CSC2414 - Metric Embeddings[*]
Lecture 9: Dimension reduction in ℓ_1 and Planar Metrics

Notes taken by Ilya Sutskever
revised by Hamed Hatami

Summary: We provide a neat proof that ℓ_1, unlike ℓ_2, does not have good dimension reduction. We also show that the existence of a certain type of partition on a graph yields a good embedding of the planar graphs to ℓ_2.

1 Dimension Reduction in ℓ_1

The aim of this section is to prove that ℓ_1 does not have good dimension reduction. See [BC03]. We follow [LN04].

Theorem 1.1. *There is an n-point metric space that is ℓ_1 and that an embedding with distortion at most D into ℓ_1 requires dimension $n^{\Omega(1/D^2)}$.*

In particular, to get $\log n$ dimensions, D^2 must be $\log n / \log\log n$.

1.1 The Idea

The key idea is to relate distortion with dimension. Observe that if d embeds to ℓ_1^m with distortion D then d embeds to ℓ_p with distortion $D \cdot m^{1-1/p}$.

For the proof, we will make use of the identity embedding from ℓ_1 to ℓ_p.

$$\|x\|_p \le \|x\|_1 \le m^{1-1/p} \cdot \|x\|_p,$$

where the second inequality follows from Hölder's inequality, specifically

$$\|x\|_1 = \sum_{i \le m} 1 \cdot |x_i| \le \left(\sum_{i \le m} |x_i|^p \right)^{1/p} \cdot \left(\sum_{i \le m} 1^{(1-1/p)^{-1}} \right)^{1-1/p} = \|x\|_p \cdot m^{1-1/p}.$$

Thus the identity embedding has distortion $m^{1-1/p}$, and the embedding from d to ℓ_1 has distortion D, therefore the distortion of the composition is at most $D \cdot m^{1-1/p}$.

---

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

This is the link from distortion to dimension.

Now for $p = 1 + 1/\log m$ we get

$$m^{1-1/p} = m^{1/(1+\log m)} \leq m^{1/\log m} = e^{\log m/\log m} = O(1). \tag{1}$$

Suppose that we produce a metric embeddable to ℓ_1 with constant distortion and in addition, for all $p \in [1,2]$, $c_p(d) = \Omega(\sqrt{p-1}\sqrt{\log n})$ where $c_p(d)$ is the best distortion of an embedding of the metric d to ℓ_p.

Then let $p = 1 + 1/\log m$ where m is the number of dimensions. We get $c_p(d) = \Omega(\sqrt{\log n/\log m})$.

If d embeds to ℓ_1^m with distortion D, then d embeds with distortion $O(D)$ to ℓ_p by (1), which must be greater than $c_p(d)$:

$$D \geq \Omega\left(\sqrt{\frac{\log n}{\log m}}\right),$$

and for this to be true, m must be $n^{\Omega(1/D^2)}$.

An example of such a metric space is the diamond graph discussed in the tutorial from week 6.

2 Planar Metrics

Definition 2.1. A metric d is called planar if it is the metric induced by a (weighted) planar graph

Theorem 2.2. *Every planar metric d embeds to ℓ_2 with distortion $O(\sqrt{\log n})$*

The theorem is due to [Rao99].

2.1 Conventions and notations

For the proof, we will use a distribution over partitions, similar to the one used in Lecture 4.

Let (X,d) be our n-points planar metric space. Consider a probability distribution over partitions of X, associated with parameters (Δ, α), satisfying the following properties.

1. $\mathrm{diam}(P) = O(\Delta)$ with probability 1.
2. For all x, the probability that the ball centered at x of radius Δ/α is entirely in one set of the partition is bigger than $1/2$ (the partitions are "solid").

Clearly such a partition can not exist for all settings of the parameter. In particular, we want Δ to be large and α to be small. Part of the proof involves showing that α is unusually small for planar metrics, which we do not show here.

