CSC2414 - Metric Embeddings*
Lecture 7: Lower bounds on the embeddability in ℓ_2 via expander graphs and some algorithmic connections to ℓ_1

Notes taken by Periklis Papakonstantinou
revised by Hamed Hatami

Summary: In view of Bourgain's upper bound a central question in finite metric embeddings concerns explicit constructions of families of metric spaces that incur distortion $\Omega(\log n)$ when embedded into ℓ_p. We will see that embedding constant degree expanders into ℓ_2 requires distortion $\Omega(\log n)$. On an independent direction we begin investigating algorithmic questions regarding the embeddability of metric spaces into ℓ_2 and ℓ_1. We show that computing a minimum distortion embedding of a metric space into ℓ_2 can be easily done in polynomial time. The corresponding question for ℓ_1 appears to be a computationally hard task (unless $\mathbf{P} = \mathbf{NP}$). Even the weaker question of deciding whether a metric space embeds isometrically into ℓ_1 is an $\mathbf{NP}$-complete problem. ℓ_1 seems to intuitively maintain a strong connection with $\mathbf{NP}$-hard combinatorial optimization problems. We begin our investigation around algorithmic questions for ℓ_1 by seeing how to represent an ℓ_1 metric as a weighted sum of cut metrics. Furthermore, we introduce the sparsest cut problem which is of particular importance to the theory of approximation algorithms.

1 Terminology - notational conventions

We use boldface to denote vectors. For example $\mathbf{x_i}$ corresponds to a vector whereas x_i corresponds to a real number. We denote by $Q_n = \{0,1\}^n$ the n-dimensional hamming cube; i.e. the hamming cube with 2^n points. Given a finite set of reals $A = \{a_1, \ldots, a_n\}$ we denote by $\text{Avg}_{i=1}^{n} a_i$ the arithmetic mean of A. We consider graphs with vertex set $V = \{1, 2, \ldots, n\}$ and denote the edges by ij, where $i, j \in [n]$. The term "expander graph" corresponds to a regular, constant degree, and of constant expansion graph. For unweighted graphs the distance between two vertices is the length of the unweighted shortest path i.e. the minimum number of edges of a path connecting

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

them (if they are not connected the distance is ∞). Let $G = (V, E)$ be an unweighted graph. We denote by d_G the shortest path metric on G.

2 Lower bounds in the distortion into ℓ_2

2.1 Introduction

Bourgain's theorem provides an $O(\log n)$ upper bound in the distortion of embedding any metric space of n points into ℓ_p. Of certain interest is the tightness when embedding into ℓ_2. Bourgain used the probabilistic method to show that there exists a family of finite metric spaces which require distortion $\Omega(\frac{\log n}{\log \log n})$ when embedding into ℓ_2 (e.g. see [Mat02] p.366). However this is not sharp and in fact we will show that his logarithmic upper-bound is sharp. In the previous lecture we saw that embedding the hamming cube of n points (i.e. of dimension $\log_2 n$) suffers distortion $\Omega(\sqrt{\log n})$. We also saw that with this construction we can not go any further since the standard embedding of the hamming cube into ℓ_2 has distortion $\sqrt{\log n}$. In this lecture we present an explicit[1] construction of families of metric spaces which require distortion $\Omega(\log n)$ when embedding into ℓ_2. In particular, we will show that embedding constant degree expander graphs equipped with the shortest path metric into ℓ_2 incurs distortion which matches the Bourgain's upper bound.

In Lecture 6 (Theorem 2.5) we presented a general framework based on Poincaré inequalities involving coefficients of positive semi-definite (PSD) matrices for proving lower bounds when embedding into ℓ_2. The second part of this theorem says that this framework is complete in the sense that if a specific metric d embeds with minimum distortion $c_2(d)$ in ℓ_2 then there exists a PSD matrix that can be applied to prove a lower bound of $c_2(d)$. We will prove the lower bound for embedding a family of expander graphs into ℓ_2 using similar inequalities. First we give a "direct" proof and then we will show how this proof can be easily fitted in the framework of the previous lecture.

