Metric Embeddings - What, Why, How?

Instructor: Sylvester Eriksson-Bique

1 Background

1.1 Literature and overview

These were notes written for a minicourse held at Okinawa Institute of Science and Technology in Spring 2024. The course consisted of 4 lectures of two hours each.

Some further reading, and notes which I have used in the preparation of these lectures:

1. Matoušek, Jiří, *Lecture notes on metric embeddings.*

2. Avner Magen's course on metric embeddings: `http://www.cs.toronto.edu/~avner/teaching/S6-2414/`

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

4. Nice notes by Yury Makarychev, see `https://home.ttic.edu/~yury/courses/geometry/`. Especially the section on the sparsest cut problem is very well written.

5. A good reference on the Goemans-Linial semidefinite relaxation is Arora, Lee and Naor's paper: `https://www.ams.org/journals/jams/2008-21-01/S0894-0347-07-00573-5/S0894-0347-07-00573-5.pdf`

1.2 General background

Some notation and background is listed here. It is advised to merely skim through it, as it should be standard, and come back to it if needed. These also serve the purpose of listing some basic pre-requisites, and if the notions here feel very difficult, the course may be quite difficult to follow.

Given a mapping $f : X \to Y$ between two sets, we write $\mathrm{Im}\,(f)$ for its image.

The volume/Lebesgue measure of a subset $A \subset \mathbb{R}^n$ is denoted $|A|$. The volume of a unit ball is given by a dimension dependent constant $\omega_n = B(0, 1)$.

If $f, g : \mathbb{N} \to (0, \infty)$ are two functions, we write $f(n) = O(g(n))$ if there exists a constant C and an $n_0 \in \mathbb{N}$ such that $f(n) \leq Cg(n)$ for all $n \geq n_0$.

1.3 A bit on vector spaces

A (real) vector space V is a space together with two operations: addition $v + w$, $v, w \in V$ and scalar multiplication tv for $t \in \mathbb{R}, v \in V$. These satisfy the following axioms.

1. Commutativity and Distributivity: $v + w = w + v$, $u + (v + w) = (u + v) + w$, for all $u, v, w \in V$
2. Zero: $\exists 0 \in V : 0 + v = v$ for all $v \in V$
3. Negation: For all $v \in V$ exists $-v \in V : v + (-v) = 0$.
4. Distributivity of scalar multiplication: $(a + b)v = av + bv, a(v + w) = av + aw$ for all $v, w \in V, a, b \in \mathbb{R}$.

For most purposes $V = \mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}, i = 1, \ldots, n\}$, with componentwise addition and multiplication, suffices for this course, although we will mention some infinite dimensional normed vector spaces.

1.4 A bit of probability

A probability space $(\Omega, \Sigma, \mathbb{P})$ is a *measure space* (Ω, Σ), where Σ is a σ-algebra of sets, and $\mathbb{P}$ is a probability measure. For most of the analysis, it will suffice to consider Ω a finite set, Σ the power set of Ω and $\mathbb{P}$ some discrete probability measure on Ω. However, we will also use a bit of Gaussian random variables, and there the continuous framework is necessary. I will not review this here, but any standard text on probability theory will suffice.

A random variable $X : \Omega \to \mathbb{R}$ is measurable, if $X^{-1}(O) \in \Sigma$ for every open set $O \subset \mathbb{R}$. For such variables, we can define integrals, which are also called expected values,

$$\int X d\mathbb{P} = \mathbb{E}(X),$$

which exist provided that

$$\int |X| d\mathbb{P} < \infty.$$

For non-negative functions, the integrals of measurable functions can always be defined and may be infinite.

We say that A and B are independent events, if $\mathbb{P}(A \cap P) = \mathbb{P}(A)\mathbb{P}(B)$. We say that X, Y are independent random variables, if for all open sets A, B the events $\{X \in A\} = X^{-1}(A)$ and $\{X \in B\} = X^{-1}(B)$ are independent. We say that X and Y are identically distributed, if for all open sets A we have $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$. This guarantees, that $\mathbb{E}(X) = \mathbb{E}(Y)$, provided that at least one of these expected values exists.

