

Arora–Rao–Vazirani Approximation for Expansion

In this lecture, we consider the problem of finding a set with smallest possible expansion¹ in a given graph. Earlier we saw that degree-2 pseudo-distributions allow us to achieve non-trivial approximation guarantee, which turns out to be closely related to Cheeger’s inequality from spectral graph theory (see [lecture 2](#)). We also showed that this guarantee is tight for degree-2 pseudo-distributions (see [lecture 3](#)). In this lecture, we investigate the question if higher-degree pseudo-distributions allow us to achieve better approximation guarantee. An indication that higher degree might indeed help is that certain linear programming approaches, that are captured by higher-degree pseudo-distributions, achieve approximation guarantees for expansion that are better than those from Cheeger’s inequality in some range of parameters ([Leighton and Rao \[1988\]](#)). (In particular, those linear programming approaches certify tight lower bounds on the expansion of cycle graphs, for which no good degree-2 sos certificates exist.)

We will show that degree-4 pseudo-distributions achieve approximation guarantees for expansion ([Arora et al. \[2004\]](#)) that for a large range of parameters are significantly stronger than both the guarantees of degree-2 pseudo-distributions and the guarantees of previous linear programming approaches.

Our exposition (especially the proof of the “structure theorem”) differs significantly from others in the literature. However, it borrows heavily from lectures notes of [Rothvoss \[2016\]](#).

Pseudo-distributions for expansion

Let G be a d -regular graph with vertex set $[n]$.² Let $f_G: \{0, 1\}^n \rightarrow \mathbb{R}$ be the function that outputs for every vector $x \in \{0, 1\}^n$ the number of edges $f_G(x)$ cut by x (viewed as a bipartition of the vertex set),

$$f_G(x) = \sum_{\{i,j\} \in E(G)} (x_i - x_j)^2. \quad (1)$$

We identify the vector $x \in \{0, 1\}^n$ with the subset $S = \{i \in [n] \mid x_i = 1\}$ indicated by x . For every $i, j \in [n]$, the function $(x_i - x_j)^2$ indicates whether i and j are on different sides of the bipartition corresponding to x . Recall the expansion of G ,

$$\varphi(G) = \min_{x \in \{0,1\}^n} \frac{f_G(x)}{\frac{d}{n}|x| \cdot (n - |x|)}. \quad (2)$$

¹ In the context of this algorithm, the computational problem being solved is typically called *Sparsest Cut*. We will refer to this problem as *Min Expansion* in order to be consistent with earlier lectures.

² The treatment of expansion here also works for non-regular graph. We assume regularity in order to be consistent with earlier definitions in the context of Cheeger’s inequality and degree-2 sum-of-squares.

Here, $|x| = \sum_{i=1}^n x_i$ is the size of the set indicated by x .

Note that actual probability distributions $\mu: \{0,1\}^n \rightarrow \mathbb{R}_{\geq 0}$ satisfy $\mathbb{E}_\mu f_G \geq \varphi(G) \cdot \mathbb{E}_{\mu(x)} \frac{d}{n} \cdot |x|(n - |x|)$ and equality is achieved by distributions supported on sets with minimum expansion. If we consider degree-4 pseudo-distributions, we can find in polynomial time one such that the ratio of $\tilde{\mathbb{E}}_\mu f_G$ and $\tilde{\mathbb{E}}_{\mu(x)} \frac{d}{n} |x|(n - |x|)$ is as small as possible—certainly no more than $\varphi(G)$. The following theorem shows that this ratio can never be much smaller than $\varphi(G)$ and that we can extract from the pseudo-distribution a set with small expansion at most $\varphi(G) \cdot O(\sqrt{\log n})$.

1. Theorem (Arora–Rao–Vazirani approximation for Min Expansion). *Let G be a d -regular graph with vertex set $[n]$ and let $\mu: \{0,1\}^n \rightarrow \mathbb{R}$ be a degree-4 pseudo-distribution. Then,*

$$\varphi(G) \leq O\left(\sqrt{\log n}\right) \cdot \frac{\tilde{\mathbb{E}}_\mu f_G}{\tilde{\mathbb{E}}_{\mu(x)} \frac{d}{n} |x| \cdot (n - |x|)}. \quad (3)$$

Furthermore, there exists a polynomial-time algorithm that given G and μ finds a vector $x \in \{0,1\}^n$ witnessing this bound on $\varphi(G)$.

