CSC2414 - Metric Embeddings*
Lecture 10: ARV $O(\sqrt{\log n})$ approximation of Sparsest Cut

Notes taken by Igor Naverniouk

Summary: We describe a $O(\sqrt{\log n})$ approximation algorithm for the Sparsest Cut problem, due to Arora, Rao and Umesh Vazirani [ARV04]. The algorithm uses Semidefinite programming and its analysis relates to notions of distortion, or more accurately average distortion of a negative type metric into ℓ_1.

1 Overview

1. We construct a Semidefinite Program (SDP) for the Sparsest Cut problem. Its solution can be viewed as an ℓ_2^2 metric in the usual way ($d(i,j) = \|v_i - v_j\|^2$).

2. We show that if a solution embeds into ℓ_1 with average distortion D, then the integrality gap of the SDP is at most D, and we can find a cut which is a $2D$-approximation to the optimal solution. In order to prove that such an embedding exists, we prove the Main Structure Theorem (MST):

3. The Main Structure Theorem (MST) says that given a metric space $X \subset \mathbb{B}^m(0,1)$ with $|X| = n$ and $\frac{1}{n^2}\sum \|x_i - x_j\|^2 \geq \beta > 0$, will find a pair of subsets $S, T \subseteq X$, such that $|S|, |T| = \Omega(n)$ and

$$d(S,T) = \Delta = \Omega\left(\frac{1}{\sqrt{\log n}}\right).$$

   Trivially the MST implies that $d \hookrightarrow \ell_1$ with average distortion $O(\sqrt{\log n})$: Map every $x \in X$ to $d(x, S)$. Thus the embedding is in fact into the line $\mathbb{R}$.

4. To prove MST we describe an algorithm to find the desired sets S and T. It is a probabilistic algorithm that makes use of the *alternation* method. Specifically, the algorithm consists of two phases:

   - Phase 1: randomly pick a hyperplane, and separate X to two sets of points that are relatively far ($\Theta(1/\log n)$) from the hyperplane, and are on different side of it.

* Lecture Notes for a course given by Avner Magen, Dept. of Computer Sciecne, University of Toronto.

- Phase 2: Repeatedly remove pairs of points that are in different halves, but are too close to each other (closer than $\frac{O(1}{\sqrt{logn)}}$).

At the end, we will get two sets that have the desired separation, by construction. The hard part will be to prove that both have $\Omega(n)$ points.

5. Much of the discussion will revolve around that difficult part. A very high-level sketch is as follows: if we assume that $|S| = o(n)$ or $|T| = o(n)$, then we will get an impossible geometric configuration. Namely, we will get a set Y, which is a "core", and the contradiction will come from an application of a theorem of Lee [Lee05].

2 Semidefinite Programming for the Sparsest Cut problem

The Sparsest Cut of a graph G, also called the *edge expansion* of G is defined as

$$SC(G) = \min \left\{ \frac{|E(S, \overline{S})|}{|S|} : S \in V, |S| \leq \frac{n}{2} \right\},$$

where $E(A, B)$ is the number of edges crossing the cut $\{A, B\}$. The following algorithm is due to Arora, Rao and U.Vazirani [ARV04].

Define

$$\eta(G) = \min_{d=\delta_S} \left\{ \frac{\sum_{ij \in E} d(i,j)}{\sum_{i,j} d(i,j)} \right\},$$

where δ_S is the cut metric corresponded to $S \subseteq V$, and E is the set of edges of G. From last time, recall that

$$SC(G) \leq n\eta(G) \leq 2SC(G)$$

because

$$\eta(G) = \frac{|E(S, \overline{S})|}{|S||\overline{S}|}.$$

In the previous class, we relaxed the condition of d being a cut metric to a more general condition of d being an ℓ_1 metric with no loss. Then we relaxed it to a strictly more general condition of d being any metric and built a Linear Program using all n^3 triangle inequalities. Then by Bourgain's theorem we found a $O(\log n)$-distortion embedding to ℓ_1, which gave us a $O(\log n)$ approximation algorithm for Sparsest Cut.