If $X_1, \ldots, X_n$ are n independent identically distributed random variables, and if they are all distributed as X, for which $\mathbb{E}(X)$ exists, then

$$\mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\mathbb{E}(X).$$

We thus expect, $\sum_{i=1}^n X_i$ to be close to $n\mathbb{E}(X)$, since $\mathbb{E}(X)$ corresponds to the "expected behavior of X". Indeed, this can be proven and is the content of the Law of large numbers. What we will need is the following bound.

Lemma 1.1. *Let $X_1, \dots, X_m$ be identical, idependently distributed random variables with values in $[0,1]$. Then, for every $\epsilon > 0$, we have*

$$\mathbb{P}(X \leq (1-\epsilon)\mu m) \leq e^{-\epsilon^2 \mu m/2}.$$

2 What? - Lecture 1

This is a course about embeddings of metric spaces $f : X \to Y$, where X is a metric space and Y is a normed vector space. We have already touched upon the motivation to these in the general lecture, and will not re-iterate it right away. Instead, let's dive to the objects of interest.

2.1 Metric spaces

The domains X in our consideration are metric spaces.

Definition 2.1. *A metric space (X, d) is a set X together with a distance function $d : X \times X \to [0, \infty)$, which satisfies*

1. *symmetry: $d(x, y) = d(y, x)$,*
2. *positivity: $d(x, y) = 0 \iff x = y$,*
3. *the triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.*

Elements $x \in X$ of a metric space are also called points. The function d is called a metric. The set X may be either finite, or infinite. It's size is denoted by $|X|$. Balls in metric spaces are denoted $B(x, r) = \{y \in X : d(x, y) < r\}$ for $x \in X$ and $r > 0$. If $A \subset X$, we write $\mathrm{diam}(A) = \sup_{a,b \in A} d(a, b)$ for its diameter.

Here are a few examples of finite metric spaces.

1. $X = \{0,1\}^n$, and $d(w, v)$ is *the Hamming distance* between $w = (w_1, \dots, w_n), v = (v_1, \dots, v_n) \in X$, that is:
$$d(w, v) = \sum_{i=1}^n |w_i - v_i|,$$
which equals the number k of bits where w and v differ from each other. This metric arises in the context of error correcting codes, since $d(w, v)$ measures how many modifications one needs to perform to get from w to v.

2. *A graph* $G = (V, E)$ consists of a finite set of vertices V and a set $E \subset \{\{v, w\} : v, w \in V, v \neq w\}$ of edges. We denote $\{v, w\} = e_{v,w}$, and we say that v is an end point of $e \in E$ if $v \in e$. We will focus on graphs without an orientation, although much of what we say is also applicable to oriented graphs.

   An edge walk in a graph is a sequence of edges $(e_1, \ldots, e_n)$ with e_i and e_{i+1} sharing an end point. We say that the edge path $(e_1, \ldots, e_n)$ connects v to w, where $v, w \in V$, if v is an end point of e_1 and w an end point of e_n. The parameter n is the length of the edge path. We say that a graph G is *connected* if every pair of vertices can be connected by an edge path. In a connected graph, we can define the (unweighted) graph distance as the shortest length of an edge path connecting pairs of points:

$$d(v, w) = \min\{n : (e_1, \ldots, e_n) \text{ is an edge path connecting } v \text{ to } w\}.$$

   To increase generality, we attach for associate edge e a weight $c(e) > 0$. We then define the (weighted) graph metric as

$$d_c(v, w) = \min\{\sum_{i=1}^{n} c(e_i) : (e_1, \ldots, e_n) \text{ is an edge path connecting } v \text{ to } w\}.$$