2.2 Proof of the $\Omega(\log n)$ lower bound

Consider a family of expander graphs $G = (V, E)$, $V = [n]$ of constant degree $k \geq 3$ and spectral gap $\lambda_2 > 0$[2]. Fix an arbitrary embedding f of G into ℓ_2 of distortion $D \geq 1$ and let $f(i) = \mathbf{x_i}$. In particular, the inequality we are going to use is the following.

Note that on the left hand side i and j are adjacent vertices.

Lemma 2.1. *Let G be a k-regular expander graph with spectral gap $\lambda_2 > 0$. Then*

$$\mathit{Avg}_{ij \in E} \|\mathbf{x_i} - \mathbf{x_j}\|_2^2 \geq \frac{\lambda_2}{k} \mathit{Avg}_{i,j \in V} \|\mathbf{x_i} - \mathbf{x_j}\|_2^2, \tag{1}$$

where the first average is taken over all edges (every edge appears only once) while the second average is taken over all n^2 ordered pairs of vertices.

[1] In theoretical computer science the term "explicit" corresponds to mathematical objects of length n that can be computed in time polynomial in n (where n is given in unary). Our constructions are explicit per se since there are efficient constructions of constant degree expander graphs.

[2] λ_2 is the second smallest eigenvalue of the Laplacian. See Tutorial notes 2.

Proof. Since the graph is k-regular it has $nk/2$ edges.

$$\begin{aligned}
& \mathrm{Avg}_{ij\in E}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2 \geq \tfrac{\lambda_2}{k}\mathrm{Avg}_{i,j\in V}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2 \\
\Leftrightarrow \quad & \tfrac{\sum_{ij\in E}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2}{nk/2} \geq \tfrac{\lambda_2}{k}\tfrac{\sum_{i,j\in V}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2}{n^2} \\
\Leftrightarrow \quad & \tfrac{\sum_{ij\in E}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2}{nk/2} \geq \tfrac{\lambda_2}{k}\tfrac{\sum_{i,j\in V}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2}{n^2} \\
\Leftrightarrow \quad & 2\sum_{ij\in E}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2 \geq \tfrac{\lambda_2}{n}\sum_{i,j\in V}\|\mathbf{x_i}-\mathbf{x_j}\|_2^2 \qquad (2)
\end{aligned}$$

The technical reason for taking squares of Euclidean norms instead of Euclidean norms is that we can argue regarding (2) by proving it for real numbers instead of n-dimensional points (since afterwards we can sum up the corresponding inequalities). We abuse notation an from now on x_i refers to a real number.

Note that to prove (2) we can assume $\sum x_i = 0$. Denote by $\mathbf{x}$ the vector whose coordinates are the x_i's.

$$\begin{aligned}
\sum_{i,j\in V}(x_i-x_j)^2 &= 2\left((n-1)\sum_{i\in V}x_i^2 - \sum_{i\neq j}x_ix_j\right) \\
&= 2\left(n\sum_{i\in V}x_i^2 - (\sum_{i\in V}x_i)^2\right) = 2n\|\mathbf{x}\|_2^2 \qquad (3)
\end{aligned}$$

On the other hand recall from Tutorial 2 that

$$\sum_{ij\in E}(x_i-x_j)^2 = \mathbf{x}^t L_G \mathbf{x} \geq \lambda_2\|\mathbf{x}\|_2^2. \qquad (4)$$

Combining (2), (3) and (4) proves the lemma. □

Relying solely on the constant degree of the graph we can easily show the following fact.

Claim 2.2. *Let $G=(V,E)$ be a constant degree $k>0$ graph. Then, from every vertex $v\in V$ at least half of the vertices are at distance $\Omega(\log n)$.*

Proof. From an arbitrary vertex $v\in V$ there are at most $1+k+k(k-1)+\ldots+k(k-1)^{r-1}\leq k^r+1$ vertices at distance r. Therefore, if $r\leq\log_2\frac{n-1}{2}$ we have that at distance r there are at most $n/2$ vertices at distance $O(\log_2 n)$. □

Claim 2.2 implies that $\mathrm{Avg}_{i,j\in V}d(i,j)=\Omega(\log n)$. So

$$1 = \mathrm{Avg}_{ij\in E}d(i,j) \leq \frac{1}{\Omega(\log n)}\mathrm{Avg}_{i,j\in V}d(i,j).$$

Suppose that $d_G \overset{D}{\hookrightarrow} \ell_2^2$. Thus, since the embedding has distortion D and since (1) holds we have that $D^2 \geq \frac{\lambda_2}{k}\Omega(\log^2 n)$ which implies that $D=\Omega\left(\sqrt{\frac{\lambda_2}{k}}\log n\right)$. Therefore, since G is a constant degree expander we have that $D=\Omega(\log n)$.