Now, we can write $\eta(G)$ as

$$\eta(G) = \min_{x_i \in \{-1,1\}} \frac{\sum_{ij \in E} (x_i - x_j)^2}{\sum_{i,j} (x_i - x_j)^2}.$$

This is still NP-hard to solve, but look instead at

$$\eta^*(G) = \min_{x_i \in \mathbb{R}^n} \frac{\sum_{ij \in E} \|x_i - x_j\|^2}{\sum_{i,j} \|x_i - x_j\|^2}.$$

This is a relaxation of the x_i's from integers in the set $\{-1,1\}$ to n-dimensional vectors. As as a result, we get a Semidefinite Program (SDP):

$$\min \textstyle\sum_{ij\in E} \|x_i - x_j\|^2,$$
$$s.t. \textstyle\sum_{i,j} \|x_i - x_j\|^2 = 1.$$

Contrast this set-up with the one before, where we relaxed cut metrics to general metrics and got a Linear Program. Now we relax cut metric to ℓ_2^2 distances and get a Semidefinite Program. Note that ℓ_2^2 distances do not necessarily define a metric. (Take, for example, 3 points on a line.)

Next, we combine both ideas and create an SDP^+ by adding all n^3 triangle inequalities:

$$\forall i,j,k : \|x_i - x_j\|^2 + \|x_j - x_k\|^2 \geq \|x_i - x_k\|^2.$$

This causes the solutions to be restricted to ℓ_2^2 metrics. From now on we refer to $\eta^*(G)$ to the solution to the above SDP in the presence of triangle inequalities.

3 Geometry of ℓ_2^2 metrics

What is the geometry of this SDP^+? When does a set of points in $\mathbb{R}^n$ define an ℓ_2^2 metric?

Consider 3 points on a line with ℓ_2 distances a, b and $c = a + b$. Then their ℓ_2^2 distances are a^2, b^2 and $a^2 + b^2 + 2ab$, respectively. Assuming that $a > 0$ and $b > 0$, this 3-point set violates the ℓ_2^2 triangle inequality and is thus not a metric.

To see which point sets do define ℓ_2^2 metrics, take an arbitrary triangle. Then, by the law of cosines,

$$c^2 = a^2 + b^2 - 2ab\cos\gamma,$$

and we have $c^2 \leq a^2 + b^2$ when $\cos\gamma \geq 0$, *i*.e. $0 \leq \gamma \leq \frac{\pi}{2}$. This has to hold for all triples of points in the set X if (X, ℓ_2^2) is to be a metric space.

Danzer and Grünbaum [DG62] showed that the maximum number of points in $\mathbb{R}^n$ without obtuse angles between them is 2^n and is realised by a hypercube.

4 Application to Sparsest Cut

What can we say about the integrality gap? If for any ℓ_2^2 metric d, $d \overset{D}{\hookrightarrow} \ell_1$, then the integrality gap is

$$IG = \frac{\eta(G)}{\eta^*(G)} \leq D.$$

From Bourgain's theorem we know that D can be taken to be equal to $O(\log n)$. Can we do better using the fact that the metric is ℓ_2^2?

Definition 4.1. Let $d \hookrightarrow d'$ s.t. $\forall i,j : d'(i,j) \leq d(i,j)$ (a non-expanding embedding). Then the *a*verage distortion is

$$\frac{\sum_{i,j} d(i,j)}{\sum_{i,j} d'(i,j)}.$$

Claim 4.2. *If every ℓ_2^2 metric embeds into ℓ_1 with average distortion D, then*

$$\frac{\eta(G)}{\eta^*(G)} \leq D.$$

If we use a polynomial number of dimensions in the embedding into ℓ_1, then we can efficiently find a cut with cost at most $D\eta^*(G)$.

Proof. Given G, solve the SDP^* and get a metric $d \in \ell_2^2$ s.t.

$$\sum_{ij \in E} d(i,j) = \eta^*(G).$$

Let $d' \in \ell_1$ be s.t.

- $\forall i,j : d'(i,j) \leq d(i,j)$,
- $\sum_{i,j} d'(i,j) \geq \frac{1}{D} \sum_{i,j} d(i,j)$.

Then

$$Z \stackrel{def}{=} \frac{\sum_{ij \in E} d'(i,j)}{\sum_{i,j} d'(i,j)} \leq D \frac{\sum_{ij \in E} d(i,j)}{\sum_{i,j} d(i,j)} = D\eta^*(G).$$

If d' has a polynomial number, m, of dimensions, then we can split it into mn cuts on which it is supported. An algorithm A that checks these cuts will get a cut of value at least Z, which is at least $D\eta^*(G)$. □

We have n points on the unit sphere in m dimensions with $\sum_{i,j} \|x_i - x_j\|_2^2 = n^2$. We want two linear-size sets S and T s.t.

$$d_{\ell_2^2}(S,T) = \min_{i \in S, j \in T} \|x_i - x_j\|^2 \geq \Delta,$$

We will show that we can have $\Delta = \Omega((\log n)^{-1/2})$.

