CSE 254: Seminar on Learning Algorithms: Embeddings March 13th, 2007

Dimension reduction in L_1: a negative result

Professor: Sanjoy Dasgupta *Lecturer: Claire Monteleoni*

The Johnson-Lindenstrauss lemma shows that only $d = O(\frac{1}{\epsilon^2}\log n)$ dimensions are needed to embed any set of n points in L_2 into ℓ_2^d with distortion at most $(1+\epsilon)$. We will show that such dimension reduction is *not* possible in L_1. In particular, we will give a set of n points in L_1 that cannot be D-embedded into ℓ_1^d unless $d \geq n^{\Omega(1/D^2)}$.

This result was originally shown by Brinkman and Charikar [1], providing a negative answer to whether a Johnson-Lindenstrauss analog exists for L_1, a previously open question (see e.g. [5]). Our lecture will follow a different proof by Lee and Naor [6].

The high-level overview of both proofs is simply to show that a particular point set cannot be embedded without the stated number of dimensions. The point set is the recursive diamond graph, which can be defined on n vertices, for n that is any power of 2. Values of n that are not powers of 2 are handled by noting that there exists a power of 2 such that the associated diamond graph is O(1)-equivalent to a point set of size n.

1 Preliminaries: the recursive diamond graph

- Planar graphs.
- Series parallel graphs.

Definition 1 *We will define these graphs recursively. A single edge (s,t) is a series parallel graph. Given two series parallel graphs, with source and target notes s,t and s',t', they can be combined as follows:*

In series: Merge t and s' without deleting edges. Set new target to be t'.

In parallel: Merge s and s', and merge t and t', without deleting edges.

- The recursive diamond graph.

Definition 2 *We will define these graphs recursively. A single edge (s,t) of weight one, is the recursive diamond graph G_0. To create the recursive diamond graph G_i from the recursive diamond graph G_{i-1}, each edge $(x,y) \in E(G_{i-1})$ is replaced by the following four edges (x,p), (p,y), (y,q), (q,x), and p and q are added to $V(G_i)$. The edge-weights of these four new edges are each equal to half of the weight of the original edge $(x,y) \in E(G_{i-1})$.*

These graphs have the following properties:

1. The weight of any edge in G_i is 2^{-i}.
2. $|E(G_i)| = 4^i$.
3. The total edge weight of G_i is thus $4^i 2^{-i} = 2^i$.
4. The weight of paths is preserved. I.e., defining $d_i(x,y)$, the distance between $x,y \in V(G_i)$, as the shortest path metric, with respect to edge-weights in the graph G_i, it is the case that $d_j(x,y) = d_i(x,j)\ \forall j > i$. Thus it will suffice to denote distance as $d(x,y)$ which is valid only when both nodes x and y are present in the current diamond graph.
5. We define the anti-edge at level i using the terminology in the definition above. For (x,y) the original edge in level $i-1$, its anti-edge at level i is (p,q). Thus $d(p,q) = 2 \times 2^{-i} = 2^{1-i}$.

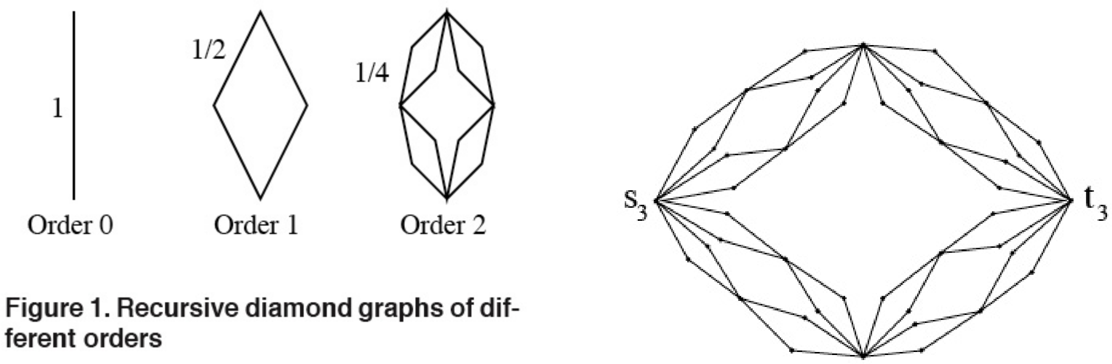

Figure 1: credit: [1], Order 3 graph: [8].

2 Lee and Naor proof [6]

Given that we will use the recursive diamond graph as the point-set requiring the lower bound on the embedding dimension (originally discovered by [1]), Lee and Naor provide a rather short proof in [6] based on geometric arguments, using some facts from convexity.

Theorem 1.1 *For every* $1 < p \leq 2$*, any embedding of* G_k *into* L_p *incurs distortion* $D \geq \sqrt{1+(p-1)k}$.