A birds-eye view of the proof

The proof of [Theorem 1](#) consists of several part. We describe each of them at a high level and explain how they fit together.

The first component is an algorithm called *region growing*. Given two subsets $A, B \subseteq [n]$ this algorithm tries to find the best cut of the graph between A and B guided by the pseudo-distribution μ . As we will see, this algorithm achieves a $O(1/\Delta)$ -approximation (compared to the pseudo-distribution) if the sets A, B are Δ -separated in the sense that $|A|, |B| \geq \Omega(n)$ and $\mathbb{E}_{\mu(x)} (x_i - x_j)^2 \geq \Delta$ for every $i \in A$ and $j \in B$.

The second component is a *structure theorem* for pseudo-distributions that shows that for some $\Delta \geq \Omega(1)/\sqrt{\log n}$ such Δ -separated sets always exist.³

To prove this structure theorem we apply the quadratic sampling lemma (from [lecture 1](#)) in order to obtain Gaussian variables X_1, \dots, X_n with the same first two moments as the pseudo-distribution. A natural first attempt to construct separated sets from those variables is to choose A to be all vertices $i \in [n]$ such that X_i is significantly below its mean, say $X_i \leq \mathbb{E} X_i - 1$ and to choose B to be all vertices $j \in [n]$ such that X_j is significant above its mean,

³ The actual proof requires a small case distinction here which we are going to ignore for this high-level discussion.

say $X_j \geq \mathbb{E} X_j + 1$. Indeed, this construction ensures that A and B are Δ -separated for some $\Delta \geq \Omega(1)/\log n$ (which in turn gives an $O(\log n)$ approximation).

To improve over this first attempt, we post-process the sets A and B by iteratively removing pairs $i \in A$ and $j \in B$ that violate the separation condition, i.e., $\mathbb{E}(X_i - X_j)^2 \leq \Delta$. The heart of the proof is to upper bound the number of disjoint pairs of this form. We will show that its expectation is at most $O(n) \cdot (\Delta \cdot \sqrt{\log n})^{\Omega(1)}$. It follows that if we choose $\Delta = c\sqrt{\log n}$ for a sufficiently small constant $c > 0$, then only a small fraction of vertices are removed in the post-processing phase in expectation.

To upper bound their expected number, we relate the above disjoint pairs to maxima of Gaussian processes defined in terms of the variables X_1, \dots, X_n . A crucial ingredient is an upper bound on the variance (and concentration) of the maximum of a Gaussian process and the fact that the Gaussian variables X_1, \dots, X_n satisfy squared triangle inequalities $\mathbb{E}(X_i - X_j)^2 \leq \mathbb{E}(X_i - X_k)^2 + \mathbb{E}(X_k - X_j)^2$ (because the pseudo-distribution μ satisfies this inequality and X_1, \dots, X_n have the same first two moments as μ).

Region growing

Let G be a d -regular graph with vertex set $[n]$. Let $\mu: \{0, 1\}^n \rightarrow \mathbb{R}$ be a degree-4 pseudo-distribution. Let $d(i, j) = \tilde{\mathbb{E}}_{\mu(x)}(x_i - x_j)^2$ be the probability under μ that i and j are on different sides of the bipartition. For a subset $S \subseteq [n]$, we let $d(i, S) = \min_{s \in S} d(i, s)$. Since the inequality $(x_i - x_j)^2 + (x_j - x_k)^2 \geq (x_i - x_k)^2$ holds over the hypercube for all $i, j, k \in [n]$ and this inequality as a deg-4 sos proof, we have the following triangle inequality for all $i, j, k \in [n]$

$$d(i, j) + d(j, k) \geq d(i, k). \quad (4)$$

The following lemma is the main tool for extracting a set with small expansion out of a degree-4 pseudo-distribution.