Theorem 4.3. Main Structure Theorem. *Let X be a set of n points on the unit sphere with a metric d, such that*

$$\frac{1}{n^2} \sum_{i,j} d(i,j) = \frac{1}{n^2} \sum_{i,j} \|x_i - x_j\|^2 \geq \gamma > 0.$$

Then there exists a partition of X into S and T, such that

- $|S| = \Omega(n)$, $|T| = \Omega(n)$;
- $d(S,T) = \Omega((\log n)^{-1/2})$.

Proof. To come later... □

First, a few remarks.

- It is enough to assume that the diameter of $X \leq 1$.

- The condition
$$\frac{1}{n^2}\sum_{i,j} d(i,j) \geq \gamma > 0$$
is essential. Otherwise, we could put everything into one point.

- The condition on the diameter of X being less than 1 is essential in the proof. (why?)

- Without the triangle inequality, we can only get a bound of $\Delta = (\log n)^{-1}$.

- The hypercube, normalized to fit into the unit sphere, makes the Main Structure Theorem tight, so we cannot hope to get a better separation. For example take two Hamming balls around two opposite corners of the hypercube. Isoperimetric inequality says that they will give the best separation we can get.

Assuming that the theorem is true, we will get a good average distortion. To get an approximation algorithm for the Sparsest Cut problem, we start with a general ℓ_2^2 metric (not one that lives on the unit sphere).

Claim 4.4. *The Main Structure Theorem implies a $O(\sqrt{\log n})$ approximation algorithm for Sparsest Cut.*

Proof. We know that $\frac{1}{n^2}\sum d(i,j) = 1$.

Case 1: There exists a radius $\frac{1}{4}$ ball of size $\geq \frac{n}{4}$. Call it L. Take a Fréchet embedding with respect to L, *i*.e. take $f : X \to L$, $f(x) = d(x, L)$. We get a metric; call it d'. Then

$$\sum d'(i,j) \geq \sum_{\substack{i \notin L \\ j \in L}} d'(i,j) \geq |L| \sum_i d(i,L) \geq \frac{n}{4}\sum_i d(i,L).$$

$$n^2 = \sum_{i,j} d(i,j) \leq \sum_{i,j} d(i,k) + d(k,j),$$

where k is the center of L. So

$$n^2 \leq 2n \sum_i d(i,k) \leq 2n \sum_i \left(d(i,L) + \frac{1}{4}\right),$$

and so $\sum d(i,L) \geq \frac{n}{4}$. Therefore,

$$\sum d'(i,j) \geq \frac{n^2}{16}.$$

Case 2: There is no ball of radius $\frac{1}{4}$ containing at least $\frac{n}{4}$ elements. Notice, that it is always true that for every point l, $|B(l,2)| \geq n/2$ (else the average of d is bigger than 1). We now claim that the average distsance among points in $B(l,2)$ is constant. Indeed, for every point, at least $n/2 - n/4$ points are of distance at least $n/4$ ensuring an average of at least $1/16$ in $B(l,2)$. Scaling down this set by $\sqrt{2}$ we get a set of points in the unit ball with average distance at least $1/32$. We can now apply the M.S.T to the set of points $B(l,2)$ and get sets S and T. (The proof that the Main Structure Theorem holds even when the points are inside the unit ball, and not just on the unit sphere, is given as a question on one of the homework assignments.) Next, we construct a Fréchet embedding with respect to S. This gives us a distance function

$$d'(i,j) = |d(i,S) - d(j,S)|.$$

By the properties of Fréchet embeddings, $d' \leq d$ and

$$\begin{aligned}
\sum_{i,j} d'(i,j) &\geq \sum_{i \in S, j \in T} d'(i,j) \\
&\geq |S||T| d(S,T) \\
&= \Omega\left(\frac{n^2}{\sqrt{\log n}}\right) \\
&= \sum_{i,j} d(i,j) \Omega\left(\frac{1}{\sqrt{\log n}}\right).
\end{aligned}$$

So the distortion is $D = O(\sqrt{\log n})$.