2.3 Presenting the proof in the framework of Theorem 2.5

Actually the proof we presented in the previous section follows the framework of Theorem 2.5. To apply Theorem 2.5 (Lecture 6) it suffices to construct a good PSD matrix. We know

$$\sum_{ij \in E} (x_i - x_j)^2 = \mathbf{x}^t L_G \mathbf{x},$$

and

$$\sum_{i,j \in V} (x_i - x_j)^2 = \mathbf{x}^t (2nI - J)\mathbf{x}.$$

Thus, by (2) the matrix

$$Q = L_G - \lambda_2 (I - \frac{1}{n} J) I,$$

is PSD. Also $Q\mathbf{1} = \mathbf{0}$ since $\mathbf{1}$ is an eigenvector of L_G for $\mu_1 = 0$ and also $\frac{1}{n} J\mathbf{1} = \mathbf{1}$ which means $(I - \frac{1}{n} J)\mathbf{1} = \mathbf{0}$. We leave to the reader the details of applying the inequality of Theorem 2.5 (Lecture 6).

3 A polynomial time algorithm for embedding into ℓ_2

It is not hard to see that deciding whether a finite metric space (X, d) can be isometrically embedded into ℓ_2 is equivalent to deciding whether a symmetric matrix A (which can be efficiently computed) is positive semi-definite. This can be easily done by applying an algorithm that computes a factorization of $A = MM^t$ for some matrix M. This factorization is called Cholesky decomposition and an efficient algorithm can be found in almost every introductory Linear Algebra or Numerical Linear Algebra text.

We will not get into the details of dealing with this issue since we will answer the following generalization of the question.

Theorem 3.1. *Let (X, d) be a metric space of n points. Then, there exists an algorithm that computes an embedding of distortion $D := c_2(X)$ in time polynomial in n, where $c_2(X)$is the minimum required distortion distortion to embed X into ℓ_2.*

Proof. Note that since the distortion $D \geq 1$ minimizing D is the same as minimizing its square. Also, since the space is normed without loss of generality we can assume that the embedding is an expansion. Let f be an embedding of (X, d), $X = [n]$, s.t. $f(i) = \mathbf{x_i}, i \in [n]$. Therefore, finding the minimum D is equivalent to the following program.

$$\begin{array}{ll} \text{minimize} & D^2 \\ \text{subject to} & d(i,j)^2 \leq \|\mathbf{x_i} - \mathbf{x_j}\|_2^2 \leq d(i,j)^2 D^2 \\ & \mathbf{x_i} \in \mathbb{R}^n \end{array}$$

Now noting that $\|\mathbf{x_i} - \mathbf{x_j}\|_2^2 = \|\mathbf{x_i}\|_2^2 + \|\mathbf{x_j}\|_2^2 - 2\langle \mathbf{x_i}, \mathbf{x_j} \rangle$, it is not hard to obtain a semi-definite program (SDP). $\square$

4 Contrastive summary for ℓ_2 and ℓ_1

Before getting into algorithmic questions for ℓ_1 metrics we present a contrastive summary of the properties and meta-properties of ℓ_2 and ℓ_1 metrics.

ℓ_2 metrics	ℓ_1 metrics
Good dimension reduction	Does not have good dimension reduction
Efficient algorithm to compute an embedding of any finite metric space in polynomial time	It is NP-hard to compute an optimal embedding [Kar85]

Table 1: Contrastive comparison of ℓ_1 and ℓ_2 metrics.