   Remark 2.2. The metric d_c is not substantially more general than the metric d. Indeed, by modifying and scaling, we can convert d_c to d. Indeed, if $c(e) \in \mathbb{N}$ for $e \in E$, then we can form a new graph G_c which is obtained by subdividing each edge $e \in E$ by $c(e)$. Then $d_c(v, w)$, for $v, w \in V$ equals the (unweighted) graph distance in this new graph. On the other hand, if $c(e) \in \mathbb{Q}$ for all $e \in E$, then $Md_c(v, w) = d_{Mc}d(v, w)$ for M the least common multiple of all denominators of $c(e)$. By the first part, $d_{Mc}d(v, w)$ is an unweighted graph distance.

   Thus, mathematically speaking d_c is not more general than d. However, in many cases, d_c is convenient to use. Further, in actual computations it is likely easier to incorporate a weight, than to increase the complexity of the graph by subdivision.

3. *A complete graph* $G = (V, E)$ is a graph, where all pairs of vertices are connected by an edge, that is $E = \{\{v, w\} : v, w \in V, v \neq w\}$. A complete graph is clearly connected, and $d(v, w) = 1$ for the unweighted distance d for all $v, w \in V$. If further (V, d_V) is a metric space and we define a weight $c(e_{v,w}) = d_V(v, w)$ for all $v, w \in V$, then

$$d_c = d_V.$$

   That is, all finite metric spaces can be thought of as graph metrics.

   Remark 2.3. The complexity of storing a very large graph or metric space can be prohibitive. Indeed, if $|V| = n$, then $|E| \leq \frac{n(n-1)}{2}$. That is, we may need $O(n^2)$ storage for the graph. Similarly, to store the full distance function d of a space of size n, we need $O(n^2)$ bits. For small or moderate sized data this is fine. However, if say we consider the space of all websites, we have $n \sim 2 \cdot 10^9$, and $n^2 \sim 10^{18}$. While both numbers are large, the first is still reasonable (of the order 1 Gb, if we give each website its own codename), but the latter is quite far from reasonable (1 exabyte, or 10^6 terabytes). This is one of the reasons we may want to consider *efficiently* representing our data E in some (hopefully low dimensional) space.

There are a few standard constructions of metric spaces from other metric spaces.

Definition 2.4. *If $A \subset X$ is a subset of a metric space, then $(A, d|_{A\times A})$ is the restricted metric space. We often simplify notation and write d instead of $d|_{A\times A}$.*

If $(X, d_X), (Y, d_Y)$ are two metric spaces, their product $Z = X \times Y$ is equipped with the following product metric:

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

Further, there is a somewhat more complex modification of a metric. Let $\phi : [0, \infty) \to [0, \infty)$ be a non-negative concave function with $\phi(0) = 0$, and $\phi(t) > 0$ for all $t > 0$. Recall, that ϕ is concave, if for all $a, b \in [0, \infty)$ and all $t \in [0, 1]$, we have

$$\phi(ta + (1 - t)b) \geq t\phi(a) + (1 - t)\phi(b).$$

Another way to write this inequality is

$$\frac{\phi(\xi_t) - \phi(a)}{\xi_t - a} \geq \frac{\phi(b) - \phi(\xi_t)}{b - \xi_t}$$

whenever $\xi_t = ta + (1 - t)b$ for $t \in [0, 1]$. The point ξ_t thus represents an arbitrary point between a, b, and the inequality states that the slope of the secant line between a and ξ_t is greater than the slope of the secant line between ξ_t and b. If ϕ is differentiable, then this property is equivalent to the fact that ϕ' is increasing - as can be readily seen from the mean value property, since then ϕ' is increasing. Further, if ϕ is twice continuously differentiable, this property is equivalent to $\phi''(t) < 0$. The most important example of a concave function for us is $\phi(t) = t^\epsilon$ for any $\epsilon \in (0, 1)$, and the concavity can be directly seen by differentiation.