$\square$

5 The Algorithm for proving MST

We have a unit ball with a set of points X on its surface. The average distances between the points of X are large, and there are no obtuse angles. We know that $X \subseteq S^{m-1}$, and that

$$\frac{1}{n^2} \sum_{i,j} \|x_i - x_j\|^2 \geq \beta > 0.$$

The algorithm consists of two phases.

- **Phase 1**: Pick a uniformly random direction (a vector in S^{m-1}) and define the sets

$$\begin{aligned}
S_u &\stackrel{def}{=} \{x_i : \langle x_i, u \rangle \geq \frac{\sigma}{\sqrt{m}}\}, \\
T_u &\stackrel{def}{=} \{x_i : \langle x_i, u \rangle \leq -\frac{\sigma}{\sqrt{m}}\}, \\
R_u &\stackrel{def}{=} \{x_i : |\langle x_i, u \rangle| < \frac{\sigma}{\sqrt{m}}\} = X \backslash S_u \backslash T_u,
\end{aligned}$$

  where σ is a constant to be picked later.

- **Phase 2**: As long as there exist pairs of points $x \in S_u, y \in T_u$, s.t. $\|x - y\|^2 < \Delta$, remove them. The order of removal is unimportant.

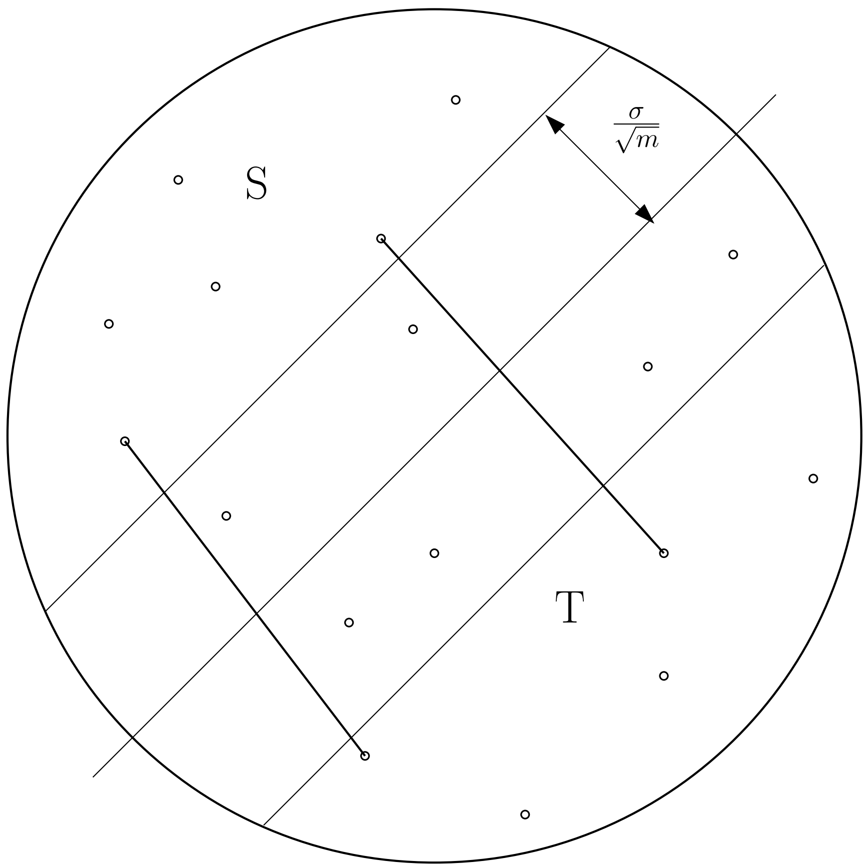

Figure 1: Separating a sphere by a hyperplane with a margin. Pairs of points that are too close to the margin are removed in Phase 2 of the algorithm.

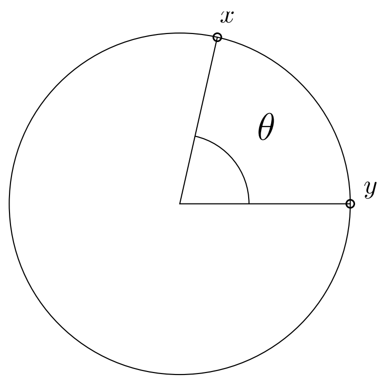

Figure 2: Two points, x and y, on the surface of a sphere, separated by an angle of θ.

