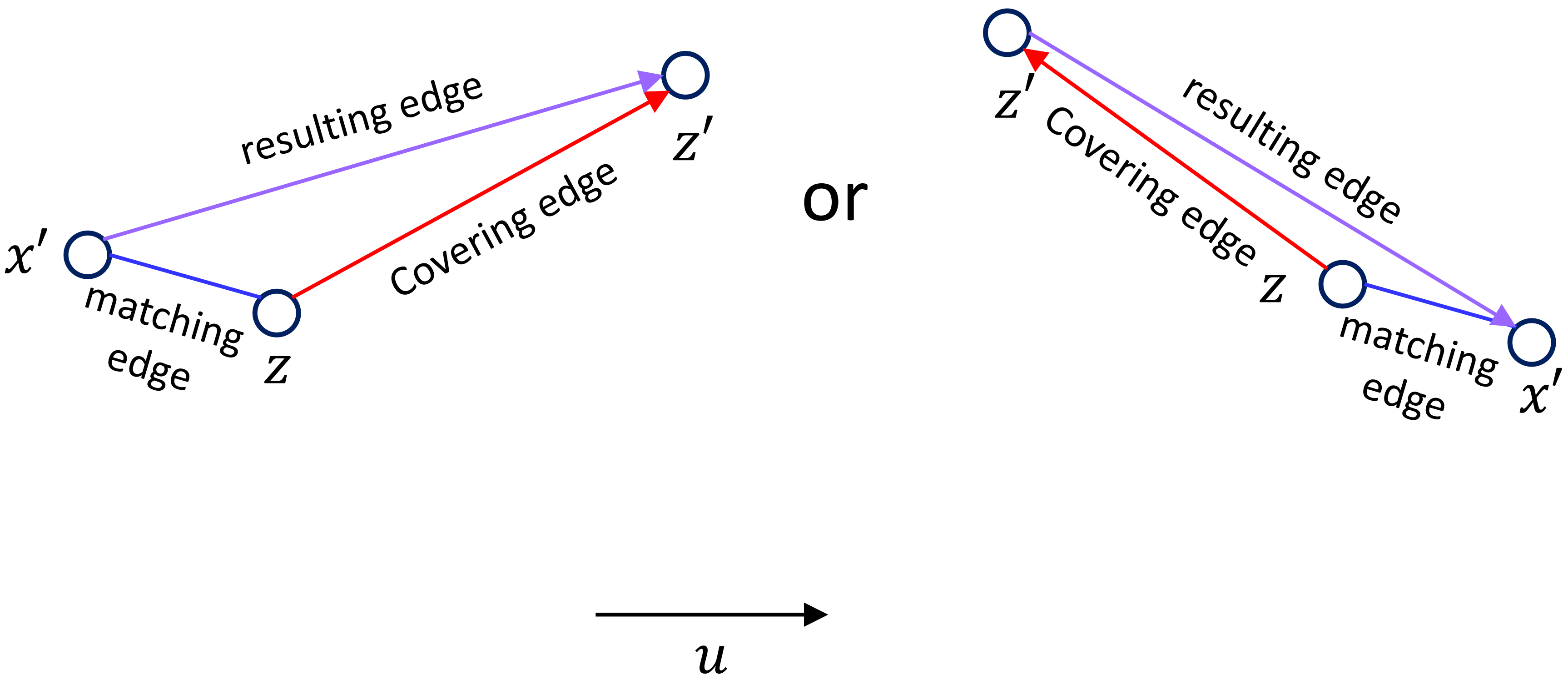
where $d_M(z,z')$ is the number of steps required to reach z' from z in the matching graph

- Note: This corresponds to Definitions 6,8 of [ARV]

Setup Rephrased

- X is a set of points where every $x \in X$ is matched to another $x' \in X$ for $\geq \delta'$ fraction of the directions
- Have a subset $Z \subseteq X$ of size $\geq \tau|X|$ where each $z \in Z$ is "covered" in $\geq 1 - \frac{\delta'}{4}$ fraction of the directions by points which are $\leq k$ steps away in the matching graph whose projected distance is $\geq \frac{k\sigma}{\sqrt{d}}$

Composition Step



Composition Step

- Given a direction u, for each point $z \in Z$:
  1. Check if z is matched in $M_u = M_{-u}$
  2. If so, let x' be the point z is matched with. $|(z - x') \cdot u| \geq \frac{2\sigma}{\sqrt{d}}$
  3. If $(z - x') \cdot u > 0$, check if z is covered in direction u. If $(z - x') \cdot u < 0$ check if z is covered in direction $-u$. With probability $\geq 1 - \frac{\delta'}{2}$, z is covered in both directions. Let $z' =$ covering point.
  4. Observe that $|(z' - x') \cdot u| \geq \frac{k\sigma + 2\sigma}{\sqrt{d}}$ and $d_M(x', z') \leq k + 1$

Composition Step

- Have that the density of the new covering edges is at least $\frac{\tau\delta'}{2}$.
- Following the same kind of logic we used to go from average to minimal degree, can find a subset $Z' \subseteq X$ of size $\geq \frac{\tau\delta'}{8}|X|$ where every vertex $z' \in Z'$ is covered in $\geq \frac{\tau\delta'}{8}$ of the directions.
- Note: Corresponds to Lemma 11 of [ARV]

Boosting Lemma

- How can we recover the inductive hypothesis?
- Can boost the covering probability to almost 1 with a small loss in the projection length!
- Corollary 12 of [ARV] rephrased: If the covering vectors have length at most $\frac{\sigma}{16\sqrt{\log\left(\frac{16}{\tau\delta'}\right)}+8\sqrt{\log\left(\frac{8}{\delta'}\right)}}$ then if z is covered with probability $\frac{\tau\delta'}{8}$ with projection length $\frac{k\sigma+2\sigma}{\sqrt{d}}$, it is covered with probability $1-\delta'/4$ with projection length $\frac{(k+1)\sigma}{\sqrt{d}}$