The proof will make use of the follow lemmas:

Lemma 2.1 ("Short diagonals lemma" when $p = 2$**)** *Fix* $1 < p \leq 2$ *and* $x, y, z, w \in L_p$*. Then*

$$\|y-z\|_p^2 + (p-1)\|x-w\|_p^2 \leq \|x-y\|_p^2 + \|y-w\|_p^2 + \|w-z\|_p^2 + \|z-x\|_p^2$$

Proof. We start with the uniform convexity lemma:
For every $a, b \in L_p$, $\|a+b\|_p^2 + (p-1)\|a-b\|_p^2 \leq 2(\|a\|_p^2 + \|b\|_p^2)$. A proof of this classifical fact can be found in [7]. We can rewrite this with the following substitutions: $a = y - x$, $b = x - z$ as,

$$\|y-z\|_p^2 + (p-1)\|y-2x+z\|_p^2 \leq 2\|y-x\|_p^2 + 2\|x-z\|_p^2$$

and letting $a = y - w$, $b = w - z$ as,

$$\|y-z\|_p^2 + (p-1)\|y-2w+z\|_p^2 \leq 2\|y-w\|_p^2 + 2\|w-z\|_p^2$$

Summing these two inequalities and dividing by 2 yields:

$$\|y-z\|_p^2 + (p-1)\frac{\|y-2x+z\|_p^2 + \|y-2w+z\|_p^2}{2} \leq \|y-x\|_p^2 + \|x-z\|_p^2 + \|y-w\|_p^2 + \|w-z\|_p^2$$

The lemma follows, since

$$\|x-w\|_p^2 \leq \frac{\|y-2x+z\|_p^2 + \|y-2w+z\|_p^2}{2}$$

by convexity of $\|\cdot\|_p^2$ (Jensen's inequality). □

Lemma 2.2 *Let A_i denote the set of anti-edges at level i and set $\{s,t\} = V(G_0)$, then for $1 < p \le 2$ and any $f : G_k \to L_p$,*

$$\|f(s)-f(t)\|_p^2 + (p-1)\sum_{i=1}^{k}\sum_{(x,y)\in A_i}\|f(x)-f(y)\|_p^2 \le \sum_{(x,y)\in E(G_k)}\|f(x)-f(y)\|_p^2$$

Proof. Apply Lemma 2.1 to (a,b) an edge at level i and (c,d), its corresponding anti-edge, yielding:

$$\|f(a)-f(b)\|_p^2 + (p-1)\|f(c)-f(d)\|_p^2 \le \|f(a)-f(c)\|_p^2 + \|f(c)-f(b)\|_p^2 + \|f(b)-f(d)\|_p^2 + \|f(d)-f(a)\|_p^2$$

Summing over all such edges, and all levels $i \in \{0,\dots,k-1\}$, and defining A_i as the set of anti-edges at level i, we attain:

$$\sum_{i=0}^{k-1}\sum_{(a,b)\in E(G_i)}\|f(a)-f(b)\|_p^2 + (p-1)\sum_{i=1}^{k}\sum_{(c,d)\in A_i}\|f(c)-f(d)\|_p^2 \le \sum_{i=1}^{k}\sum_{(x,y)\in E(G_i)}\|f(x)-f(y)\|_p^2$$

We then subtract $\sum_{i=1}^{k-1}\sum_{(a,b)\in E(G_i)}\|f(a)-f(b)\|_p^2$ from each side. □

We can now proceed to prove Theorem 1.1.

Proof of Theorem 1.1. Let $f : G_k \to L_p$ be a non-expansive D-embedding. We apply Lemma 2.2, and note the following. By definition of distortion, $\frac{1}{D} \le \frac{\|f(s)-f(t)\|_p}{d(s,t)} = \|f(s)-f(t)\|_p$ where the second step is by definition of G_0. This provides a lower bound of $\frac{1}{D^2}$ on the first term of the LHS. To analyze the second term of the LHS, we note that, by construction of the recursive diamond graph, $|A_i| = 4^{i-1}$. Additionally since the edge lengths at level i are each 2^{-i}, the length of a diagonal (or anti-edge) is $2 \times 2^{-i} = 2^{1-i}$. Thus letting (x_i, y_i) be any anti-edge at level i, the second term of the LHS can be expressed and lower bounded as follows,

$$(p-1)\sum_{i=1}^{k}|A_i|\,\|f(x_i)-f(y_i)\|_p^2 \ge (p-1)\sum_{i=1}^{k}|A_i|\left(\frac{d(x_i,y_i)}{D}\right)^2 = (p-1)\sum_{i=1}^{k}4^{i-1}\left(\frac{2^{1-i}}{D}\right)^2 = \frac{(p-1)k}{D^2}$$

where the first inequality is by definition of distortion. Thus we lower bound the whole LHS by $\frac{1}{D^2}(1+(p-1)k)$. To complete the proof, we upper bound the RHS by one, noting that $f(\cdot)$ is not an expansion, so we can upper bound the RHS by $\sum_{(x,y)\in E(G_k)}(d(x,y))^2 = 4^k(2^{-k})^2$ □

Corollary 1.2 *For every $n \in \mathbb{N}$, there exists an n-point subset $X \subseteq L_1$ such that for every $D > 1$, if X D-embeds into ℓ_1^d then $d \ge n^{\Omega(1/D^2)}$.*