5 Representation of ℓ_1 metrics as conical combinations of cut metrics

We provide a representation of ℓ_1 metrics as conical combinations[3] of cut metrics. This representation finds many applications. In particular, cut metrics are the extreme rays of the ℓ_1 cone. This is very useful as minimizing the ratio of linear functions over a convex cone is the same as minimizing over the extreme rays of the convex cone.[4]

A line metric is a one dimensional ℓ_1 metric. Let $x, y \in \mathbb{R}^n$. We denote by $d^{(i)}(x, y) = |x_i - y_i|$. Let d be an ℓ_1 metric. Clearly, d can be written as the sum of line metrics $d^{(i)}$. It is straightforward to verify the following fact.

Claim 5.1. *The set of ℓ_1 metrics is a convex cone. That is, if d_1, d_2 are ℓ_1 metrics and λ_1, λ_2 are non-negative reals then $\lambda_1 d_1 + \lambda_2 d_2$ is an ℓ_1 metric.*

Recall that we call a metric δ_S a cut (semi)metric if $\delta_S(i, j) = |\chi_S(i) - \chi_S(j)|$, where

$$\chi_S(i) = \begin{cases} 1 & i \in S \\ 0 & \text{otherwise} \end{cases}$$

The main representation theorem we want to show is:

Theorem 5.2. *Let d be a finite ℓ_1 metric. Then, d can be represented as $\sum_{S \subseteq [n]} a_S \delta_S$, for some constants $a_S \geq 0$.*

The proof of this theorem is transparent in the following illustrative example and follows as a corollary of this and the previous claim. Consider the line metric depicted in Figure 1.

By separating the line into two clusters (as in the figure) we get that this line metric d can be represented as

$$d = 3\delta_{\{a\}} + 10\delta_{\{a,b\}} + 5\delta_{\{a,b,c\}}$$

[3] A conical combination of vectors in $\mathbb{R}^n$ is a non-negative linear combination.

[4] Using this fact in order to prove Poincaré inequalities for ℓ_1 we only need to consider cut metrics, i.e. $f : X \to \{0, 1\}$ instead of $f : X \to \ell_1$ (this is similar to the situation that we showed in order to prove a Poincaré inequalities for ℓ_2, instead of considering $f : X \to \ell_2^n$ we could consider $f : X \to \ell_2^1$).

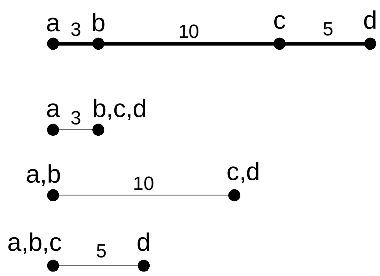

Figure 1: Example of a line metric

Now, it is clear that the theorem follows for every ℓ_1 metric.

Another way of viewing the theorem is that the cut metrics are the extreme rays of the ℓ_1 cone. Here is a lemma, useful when minimizing fractions of linear functions over a convex cone.

Lemma 5.3. *Let $C \subset \mathbb{R}^n$ be a convex cone and $f, g : \mathbb{R}^{n+} \to \mathbb{R}^+$ be linear functions and suppose that $\min_{x \in C} \frac{f(x)}{g(x)}$ is defined. Then,*

$$\min_{x \in C} \frac{f(x)}{g(x)} = \min_{x \text{ is in the extr. ray of } C} \frac{f(x)}{g(x)}$$

Proof. Let x_0 be the optimum. Then, since $x_0 \in C$ we have that $x_0 = \sum a_i y_i$, where y_i are in the extreme rays of C and a_i are non-negative reals.

$$\frac{f(x_0)}{g(x_0)} = \frac{f(\sum a_i y_i)}{g(\sum a_i y_i)} \frac{\sum f(a_i y_i)}{\sum g(a_i y_i)} \geq \frac{f(a_j y_j)}{g(a_j y_j)} = \frac{f(y_j)}{g(y_j)}$$

The second equality follows by linearity of f and g. The last inequality follows by the fact that for a fraction of the sum of the terms of any two finite sequences $\frac{\sum \alpha_i}{\sum \beta_i} \geq \min_j \frac{\alpha_j}{\beta_j}$ (we can show this by observing that adding numerators and denominators of two fractions of non-negative reals we get a fraction whose value is between the minimum and the maximum). □

6 The Sparsest-Cut problem

The sparsest cut problem is one of the major problems in the theory of approximation algorithms. On one hand it is related to other optimization problems. Also, its approximability is one of the central questions in approximation algorithms. The best known lower bound is that the problem does not admit a PTAS. This is still very far from the best known approximation algorithm with approximation ratio $O(\sqrt{\log n})$ [ARV04]. The first $O(\log n)$ approximation algorithm is due to [BL84] (1984). The same approximation ratio was achieved by applying methods of metric embeddings [LLR95, AR98] (1995). We will discuss an approach that uses metric embeddings and we will study the unweighted case.