For a metric space X, let α_X be the smallest possible α such that for all Δ there is a partition with parameters (Δ, α_X).

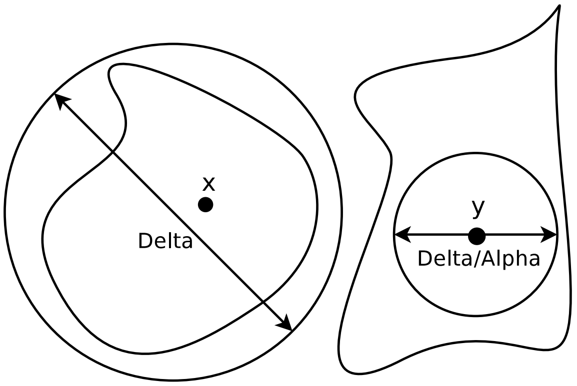

Figure 1: The partition

Theorem 2.3. *For every metric space X, we have*

$$c_2(X) = O(\alpha_X \cdot \sqrt{\log n})$$

Theorem 2.4. *If X is planar, $\alpha_X = O(1)$.*

The above theorems imply that planar graphs embed with distortion $O(\sqrt{\log n})$ to ℓ_2.

2.2 Proof of Theorem 2.3

Given a Probability distribution over the partitions of X as above, create a distribution over subsets. Fix Δ. Pick $\pi_\Delta \sim P_\Delta$, a partition. Generate a random subset of X by picking each set in the partition π_Δ with probability 1/2 independently and take the union of these sets. Call this random subset Z_Δ.

Let $x, y \in X$ with $c_1\Delta < d(x,y) < 2c_1\Delta$ (c_1 is a constant that is independent of n, and is hidden in the $O(\Delta)$). Then x and y will be separated by every partition with parameter Δ, since every element of the partition has diameter smaller than $d(x,y)$.

Let us examine

$$|d(x, Z_\Delta) - d(y, Z_\Delta)|$$

On the one hand, this is less than $d(x,y)$, which is less than $2c_1\Delta$ by assumption. On the other hand, if x happens to be in Z_Δ (an event occurring with probability 1/2), and the y is not in Z_Δ (an independent event occurring with probability 1/2 since x and y are in different elements of the partition) and that the ball of radius $c_1/\alpha_x \cdot \Delta$ around y is in the element of the partition containing y (yet an independent event with probability at least 1/2, since the previous events did not depend on the way the partition was formed) then with probability at least $1/8$

$$|d(x, Z_\Delta) - d(y, Z_\Delta)| > c_1/\alpha_X \cdot \Delta$$

Let us construct a Frechet embedding using these Z_Δ sets. Specifically, let us, for each choice of parameter Δ create $t = c_2 \log n$ independent copies of Z_Δ. In addition, let us first assume that the weights of the graph are unit weights, so the distance is in $[n]$. Therefore, we shall restrict the parameter Δ to be among $1, 2, 4, \ldots, 2^{j-1}, 2^{\lceil \log n \rceil}$. Thus we get a Frechet embedding to $O(\log^2 n)$ dimensions.

Let us analyze it:

$$\|f(x) - f(y)\|_2^2 \geq \sum_{t \leq c_2 \log n} |d(x, Z_\Delta) - d(y, Z_\Delta)|^2 \geq 1/64 \left(\frac{\Delta}{\alpha_X}\right)^2 \cdot (\log n) \cdot c'$$

where in the first inequality we choose a suitable Δ for the distance. Here c' is some constant independent of n. Then we note that since the Z_Δ are independent, and the above inequality holds with probability at least $1/8$ per Z_Δ, the probability that it holds for at least $1/64$ of all t's is negligible, so that with positive probability, there exists a choice of the Z_Δ's that makes this inequality true for all x and y.