Proof. Consider $p = 1 + \frac{1}{\log d}$, and apply Theorem 1.1, yielding distortion at least $\sqrt{1+\frac{k}{\log d}}$. Noting that ℓ_1^d is $O(1)$-isomorphic to ℓ_p^d for this value of p, we have that the distortion D for embedding in ℓ_1^d obeys $\sqrt{1+\frac{k}{\log d}} = O(D)$. For any n, by the definition of the recursive diamond graph, $k = \Omega(\log n)$ for the diamond graph G_k that is $O(1)$-equivalent to a subset $X \subseteq L_1$ of size n. Thus we have that,

$$\sqrt{1+\frac{\Omega(\log n)}{\log d}} = O(D)$$

which yields the desired result. □

3 Implications and open problems

- Lee and Naor in [6] solve the open problem posed by Brinkman and Charikar in [1], as to whether the lower bound holds for ℓ_p for $p \in (1, 2)$: YES.
- Brinkman and Charikar in [1] explain why, based on previous results, attaining a significantly tighter lower bound than $n^{\Omega(1/D^2)}$ cannot be done using a planar graph. This can be seen by composing embeddings. One can embed n points in ℓ_2 into $\ell_1^{O(\log n)}$ with distortion $(1+\epsilon)$. If it is possible to embed n points from ℓ_1 to ℓ_2 with distortion D, then combined with the first embedding this would entail dimension reduction for ℓ_1 down to $O(\log n)$ dimensions with distortion D. The known bounds on distortion for the second embedding are upper bound $O(\log n)$ using Bourgain embedding, and lower bound $\Omega(\sqrt{\log n})$. Planar graphs attain the lower bound, i.e. distortion is $O(\sqrt{\log n})$ by the result of Rao. However for distortion $D = \sqrt{\log n}$, the composed embedding yields a relation $d = n^{O((\log D)/D^2)}$, where d is the dimension of the embedding. Thus the best lower bound that could be attained by a planar graph is $n^{\Omega((\log D)/D^2)}$, which is not significantly tighter.

4 Related work

High-level idea of approach of Brinkman and Charikar [1]:

- They consider *stretch*-limited embeddings. A stretch-s embedding is a convex combination of line metrics, where distances in *any* line metric cannot be more than a factor s larger than distances in the original metric. This notion is more general than an ℓ_1 embedding with s dimensions, in that the dimension can be arbitrary, and *any* convex combination of line metrics is permitted. However they are a good proxy for ℓ_1 embedding with s dimensions because of the following two facts:
  1. *The existence of a D-distortion embedding of a metric M into ℓ_1^s implies the existence of a D-distortion stretch-limited embedding of M with stretch s.*
  2. *The existence of a D-distortion stretch-limited embedding of a metric (M, d) with stretch s implies the existence of a $D(1+\epsilon)$-distortion embedding of (M, d) into $\ell_1^{O(sD\log(n)/\epsilon^2)}$.*
- Thus it would suffice to provide a lower bound on the stretch s needed to achieve a given distortion D.
- Formulate an LP to minimize stretch, subject to the distortion-induced constraints on all the pairwise distances. Taking the dual, and finding a feasible solution establishes the lower bound on stretch. (General idea. There are some additional complications that they overcome.)
- Note that Brinkman and Charikar [1] attain a slightly more precise bound. For example, for distortion $(1+\epsilon)$, their bound is $d \geq n^{\frac{1}{2} - O(\epsilon \log(1/\epsilon))}$.

Work prior to Brinkman and Charikar [1]:

- Recursive diamond graph can embed in ℓ_1 with constant distortion but many dimensions [3].
- There exists an embedding from ℓ_1^d to $\ell_1^{d'}$ where $d' = (\frac{1}{\epsilon}\log 1/\delta)^{O(1/\epsilon)}$ with the following properties. With probability ϵ distances do not increase (so they can increase with large probability). With probability $1-\delta$ distances do not decrease with distortion $(1+\epsilon)$ [4].
- *Linear* embeddings do not obtain dimension reduction in ℓ_1. There exists a set of $O(n)$ points in ℓ_1^n such that any linear embedding incurs distortion $\Omega(\sqrt{n/d})$ [2].
- Low dimensional, low distortion embeddings exist for several ℓ_1 embeddable metrics, e.g. tree metrics, shortest path metrics on outer-planar graphs [2].

References

[1] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in ℓ_1. In *Proc. 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[2] M. Charikar and A. Sahai. Dimension reduction in the ℓ_1 norm. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002.

[3] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1 embeddings of graphs. In *Proc. 40th Annual IEEE Symposium on Foundations of Computer Science*, 1999.

[4] P. Indyk. Stable distributions, pseudorandom generators, embedding and data stream computation. In *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science*, 2000.

[5] P. Indyk. Algorithmic applications of low-distortion embeddings. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, 2001.

[6] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004.

[7] A. Naor. Proof of the uniform convexity lemma. In *http://www.cims.nyu.edu/∼naor/homepage files/inequality.pdf*, 2004.

[8] I. Newman and Y. Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. In *Proc. 18th Annual Symposium on Computational Geometry*, 2002.