Definition 6.1. We define the sparsity $\alpha(G)$ of an unweighted graph $G = (V, E)$ as

$$\alpha(G) = \min_{|S| \le \frac{|V|}{2}} \frac{|E(S, \bar{S})|}{|S|}$$

where $S \subseteq V$, $\bar{S} = V \backslash S$ and $E(S, \bar{S})$ is the set of edges crossing the cut $(S, \bar{S})$.

For example, consider the hamming cube Q_n. It is not hard to see that if we partition its vertices according to a fixed coordinate (i.e. let S be the set of points which have this coordinate set to 0), then we get the minimum $\alpha(Q_n) = \frac{2^{n-1}}{2^{n-1}} = 1$.

Intuitively the above definition suggests the definition of a combinatorial problem in which we wish to find a cut which is somehow balanced between the attempt to be minimal and to contain enough vertices. We define the following notion

$$\beta(G) = \min_{|S| \le \frac{|V|}{2}} \frac{|E(S, \bar{S})|}{|S||\bar{S}|}.$$

We are going to deal with the problem related to the later definition. The reason is that this definition is simpler and for every graph G it holds that $\alpha(G) \le n\beta(G) \le 2\alpha(G)$. The optimization problem we are interested in takes as an input a graph G and the objective is to output a cut $(S, \bar{S})$ such that $\frac{|E(S,\bar{S})|}{|S||\bar{S}|}$ is minimized.

Fix an arbitrary graph $G = (V, E)$, $V = [n]$ and an arbitrary cut of this graph $(S, \bar{S})$. We observe that

$$\frac{|E(S, \bar{S})|}{|S||\bar{S}|} = \frac{\sum_{ij \in E} \delta_S(i, j)}{\sum_{i,j \in V} \delta_S(i, j)}.$$

This observation together with Lemma 5.3 implies the following theorem.

Theorem 6.2. *Let $G = (V, E)$ be a graph. Then,*

$$\min_{|S| \le \frac{|V|}{2}} \frac{|E(S, \bar{S})|}{|S||\bar{S}|} = \min_{d \in \ell_1} \frac{\sum_{ij \in E} d(i, j)}{\sum_{i,j \in V} d(i, j)}$$

Based on the above theorem computing $\beta(G)$ is equivalent to the following mathematical program.

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{ij \in E} d(i, j) \\ \text{subject to} & \displaystyle\sum_{i,j \in V} d(i, j) = 1 \\ & d \in \ell_1 \end{array}$$

The requirement that d is an ℓ_1 metric is not known to be representable as a small set of linear constraints[5]. We "relax" (in some non-standard relaxation notion) the

[5]Note that there is some inaccuracy here. What we really require is the linear program to have a polytime separation oracle. Also note that according to what we have said before if this was true this would imply $P = NP$.

above program by only requiring that d is a metric. This introduces only a number of constraints polynomial in n; in particular $3\binom{n}{3}$ triangle inequlaities. Therefore, the relaxed linear program takes the following form which is solvable in time polynomial in n.

$$\begin{array}{lll}
\text{minimize} & \sum_{ij \in E} d(i,j) & \\
\text{subject to} & \sum_{i,j \in V} d(i,j) = 1 & \\
& d(i,j) \leq d(i,k) + d(k,j) & \text{for every } i,j,k \in [n].
\end{array}$$

In the next lecture we show that this relaxation leads to a $\log(n)$-approximation algorithm.

References

[AR98] Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231 (electronic), New York, 2004. ACM.

[BL84] Sandeep N. Bhatt and Frank Thomson Leighton. A framework for solving VLSI graph layout problems. *J. Comput. System Sci.*, 28(2):300–343, 1984.

[Kar85] A. V. Karzanov. Metrics and undirected cuts. *Math. Programming*, 32(2):183–198, 1985.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.