2. Lemma (Region growing). *Suppose for $\Delta > 0$ there exists a subset $A \subseteq [n]$ such that $|A| \cdot \sum_{i=1}^n d(i, A) \geq \Delta \cdot \sum_{i,j=1}^n d(i, j)$. Then,*

$$\varphi(G) \leq 1/\Delta \cdot \frac{\tilde{\mathbb{E}}_{\mu} f_G}{\tilde{\mathbb{E}}_{\mu(x)} \frac{d}{n} |x|(n - |x|)}. \quad (5)$$

Furthermore, one of the sets $A_t = \{i \in [n] \mid d(i, A) \leq t\}$ for $t \geq 0$ witnesses this bound on the expansion.

Proof. Consider the following distribution μ' over the hypercube:

- choose t uniformly at random in the interval $[0, 1]$
- output the vector $x' \in \{0, 1\}^n$ with $x'_i = 1$ if and only if $d(i, A) \leq t$.

Then, every pair $i, j \in [n]$ with $d(i, A) \leq d(j, A)$ satisfies

$$\begin{aligned} \mathbb{E}_{\mu'(x')} (x'_i - x'_j)^2 &= \mathbb{P}_{t \in [0,1]} \{d(i, A) \leq t \leq d(j, A)\} \\ &= |d(i, A) - d(j, A)| \end{aligned} \quad (6)$$

By the triangle inequality Eq. (4), $d(i, A) - d(j, A) \leq d(i, j)$. Therefore, $\mathbb{E}_{\mu'} f_G \leq \tilde{\mathbb{E}}_{\mu} f_G$.

At the same time, since $|x'| \geq |A|$ and $\mathbb{P}_{\mu'(x)} \{x'_i = 0\} = \mathbb{P}_{t \in [0,1]} \{t < d(i, A)\} = d(i, A)$,

$$\begin{aligned} \mathbb{E}_{\mu'(x')} |x'| (n - |x'|) &\geq |A| \cdot \sum_{i=1}^n d(i, A) \\ &\geq \Delta \cdot \sum_{i,j \in [n]} d(i, j) \\ &= \Delta \cdot \tilde{\mathbb{E}}_{\mu(x)} |x| \cdot (n - |x|). \end{aligned} \quad (7)$$

It follows that the expansion of G satisfies

$$\varphi(G) \leq \frac{\mathbb{E}_{\mu'} f_G}{\mathbb{E}_{\mu'(x')} |x'| (n - |x'|)} \leq \frac{1}{\Delta} \cdot \frac{\tilde{\mathbb{E}}_{\mu} f_G}{\tilde{\mathbb{E}}_{\mu(x)} |x| (n - |x|)}. \quad (8)$$

Furthermore, there exists a vector $x' \in \{0, 1\}^n$ in the support of μ' that witnesses this bound on the expansion. \square

It remains to show that we can always find a subset A as above for $\Delta \geq \Omega(1/\sqrt{\log n})$. For simplicity, we first focus on the following special case,⁴

$$\sum_{i,j \in [n]} d(i, j) \geq 0.1 \cdot n^2. \quad (9)$$

In this special case, we show that for some $\Delta \geq \Omega(1/\sqrt{\log n})$ there exist sets $A, B \subseteq [n]$ with $|A|, |B| \geq \Omega(n)$ such that $d(i, j) \geq \Delta$ for all $i \in A$ and $j \in B$. Indeed for this choice of A , the condition of the region growing lemma is satisfied,

$$|A| \cdot \sum_{i=1}^n d(i, A) \geq |A| \cdot |B| \cdot \Delta \geq \Omega(\Delta \cdot n^2) \geq \Omega(\Delta) \cdot \sum_{i,j \in [n]} d(i, j). \quad (10)$$

By Lemma 2, we obtain a $O(1/\Delta) \leq O(\sqrt{\log n})$ -approximation for the expansion of G .

⁴ If the graph has a sparsest cut such that the smaller side has size at least $0.2n$, we can require this inequality for the pseudo-distribution without loss of generality. Furthermore, it turns out that even the general case can be reduced to this special case. We present this reduction at the end of these notes.