Finally, call what is remained in S_u and T_u the required sets S and T.

By construction, the separation between S and T (assuming they are non-empty) is at least Δ in the ℓ_2^2 norm. The only claim left to prove is that S and T have size that is at least linear in n.

First, look at the two half-balls

$$\tilde{S_u} = \{x_i | \langle x_i, u \rangle \geq 0\},$$
$$\tilde{T_u} = \{x_i | \langle x_i, u \rangle \leq 0\},$$

and define two events:

$$E_1 = \left\{ u : |\tilde{S_u}| \geq \frac{\gamma n}{2} \wedge |\tilde{T_u}| \geq \frac{\gamma n}{2} \right\},$$
$$E_2 = \left\{ u : |R_u| \leq \frac{\gamma n}{4} \right\}.$$

We want to show that $p \stackrel{def}{=} \Pr[E_1 \cap E_2]$ is a positive constant.

Note that if $x, y \in \mathbb{R}^2$ are on a circle and are separated by an angle of θ, then the probability of separating them by a line passing through the center of the circle is $\frac{\theta}{\pi}$ (see Figure 5). The same obvious fact holds in m dimensions, by symmetry.

Hence,

$$\Pr[\langle x, u \rangle \langle y, u \rangle < 0] = \frac{\theta}{\pi}.$$
$$\|x - y\|^2 = 2(1 - \langle x, y \rangle) = 2(1 - \cos\theta).$$

Note:

$$\forall \theta : 2\frac{\theta}{1 - \cos\theta} \geq 0.878.$$

So if E_s is the event that u separates x from y, then

$$\frac{\Pr[E_s]}{\|x - y\|^2} \geq \frac{0.878}{4}.$$

To extend the claim to a set of points inside the unit ball instead of the unit sphere, we can simply move the points away from the origin (see homework).

Let $\nVdash_{x,y}$ be the indicator function of E_s in the probability space of u. Then

$$\begin{aligned} E[\sum_{x,y} \nVdash_{x,y}] &= \sum_{x,y} \Pr[E_s] \\ &\geq \sum_{x,y} \|x-y\|^2 \frac{4}{0.878} \\ &\geq \gamma n^2. \end{aligned}$$

So

$$\Pr[\sum_{x,y} \nVdash_{x,y} \leq \frac{\gamma n^2}{2}] \leq \frac{1-\gamma}{1-\frac{\gamma}{2}} \leq 1-\frac{\gamma}{2}.$$

Therefore, with probability $\frac{\gamma}{2}$,

$$|\tilde{S_u}||\tilde{T_u}| \geq \frac{\gamma n^2}{2},$$

so

$$|\tilde{S_u}|, |\tilde{T_u}| \geq \frac{\gamma n}{2}$$

because $|\tilde{S_u}| \leq n$ and $|\tilde{T_u}| \leq n$.

This gives us a bound on the probability of E_1. Now, we need to approximate the probability of E_2.

Lemma 5.1. *If v is a unit vector in $\mathbb{R}^m$ and u is a random unit vector (by the Haar measure) then*

- $\Pr[|\langle u, v\rangle| < \frac{x}{\sqrt{m}}] \leq 3x.$
- $\Pr[|\langle u, v\rangle| \geq \frac{x}{\sqrt{m}}] \leq \exp(-x^2/4).$

Proof. This follows from the Gaussian-like behavior of projections. □

Note that this implies that $\Pr[x \in R_u] \leq 3\sigma$ and

$$Pr\left[|R_u| > \frac{3\sigma n}{\gamma/4}\right] \leq \frac{\gamma}{4}.$$

So if $\sigma = \frac{1}{3}\gamma^2$, we get

$$\Pr[\overline{E_2}] = Pr\left[|R_u| > \frac{\gamma n}{4}\right] < \frac{\gamma}{4}.$$

Now we have shown that $|\tilde{S_u}|$ and $|\tilde{T_u}|$ are linear in n, and that R_u is small. Since $S_u \subset \tilde{S_u}$ and $T_u \subset \tilde{T_u}$, we get that $|S|$ and $|T|$ must be $\Omega(n)$.

What remains is to analyze Phase 2 of the algorithm. We need to show that the number of points removed from S_u and T_u is small.