Suppose that this event occurs, i.e. for all x and y the above inequality holds. Then

$$O(d(x,y)^2 \log^2 n) \geq \|f(x) - f(y)\|_2^2 \geq O(\log n) \cdot \left(\frac{\Delta}{\alpha_X}\right)^2 \geq O(d(x,y)^2 \log n)/\alpha_X^2$$

Where the first inequality holds since every coordinate $|d(x, Z) - d(y, Z)|$ of the Frechet embedding is non-expanding and there are $O(\log^2 n)$ coordinates. The second inequality holds since the above Δ differs at most by a fixed factor from $d(x, y)$.

Therefore, the distortion of d is $O(\alpha_X \sqrt{\log n})$.

2.3 The proof for weighted graphs

To prove the result for weighted graphs, we need to use a different construction of Z_Δ. In particular, we will not use a partition and then select sets at random from it, but construct the set directly. We loosely follow [Mat02].

First we describe the construction and state its properties. This construction has only one parameter, Δ. We still assume that the distances are in $\{1, \ldots, n\}$, i.e. that we have a unit weight graph. This assumption will later be removed.

Let V be the set of vertices.

The algorithm:

- Pick an arbitrary vertex $x \in V$.
- Pick a radius r_1 uniformly from $\{0, 1, \ldots, \Delta - 1\}$.
- Construct annuli around x:

$$\begin{aligned} &B(x, r_1), \\ &B(x, r_1 + \Delta) - B(x, r_1), \\ &B(x, r_1 + 2\Delta) - B(x, r_1 + \Delta), \\ &\ldots \end{aligned}$$

- Remove all the vertices that are on the boundaries between the annuli.

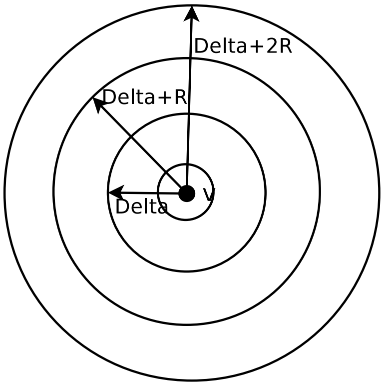

Figure 2: The making of the set

- Repeat the same for each of the connected components, *i*.e. pick a starting vertex x_i and a radius r_2 uniformly in $\{0, 1, \ldots, \Delta - 1\}$.

This procedure is only repeated once, and it is not recursive.

Denote by B the set of all vertices removed. For a planar metric, we state the properties of B and $V - B$.

It can be shown that there are universal constants c_1, c_2, c_3 such that

1. The diameter of each connected component of $V - B$ is at most $c_3\Delta$.
2. For each vertex $v \in V$, $d(v, B) > \Delta c_1$ with probability at least $c_2 > 0$.

Note how the connected components resemble the partition.

We now will define the embedding, similar to the one in the previous section. We then extend it to the case of arbitrary weighted graph.

For each connected component of $V - B$ assign randomly and independently a value from $\{-1, 1\}$, and let σ_x be the value assigned to the component of x. Let σ_x be 0 if $x \in B$.

Define $f_\Delta(w) = \sigma_w d(w, B)$, be one coordinate of the embedding. Here B was constructed as above with parameter Δ. Let

$$F(w)_{\Delta,k} = f_\Delta(w)$$

be the embedding itself, where k ranges over $O(\log n)$ independent copies of f_Δ and Δ ranges over $1, 2, \ldots, 2^{O(\log n)}$ so that $F : X \to \ell_2^{O(\log^2 n)}$. Let us investigate its properties.

Let $x, y \in V$ and let $\Delta = 2^j$, $(j \in \mathbb{Z})$ be such that

$$c_3\Delta < d(x, y) < 2 \cdot c_3\Delta,$$

and let B be the result of the above construction with parameter Δ.

Since the diameter of each component of $V - B < c_3\Delta$, x and y can't belong to the same component, so σ_x may not be equal to σ_y.