Structure theorem

The quadratic sampling lemma allows us to translate the task of finding sets A and B as above to a question about n jointly-distributed Gaussian variables. Without loss of generality we may assume that $\mathbb{E}_{\mu(x)} x = \frac{1}{2} \cdot \mathbf{1}$.⁵

Let $X = (X_1, \dots, X_n)$ be a Gaussian vector that matches the first two moments of $x - \frac{1}{2} \cdot \mathbf{1}$ under the pseudo-distribution μ . Then, X satisfies $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 \leq 1$, and $\mathbb{E}(X_i - X_j)^2 = d(i, j)$ for all $i, j \in [n]$.

The following theorem about such Gaussian vectors shows that sets A and B as above exist and that we can find them efficiently. Together with [Lemma 2](#) this theorem implies [Theorem 1](#) in the special case that the pseudo-distribution satisfies [Eq. \(9\)](#)

3. Theorem (Structure). *Let $X = (X_1, \dots, X_n)$ be a centered Gaussian vector with $\mathbb{E} X_i^2 \leq 1$ for all $i \in [n]$. Suppose $d(i, j) = \mathbb{E}(X_i - X_j)^2$ satisfies the inequalities [Eq. \(4\)](#) and [Eq. \(9\)](#). Then, there exists sets $A, B \subseteq [n]$ with $|A|, |B| \geq \Omega(n)$ such that*

$$\min_{i \in A, j \in B} d(i, j) \geq \Omega\left(1/\sqrt{\log n}\right). \quad (11)$$

Furthermore, we can find such sets A and B in polynomial time.

It will be useful to rephrase the above theorem in terms of vertex separators in certain graphs associated with the Gaussian vector X . For $\Delta > 0$, let H be the directed graph with vertex set $[n]$ such that

$$E(H) = \{(i, j) \in [n]^2 \mid d(i, j) \leq \Delta\}. \quad (12)$$

We say that subsets $A, B \subseteq [n]$ form a *vertex separator* if no edge of H goes directly between A and B , i.e., $E(H) \cap A \times B = \emptyset$. We say such a vertex separator is *good* if $|A| \cdot |B| \geq \Omega(n^2)$. With this terminology in place, we see that in order to prove [Theorem 3](#) it is enough to show that there exists $\Delta \geq \Omega(1/\sqrt{\log n})$ such that H has a good vertex separator.

Proof of structure theorem

Let $X = (X_1, \dots, X_n)$ be a centered Gaussian vector with $\mathbb{E} X_i^2 \leq 1$ for all $i \in [n]$. Let $d(i, j) = \mathbb{E}(X_i - X_j)^2$ for all $i, j \in [n]$. Suppose that d satisfies [Eq. \(4\)](#) and [Eq. \(9\)](#). For $\Delta > 0$, let H be the graph with vertex set $[n]$ and edge set [Eq. \(12\)](#).

We consider the following randomized algorithm to find vertex separators for H :

⁵ The definition of expansion [Eq. \(2\)](#) is symmetric in the sense that both x and its complement $\mathbf{1} - x$ have the same value. Therefore, we can symmetrize the pseudo-distribution in the same way as for Max Cut.

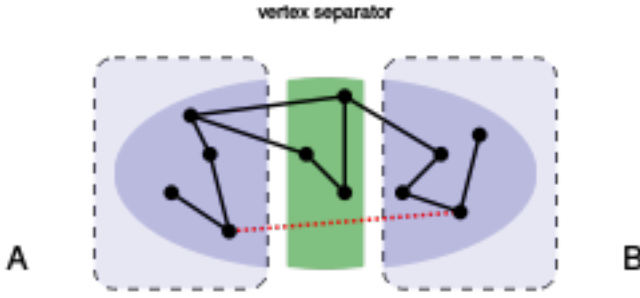


Figure 1: In this graph, the set of vertices in the middle is a vertex separator. Removing this set breaks the graph into two parts A and B that are not connected by an edge. If we added, the red edge to the graph, the set of vertices in the middle would no longer be a vertex separator.

1. Choose subsets $A^0, B^0 \subseteq [n]$ such that

$$A^0 = \{i \mid X_i \leq -1\} \text{ and } B^0 = \{j \mid X_j \geq 1\}. \quad (13)$$

2. Find a maximal matching M in the edges $E(H) \cap A^0 \times B^0$.
3. Output the sets $A = A^0 \setminus V(M)$ and $B = B^0 \setminus V(M)$.