Bound on k and Δ

- If we apply this directly:
  - $\tau \sim (\delta')^{-k}$
  - Need covering vectors to have length $O\left(\frac{1}{\sqrt{log\tau}}\right) = O\left(\frac{1}{\sqrt{k}}\right)$
  - Guaranteed to have length $\leq \sqrt{k\Delta}$
  - We can take $k = \Omega(\Delta^{-\frac{1}{2}})$. We want $\frac{k}{\Delta}$ to be a large constant times $\log(n)$, which means we can take $\Delta = \Omega((logn)^{-2/3})$

Reaching $k = \Omega\left(\sqrt{logn}\right)$

- To reach $k = \Omega\left(\sqrt{logn}\right)$, a more careful argument is needed, see [ARV].
- Note: We should not expect k to be any higher than $\mathrm{O}\left(\sqrt{logn}\right)$. Recalling that the projection length with k steps is $\frac{k\sigma}{\sqrt{d}}$, if $d = \Theta(logn)$ (matching the hypercube example) and k is $\omega\left(\sqrt{logn}\right)$ then this is $\omega(1)$, which is too large!

Part V: Reduction to the Well-Separated Case

Two Cases

- Take the scaling where $\sum_{i,j:i<j} d^2(i,j) = \binom{n}{2}$ (i.e. the average squared distance between pairs of points is 1)
- One of the following two cases holds:
  1. There exists a point x_0 such that $\frac{n}{10}$ other points are within squared distance $\frac{1}{10}$ of x_0
  2. For all points x, less than $\frac{n}{10}$ other points are within squared distance $\frac{1}{10}$ of x

Case #1

- Assume there exists a point x_0 such that $\frac{n}{10}$ other points are within squared distance $\frac{1}{10}$ of x_0
- Let $X = \{\mathrm{x}: \mathrm{d}^2(\mathrm{x}, \mathrm{x}_0) \leq \frac{1}{10}\}$
- Key idea: Take the Fréchet embedding with respect to X!
- In particular, take

$$d_X(y, z) = |d^2(y, X) - d^2(z, X)|$$

Case #1 Continued

- We will show that

$$\frac{\sum_{i,j:i<j,(i,j)\in E(G)} d_X(i,j)}{\sum_{i,j:i<j} d_X(i,j)} \text{ is } O\left(\frac{\sum_{i,j:i<j,(i,j)\in E(G)} d^2(i,j)}{\sum_{i,j:i<j} d^2(i,j)}\right)$$

- d_X is an L^1 metric, so this gives an $O(1)$-approximation!
- First note that $\sum_{i,j:i<j,(i,j)\in E(G)} d_X(i,j)$ is less than or equal to $\sum_{i,j:i<j,(i,j)\in E(G)} d^2(i,j)$
- We just need to show that $\sum_{i,j:i<j} d_X(i,j)$ is $\Omega(n^2)$

Case #1 Continued

- Proposition: The average squared distance of points outside of X from X is at least $\frac{1}{5}$
- Proof: If this were not the case then the average squared distance between points would be < 1 as for all y, z,

$$d^2(y,z) \leq d^2(y,X) + d^2(z,X) + \frac{1}{5}$$

- Corollary: $\sum_{i,j:i<j} d_X(i,j)$ is $\Theta\left(n^2\right)$. To show this, it is sufficient to consider the pairs where exactly one of i, j are in X.

Case #2

- Assume that for all points x, there are fewer than $\frac{n}{10}$ other points which are within squared distance $\frac{1}{10}$ of x
- Proposition: There is a point x_0 such that at least $\frac{n}{2}$ other points are within distance 2 of x_0
- Proof: If this were not the case then the average distance between points would be > 1.
- Let X be the set of points within distance 2 of x_0.

Case #2 Continued

- Key idea: Subtract x_0 from all vectors!
- After this translation:
  - All points in X have length ≤ 2
  - For all points $x \in X$, there are at least $\frac{\mathrm{n}}{2} - \frac{\mathrm{n}}{10} = \frac{2n}{5}$ points in X which have squared distance more than $\frac{1}{10}$ from x. Thus, the average squared distance between points in X is $\Omega(1)$
- Restricting to X and scaling down by a factor of 2, we are now in the well-spread case

Part VI: Open Problems

Lower Bounds

- Lower Bounds have been shown for this semidefinite program
- Khot and Vishnoi [KV05] proved the first super-constant lower bound.
- For weighted graphs, Naor and Young [NY17] showed an $\Omega\left(\sqrt{logn}\right)$ lower bound (which is tight up to a $loglogn$ factor).
- However, these lower bounds don’t apply even to degree 4 SOS!

Open Questions

- Is this also true for unweighted graphs?
- Does degree 4 SOS or higher degree SOS give further improvements? Can we show a superconstant lower bound for a constant number of rounds of SOS?

References

- [ALN08] S. Arora, J. R. Lee, A. Naor. Euclidean distortion and the sparsest cut. J. Amer. Math. Soc. 21 (1), p. 1–21. 2008
- [ARV] S. Arora, S. Rao, U. Vazirani. Expander Flows, Geometric Embeddings and Graph Partitioning. https://www.cs.princeton.edu/~arora/pubs/arvfull.pdf
- [KV05] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into L^1. FOCS 2005
- [NY17] A. Naor and R. Young. The integrality gap of the Goemans--Linial SDP relaxation for Sparsest Cut is at least a constant multiple of $\sqrt{logn}$. STOC 2017