Consider

$$|f_\Delta(x) - f_\Delta(y)| = |d(x,B)\sigma_x - d(y,B)\sigma_y|$$

If $\sigma_x \neq \sigma_y$ and if $d(x,B) \geq c_1\Delta$ then $|d(x,B)\sigma_x - d(y,B)\sigma_y| > c_1\Delta$, an event that occurs with probability at least $c_2/2$. Thus with 1-exponentially small probability this inequality holds for a constant fraction of the independent copies, hence

$$\|F(x) - F(y)\|_2^2 \geq \Omega(\Delta^2 \log n) = \Omega(d(x,y)^2 \log n)$$

with a very high probability. Therefore, this inequality holds simultaneously for all pairs x, y with non-zero probability.

In addition, $|d(x,B)\sigma_x - d(y,B)\sigma_y| < 2\max_{z\in x,y} d(z,B) < 2c_3\Delta < 2d(x,y)$, implying that $\|F(x) - F(y)\|_2^2 \leq O(d(x,y)^2 \log^2 n)$.

From this point, the analysis is as in the previous section: On the one hand, we have $\|F(x) - F(y)\|_2^2 \leq O(d(x,y)^2 \log^2 n)$ and on the other hand, for every pair x, y with very high probability $\|F(x) - F(y)\|_2^2 \geq \Omega(d(x,y)^2 \log n)$, thus with non-zero probability this inequality holds for all x, y. This gives the $O(\sqrt{\log n})$ distortion in ℓ_2.

2.3.1 applying the construction on weighted graphs

Now we show how to generalize this to an arbitrarily weighted graph. We modify the graph in a way that depends on Δ, so $f_\Delta(x)$ is using different graphs for different Δ's.

The first modification that we make is that when creating the set B with parameter Δ, we modify the graph so that if $(u,v) \in E$ and $d(u,v) < \Delta c_3/8n$, then we set $d(u,v) = 0$. So short edges get contracted.

This is done in order make sure that $\|F(x)-F(y)\|_2^2$ is not being made much larger by coordinates with a very large Δ, compared to $d(x,y)$.

Having contracted some edges, we modify the graph as follows:

Given Δ, in the construction of the set B we used the vertices of distances $r_1 + k\Delta$ from x to be in B. But the weighted graph may not have such vertices, so B may end up being empty which is not good.

To overcome this problem we add a virtual vertex v on every edge (u,w) such that $d(x,u) = r_1 + k\Delta - \varepsilon$ and $d(x,w) = r_1 + k\Delta + \varepsilon'$, so that the end result is $d(x,v) = r_1 + k\Delta$. Note that a very long edge on the original graph may get many virtual vertices added. Similarly, in the same way add similar virtual vertices to the connected components of $V-$ the first set of annuli (see the algorithm).

It can be shown that as a result of this construction B will satisfy exactly the same properties it satisfied for the unweighted graph, a fact that will be used implicitly in the analysis of both the upper and lower bounds.

Denote by d_Δ the graph metric that we get from these two modifications.

Let us then define the embedding: $F(x)_{\Delta,t} = f_\Delta(x)$, where $f_\Delta(x) = d_\Delta(x,B)\sigma_x$. σ_x is defined is as in the previous section. Δ ranges over a sufficiently large range of powers of 2, to contain all the distances in the graph. t ranges over $O(\log n)$ independent copies of f_Δ for each Δ.

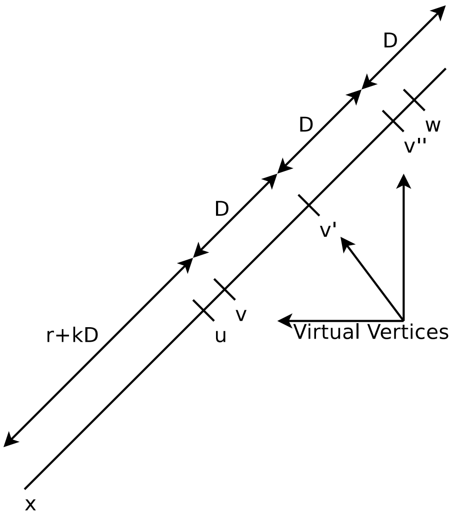

Figure 3: Adding virtual vertices

The set B is created with the above modifications for each Δ; The graph has extra vertices, and some of the original edges are contracted to 0. $F : X \to \ell_2^M$ for some possibly very large M.