Note that the objects constructed by the algorithm, in particular the sets A and B and the matching M , are jointly distributed with the Gaussian vector X . We emphasize that the edges of the matching M are directed. In particular, the event $(i, j) \in E$ implies that $X_j - X_i \geq 2$.

We observe that the output (A, B) of the algorithm forms a vertex separator in H . Indeed, by the maximality of M , every edge $(i, j) \in E(H) \cap A^0 \times B^0$ has at least one endpoint in $V(M)$, which means that (i, j) is not contained in $A \times B$.

We also observe that if we find the maximal matching in step 2 by greedily passing over the edges in a canonical order independent of the randomness of X , then for every $x \in \mathbb{R}^n$, the matching produced in the event $X = x$ is the reverse of the matching produced in the event $X = -x$. It follows that for every vertex $i \in [n]$,

$$\mathbb{P}\{i \text{ has outgoing edge in } M\} = \mathbb{P}\{i \text{ has incoming edge in } M\}. \quad (14)$$

The following lemma characterizes the success of the above algorithm in terms of the expected size of M .

4. Lemma (Vertex separator or large matching). *Either the above algorithm outputs a good vertex separator for the graph H with probability $\Omega(1)$, or the random matching M produced by the algorithm satisfies $\mathbb{E}|M| \geq \Omega(n)$.*

Proof. A similar analysis as in the Goemans–Williamson algorithm for Max Cut shows that there exists an absolute constant $c > 0$ such that $\mathbb{P}\{X_i \leq -1 \text{ and } X_j \geq 1\} \geq c \cdot d(i, j)$ for every $i, j \in [n]$. Thus, by inequality Eq. (9),

$$\mathbb{E}|A^0| \cdot |B^0| \geq 0.1c \cdot n^2. \quad (15)$$

Therefore,

$$\mathbb{E}|A| \cdot |B| \geq \mathbb{E}|A^0| \cdot |B^0| - n \cdot \mathbb{E}|M| \geq 0.01c \cdot n^2 - n \cdot \mathbb{E}|M|. \quad (16)$$

It follows that $\mathbb{E}|M| \geq \frac{1}{2} \cdot 0.1c \cdot \mathbb{E}|M|$. (Otherwise, $\mathbb{E}|A| \cdot |B| \geq \Omega(n^2)$, which would mean that the algorithm outputs a good separator for H with probability $\Omega(1)$, contradicting the premise of the lemma.) \square

We will analyze the expected size of the random matching M by relating it to the maximum of a Gaussian process.⁶

5. Lemma (Maximum of Gaussian process). *The expected maximum of the Gaussian process $\{X_j - X_i \mid i, j \in [n]\}$ satisfies*

$$\frac{\Omega(1)}{\Delta} \cdot (\mathbb{E}|M|/n)^3 \leq \mathbb{E} \max_{i, j \in [n]} X_j - X_i \leq \sqrt{2 \log n}. \quad (17)$$

Together Lemma 4 and Lemma 5 allow us to prove Theorem 3.

Proof of Theorem 3. We prove the contrapositive: if the above algorithm fails to find a good vertex separator in H , then the distance parameter Δ of the graph satisfies $\Delta \geq \Omega(1/\sqrt{\log n})$.⁷ Let M be the random matching defined by the algorithm. Since the algorithm fails to find a good vertex separator Lemma 4 gives a lower bound on the expected size of the matching, $\mathbb{E}|M|/n \geq \Omega(1)$. On the other hand, Lemma 5 allows us to upper bound the same quantity $\mathbb{E}|M|/n \leq O(\Delta \sqrt{\log n})^{1/3}$. We conclude $\Delta \cdot \sqrt{\log n} \geq \Omega(1)$, which gives the desired lower bound $\Delta \geq \Omega(1)/\sqrt{\log n}$. \square

It remains to prove Lemma 5. The proof considers the following family of Gaussian processes and relates their expected maxima in terms of the expected size of the matching M . For $i \in [n]$ and $k \in \mathbb{N}$, let $Y_i^{(k)}$ be the maximum of the Gaussian process $\{X_j - X_i \mid j \in H^k(i)\}$,

$$Y_i^{(k)} = \max_{j \in H^k(i)} X_j - X_i. \quad (18)$$

Here, $H^k(i)$ denotes the set of vertices that can be reached from i by at most k steps in the graph H . Define the following potential function,

$$\Phi(k) = \sum_{i=1}^n \mathbb{E} Y_i^{(k)}. \quad (19)$$

⁶ We remark that the lemma only uses the triangle inequalities Eq. (4) but not inequality Eq. (9).