What can we say if x and y were removed by u? Let's call $M(u)$ the set of pairs removed when the separating hyperplane is u. Also let $l = \sqrt{\Delta}$, that is the **Euclidean** needed separation. For $(x, y) \in M(u)$ we have that

- $\|x - y\| \leq l \stackrel{def}{=} \sqrt{\Delta}$.
- $|\langle x - y, u\rangle| \geq \frac{2\sigma}{\sqrt{m}}$.

The first condition follow from that the pair was too close to be left untouched, while the second follow from the fact that the two points were on different side of the "fat" cut. Here is a major observation. We expect $|\langle x - y, u\rangle|$ to be about $\frac{\|x-y\|}{\sqrt{m}}$, in general. But here, we have

$$|\langle x - y, u\rangle| \geq \frac{\|x - y\|}{\sqrt{m}} \cdot \frac{2\sigma}{l}.$$

So we say that $x - y$ has a "stretch" of $\frac{2\sigma}{l}$ with respect to u, which precisely means that this is the factor by which its projection is larger (asymptotically) from the expected length.

Let $K_u(x, y)$ be the event that the pair $\{x, y\}$ is removed by u. Then

$$\Pr[K_u(x, y)] = \exp\left(-\Omega((2\sigma/l)^2)\right).$$

If we required the "modest" separation of $l = O\left(\frac{1}{\sqrt{\log n}}\right)$, then that would be easy

$$\Pr[K_u(x, y)] = \exp(-\Omega(\log n)) = o(1/n)$$

But with $l = \Theta((\log n)^{-1/4})$, we get

$$\Pr[K_u(x, y)] = e^{-\Omega(\sqrt{\log n})} = \Omega(1).$$

What does this tell us? that the expected number of pairs that are locally (per the above conditions) candidate for deletions is large. This means that we have no choice but to understand the stochastic behavior of $|M_u|$ which is, recall, a matching. In a way we learn that a global approach is inventible.

We would like to show that

$$q = \Pr[|M_u| \geq \gamma n/8]$$

is small, more precisely that $q = o(1)$. If this is the case we are happy as then with constant probability phase one is successful, and to say that on top of that the size of the matching is small enough to leave linear size sets, doesn't "cost" more than q, and so with constant probability we get two sets. From now on we assume for contradiction that

$$q = \Omega(1).$$

Now think of the following graph on X. For a pair $x, y \in X$ we have an edge labelled with a subset of S^{m-1} defined by $\{u : (x, y) \in M_u\}$. Notice that since M_u is a matching, the sets corresponding to edges out of a vertex x are disjoint. We define the degree of $x \in X$ as the measure of the union of the sets corresponding to the edges out of it, which is, by the above observation, the same as the sum of the measures of the edges out of it. Clearly, since $q = \Omega(1)$ we get the the average degree of the graph is some constant $\nu > 0$.

We next apply a simple greedy procedure that turns a graph with average degree ν to a graph with *minimum* degree $\nu/2$. This is done by removing every vertex with degree strictly less than $\nu/2$. By doing so the remaining total degree is strictly bigger than $|X|\nu - 2|X|(\nu/2) = 0$, and therefore we must remain with a nonempty graph with the guaranteed minimum degree.

Let Y be the surviving set of vertices. We claim that Y is a a $(\sigma, \nu/2, l)$ core as is defined in the tutorial notes. We simply note that the degree of $x \in Y$ is exactly $\Pr[\exists y \in Y(x, y) \in M_u]$ to establish that.

The point now is that Y is a very constrained object. In fact, it is so constrained that it has no choice but to be *very large*. This is well formulated in Lee's Big Core Theorem. We get $|Y| \geq \exp(\Omega(\frac{\sigma^6}{l^4 \log^2(2/\nu)})) = \exp(\Omega(\log n))$ and whenever $l = c(\log n)^{-1/4}$ with c small enough, the hidden constant in the Ω as as large as we want, hance we can get with an appropriate c that $|Y| > n$. Contradiction.

References

[ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, June 13-16, Chicago, IL, USA*, pages 222–231, 2004.

[DG62] L Danzer and B. Grünbaum. On two problems of p.erdös and v.l. klee concerning convex bodies *(in German)*. *Math. Zeitschrift*, 79:95–99, 1962.

[Lee05] J.R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, Vancouver*, 2005.