Let us analyze this embedding.

$$||F(x) - F(y)||_2^2 = \sum_{\Delta,t} |f_\Delta(x) - f_\Delta(y)|^2$$

If $\Delta c_3 > 8nd(x,y)$, then $|f_\Delta(x) - f_\Delta(y)| = 0$, since for such a large Δ, B was constructed with $d_\Delta(x,y) = 0$. So x and y are in the same connected component of $V - B$ (or both are in B) with the appropriate Δ, so $\sigma_x = \sigma_y$ and had $d_\Delta(x,B) = d_\Delta(y,B)$. So $|d_\Delta(x,B)\sigma_x - d_\Delta(y,B)\sigma_y| = 0$.

If $\Delta < d(x,y)$, then $|f_\Delta(x) - f_\Delta(y)| < \Delta$ since in the event where Δ is much smaller than $d(x,y)$, x and y will necessarily be in different connected components which have small diameter (or both are in B, so $d(x,B) = 0$ and $d(y,B) = 0$).

Therefore, the sum-total of the small Δ's contribute little and the Δ's above $2nd(x,y)$ don't contribute at all to $||F(x) - F(y)||_2^2$. This leaves us with $O(\log n)$ Δ's, each of which has $O(\log n)$ independent copies of f_Δ, and the contribution of each coordinate is bounded by $O(d(x,y)^2)$.

Therefore, we conclude that $||F(x) - F(y)||_2^2 \leq O(d(x,y)^2 \log^2 n)$.

To have a good distortion, we need to show that the embedding doesn't contract too much. We will show that $|f_\Delta(x) - f_\Delta(y)| \geq \Omega(d(x,y))$ for $2c_3\Delta \leq d(x,y) \leq 4c_3\Delta$ with probability bounded below by a constant (Note the slightly different choice of Δ). This will imply that with non-zero probability, $||F(x) - F(y)|| \geq \Omega(d(x,y)^2 \log n)$ for all x, y simultaneously, which implies that the distortion is $O(\sqrt{\log n})$.

For this to be true, the only thing we need is that x and y belong to different connected components of $V - B$ in the modified graph d_Δ. This is the only way σ_x may be not equal to σ_y with probability $1/2$. If x and y indeed belong to different components with probability 1, then the analysis of the last section shows that the desired inequality holds with probability bounded from zero by a constant.

2.3.2 Some details

We will show that $d(x,y) < 2d_\Delta(x,y)$ if $2c_3\Delta \leq d(x,y) \leq 4c_3\Delta$. From that we get $c_3\Delta < d_\Delta(x,y)$, i.e. x and y belong to different connected components in the modified graph d_Δ, as desired.

For the proof, consider the shortest path in the original graph between x and y. At most n edges may be contracted (since we contracted the original graph without virtual vertices), each of length at most $\Delta c_3/8n$. Therefore, $d_\Delta(x,y) \geq d(x,y) - n\Delta c_3/8n$. Since $\Delta c_3 < 4d(x,y)$, we get that $d(x,y) - n\Delta c_3/8n \geq d(x,y) - d(x,y)/2$. Thus x and y belong to different components in $V - B$ in d_Δ.

Combining the two inequalities we get $d(x,y)/2 \leq d_\Delta(x,y)$.

References

[BC03] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in l1. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[LN04] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1, July 29 2004. Comment: 3 pages. To appear in Geometric and Functional Analysis (GAFA).

[Mat02] Jiri Matousek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and euclidean metrics. In *Proceedings of the Conference on Computational Geometry (SCG '99)*, pages 300–306, New York, N.Y., June 13–16 1999. ACM Press.