⁷ That means that if we choose Δ a sufficiently small constant factor times $1/\sqrt{\log n}$, the algorithm will succeed in finding a vertex separator for the corresponding graph H .

Note that $\Phi(k)$ allows us to lower bound $\mathbb{E} \max_{i,j \in [n]} X_j - X_i$ because $\Phi(k)/n \leq \mathbb{E} \max_{i,j \in [n]} X_j - X_i$.

The following lemma is the key ingredient for the proof of [Lemma 5](#).⁸

6. Lemma (chaining). *For every $k \in \mathbb{N}$,*

$$\Phi(k+1) \geq \Phi(k) + \mathbb{E}|M| - O(n) \cdot \max_{i \in [n], j \in H^{k+1}(i)} \left(\mathbb{E}(X_i - X_j)^2 \right)^{1/2}. \quad (20)$$

This chaining lemma directly implies [Lemma 5](#).

Proof of Lemma 5. By the triangle inequality [Eq. \(4\)](#), $\mathbb{E}(X_i - X_j)^2 \leq k \cdot \Delta$ for all $i \in [n]$ and $j \in H^k(i)$. By [Lemma 6](#), it follows that there exists an absolute constant $C \geq 1$ such that for every $k \in \mathbb{N}$, $\Phi(k+1) \geq \Phi(k) + \mathbb{E}|M| - Cn \cdot \sqrt{k\Delta}$. For all $k \leq k_0 = (1/4C^2) \cdot (\mathbb{E}|M|/n)^2/\Delta$, this inequality simplifies to

$$\Phi(k+1) \geq \Phi(k) + \frac{1}{2} \mathbb{E}|M|. \quad (21)$$

If we unroll this recurrence relation, we obtain the desired lower bound

$$\mathbb{E} \max_{i,j \in [n]} X_j - X_i \geq \Phi(k_0+1)/n \geq \frac{1}{2} k_0 (\mathbb{E}|M|/n) = (1/8C^2) \cdot (\mathbb{E}|M|/n)^3/\Delta. \quad (22)$$

□

The proof of [Lemma 6](#) uses the following remarkable fact about the variance of maxima of Gaussian processes: The *variance* of the maximum of a collection of Gaussian variables is bounded by the maximum variance of a variable in the collection.⁹

7. Theorem (Variance of maxima of Gaussian processes). *Let $Z = (Z_1, \dots, Z_t)$ be centered Gaussian variables. Then, the variance of the maximum of Z_1, \dots, Z_t is bounded by*

$$\mathbb{V}[\max\{Z_1, \dots, Z_t\}] \leq O(1) \cdot \max\{\mathbb{V}[Z_1], \dots, \mathbb{V}[Z_t]\}. \quad (23)$$

The proof of this inequality also shows Gaussian-like concentration around the mean. In that form it is sometimes called *Borell's inequality* and follows from Gaussian concentration inequalities for Lipschitz functions ([Sudakov and Cirel'son \[1974\]](#); [Borell \[1975\]](#)).¹⁰

Proof of Lemma 6. The starting point of the proof is the following inequality among random variables for all $(i, j) \in E(H)$,

$$Y_i^{(k+1)} \geq Y_j^{(k)} + X_j - X_i. \quad (24)$$

⁸ We remark that the proof of this lemma neither uses the triangle inequalities [Eq. \(4\)](#) nor the inequality [Eq. \(9\)](#).

⁹ In contrast, the *expectation* of the maximum typically depends also on the number of variables in the collection.

¹⁰ See [this book chapter](#) by Pollard for different proofs of this concentration inequality.

The inequality holds because $H^{k+1}(i) \supseteq H^k(j)$. Hence, if $Y_j^{(k)} = X_h - X_j$ for some $h \in H^k(j)$, then $Y_i^{(k+1)} \geq X_h - X_i = Y_j^{(k)} + X_j - X_i$.

Let $N \subseteq [n] \times [n]$ be an arbitrary matching of the vertices not in matching M . Since matching edges $(i, j) \in M$ satisfy $X_j - X_i \geq 2$, we can derive the following set of $n/2$ inequalities,

$$\forall (i, j) \in M. Y_i^{(k+1)} \geq Y_j^{(k)} + 2 \text{ and } \forall (i, j) \in N. \frac{1}{2}Y_i^{(k+1)} + \frac{1}{2}Y_j^{(k+1)} \geq \frac{1}{2}Y_i^{(k)} + \frac{1}{2}Y_j^{(k)}. \quad (25)$$

If we sum up these $n/2$ inequalities, we obtain the inequality

$$\sum_{i=1}^n Y_i^{(k+1)} \cdot L_i \geq \sum_{j=1}^n Y_j^{(k)} \cdot R_j + 2|M|. \quad (26)$$

Here, $L_i = 1$ if i has an outgoing edge in M , $L_i = 0$ if i has an incoming edge in M , and $L_i = 1/2$ if i is not matched in M . Similarly, $R_j = 1$ if j has an incoming edge in M , $R_j = 0$ if j has an outgoing edge in M , and $R_j = 1/2$ if j is not matched in M . Since every vertex is equally likely to have an incoming edge and an outgoing edge in M (see discussion after the description of the algorithm), it follows that $\mathbb{E} L_i = \mathbb{E} R_j = 1/2$ for all $i, j \in [n]$. Using a general bound on the variance of maxima of Gaussian processes,¹¹

$$\left| \mathbb{E} Y_i^{(k+1)} \cdot L_i - \mathbb{E} Y_i^{(k+1)} \cdot \mathbb{E} L_i \right| \leq \sqrt{\mathbb{V}[Y_i^{(k+1)}]} \cdot \mathbb{V}[L_i] \leq O(1) \cdot \max_{j \in H^{k+1}(i)} \left(\mathbb{E}(X_i - X_j)^2 \right)^{1/2}. \quad (27)$$

¹¹ Here, we use the general inequality $|\mathbb{E} XY - \mathbb{E} X \cdot \mathbb{E} Y| \leq \sqrt{\mathbb{V}X \cdot \mathbb{V}Y}$, which we leave as an exercise to the reader.

Similarly,

$$\left| \mathbb{E} Y_j^{(k)} \cdot R_j - \mathbb{E} Y_j^{(k)} \cdot \mathbb{E} R_j \right| \leq \sqrt{\mathbb{V}[Y_j^{(k)}]} \cdot \mathbb{V}[R_j] \leq O(1) \cdot \max_{i \in H^k(j)} \left(\mathbb{E}(X_i - X_j)^2 \right)^{1/2}. \quad (28)$$

By taking the expectation of inequality Eq. (26) and applying the above deviation bounds, it follows as desired that

$$\sum_{i=1}^n \mathbb{E} Y_i^{(k+1)} \geq \sum_{j=1}^n \mathbb{E} Y_j^{(k)} + 4|M| - O(n) \cdot \max_{i \in [n], j \in H^{k+1}(i)} \left(\mathbb{E}(X_i - X_j)^2 \right)^{1/2}. \quad (29)$$

□

Reduction to the well-spread case

We describe now how to prove [Theorem 1](#) in case inequality [Eq. \(9\)](#) is not satisfied.

Recall the distance $d(i, j) = \mathbb{E}_{\mu(x)}(x_i - x_j)^2$. Let $\delta > 0$ be the average distance

$$\delta = \frac{1}{n^2} \sum_{i, j \in [n]} d(i, j). \quad (30)$$

The following lemma shows that if there exists a cluster of radius significantly smaller than δ that contains a constant fraction of vertices, then the condition of [Lemma 2](#) is satisfied for $\Delta \geq \Omega(1)$ (which means that we get a $O(1)$ -approximation).

8. Lemma (Heavy cluster). *Suppose there exists $i \in [n]$ such that $A = \{j \in [n] \mid d(i, j) \leq \delta/4\}$ satisfies $|A| \geq \Omega(n)$. Then, $|A| \cdot \sum_{j=1}^n d(j, A) \geq \Omega(1) \cdot \sum_{i, j \in [n]} d(i, j)$.*

Proof. By the triangle inequality,

$$\delta \leq \frac{1}{n^2} \cdot \sum_{i, j \in [n]} d(i, A) + d(j, A) + \delta/2 \leq \delta/2 + \frac{1}{n} \sum_{i=1}^n 2d(i, A). \quad (31)$$

Thus, $|A| \sum_{j=1}^n d(j, A) \geq n \cdot \delta/4 \cdot |A| \geq \Omega(1) \cdot \delta n^2$, which is the desired bound. \square

The following lemma shows that if the condition of the previous lemma is not satisfied, then after restricting to a constant fraction of the vertices there are Gaussian random variables that we can apply the structure theorem to in order to obtain sets $A, B \subseteq [n]$ with $|A|, |B| \geq \Omega(n)$ such that $d(i, j) \geq \Omega(1/\sqrt{\log n}) \cdot \delta$ for all $i \in A$ and $j \in B$.

9. Lemma (Heavy cluster or well-spread). *Either there exists a vertex $i \in [n]$ that satisfies the condition of [Lemma 8](#) or there exists a subset $U \subseteq [n]$ with $|U| \geq \Omega(n)$ and a centered Gaussian vector $Y = (Y_1, \dots, Y_n)$ such that*

- $\mathbb{E} Y_i^2 \leq 1$ for all $i \in U$,
- $\sum_{i, j \in U} \mathbb{E}(Y_i - Y_j)^2 \geq 0.05$,
- there exists $\alpha \geq \Omega(\delta)$ such every $i, j \in U$ satisfies $d(i, j) = \alpha \cdot \mathbb{E}(Y_i - Y_j)^2$.

Proof. Suppose that no vertex $i \in [n]$ satisfies the condition of [Lemma 8](#). So every $i \in [n]$ satisfies $|B(i, \delta/4)| \geq 0.1n$, where $B(i, \delta/4)$ is the set of vertices $j \in [n]$ such that $d(i, j) \leq \delta/4$. Let $k \in [n]$ be such that $\sum_{i=1}^n d(i, k) \leq \delta n$. Let $U \subseteq [n]$ be the set of vertices $i \in [n]$ such that $d(i, k) \leq 2\delta$. By Markov's inequality, $|U| \geq n/2$. Then,

$$\sum_{i, j \in U} d(i, j) \geq \sum_{i \in [n]} \delta/4 \cdot |U \setminus B(i, \delta/4)| \geq n \cdot \delta/4 \cdot 0.4n = 0.1\delta n. \quad (32)$$

By symmetry, we may assume $\tilde{\mathbb{E}}_{\mu(x)} x = \frac{1}{2} \cdot \mathbf{1}$. Therefore, if we let X_1, \dots, X_n be the Gaussian variables with the same first two moments as the pseudo-distribution μ . Then, the variables Y_1, \dots, Y_n with $Y_i = (X_i - X_k)/\sqrt{2\delta}$ satisfy $\mathbb{E} Y_i^2 \leq 1$ for all $i \in [n]$, $\sum_{i, j \in U} \mathbb{E}(Y_i - Y_j)^2 \geq 0.05$ and $\mathbb{E}(Y_i - Y_j)^2 = d(i, j)/2\delta$. \square

References

- Sanjeev Arora, Satish Rao, and Umesh V. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *STOC*, pages 222–231. ACM, 2004.
- Christer Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975. ISSN 0020-9910.
- Frank Thomson Leighton and Satish Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *FOCS*, pages 422–431. IEEE Computer Society, 1988.
- Thomas Rothvoss. Lecture notes on the ARV algorithm for sparsest cut. *CoRR*, abs/1607.00854, 2016.
- V. N. Sudakov and B. S. Cirel'son. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 41:14–24, 165, 1974. Problems in the theory of probability distributions, II.