Problem 2.5. Prove that $t \to t^\epsilon$ is concave if and only if $t \in (0, 1]$.

Lemma 2.6. *Let (X, d) be a metric space. Then $(X, d_\phi = \phi(d))$ is also a metric space, where $d_\phi(x, y) = \phi(d(x, y))$.*

Proof. It is easy to check that $d_\phi(x, y) = d_\phi(y, x)$, $d_\phi(x, y) \geq 0$ and $d_\phi(x, y) = 0 \implies x = y$. These follow directly from their corresponding properties for the metric d. Thus, as usual, we are left to show the triangle inequality.

First, we show that

$$\phi(a + b) \leq \phi(a) + \phi(b)$$

for all $a, b > 0$. For simplicity, assume first $a < b$,. Consider the three secant lines over the intervals $[0, a]$, $[a, b]$ and $[b, a + b]$. The slopes are increasing by concavity:

$$\frac{\phi(a)}{a} = \frac{\phi(a)-\phi(0)}{a} \leq \frac{\phi(b)-\phi(a)}{b-a} \leq \frac{\phi(b+a)-\phi(b)}{a}.$$

Now multiplying by a and moving terms, we get the desired inequality. If $a = b$ we can compare directly the slope over $[0,a]$ and $[b,a+b] = [a,2a]$, and get the same result.

Now, the triangle inequality readily follows. If $x,y,z \in X$, then

$$\begin{aligned} d_\phi(x,y) = \phi(d(x,y)) &\leq \phi(d(x,z)+d(z,y)) \\ &= \phi(d(x,z)) + \phi(d(z,y)) \\ &= d_\phi(x,z) + d_\phi(z,y). \end{aligned}$$

In the second line we used that ϕ is increasing, and on the third line we used the inequality just derived. □

2.2 Normed spaces

The target spaces Y are vector spaces together with a norm.

Definition 2.7. *A normed space $(Y, \|\cdot\|)$ is a vector space Y together with a norm $v \to \|v\| \in [0,\infty)$ for which the following hold.*

1. *$\|v\| \in [0,\infty)$ for all $v \in Y$.*
2. *$\|tv\| = |t|\|v\|$ for all $v \in Y, t \in \mathbb{R}$.*
3. *$\|v\| = 0$ if and only if $v = 0$.*
4. *Triangle inequality: $\|v+w\| \leq \|v\| + \|w\|$.*

A normed space is also a metric space when equipped with the metric $d(v,w) = \|v-w\|$. Additionally, if Y is infinite dimensional, we will additionally assume that the normed space is complete. Completeness is equivalent to the metric space (Y,d) being complete, or the property that for all vectors $v_i \in Y$, $i \in \mathbb{N}$, we have

$$\sum_{i=0}^{\infty} \|v_i\| < \infty \Longrightarrow \sum_{i=0}^{\infty} v_i \text{ exists }.$$

For the most part, we will not need these properties, and thus leave them for now. Also, all of the spaces that we consider and present as examples are complete, and especially all finite dimensional spaces are complete. The main spaces that we consider, which also often appear in applications, are the ℓ_p spaces.

1. ℓ_p^n is the vector space $\mathbb{R}^n$ together with the norm

$$\|(x_1, \dots, x_n)\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}},$$

for $p \in [1, \infty)$, and

$$\|(x_1, \dots, x_n)\|_\infty = \max_{i=1,\dots,n} |x_i|.$$

2. At times it is easier to have infinitely many co-ordinates $\ell_p = \ell_p^\infty$ is the vector space

$$\ell_p^\infty = \ell_p = \{(x_1, x_2, \dots,) \in \mathbb{R}^{\mathbb{N}} : \|(x_1, x_2, \dots,)\|_p < \infty\{$$

where we equip the space with the norm (which is also part of the definition)

$$\|(x_1, \dots)\|_p = \left(\sum_{i=1}^\infty |x_i|^p \right)^{\frac{1}{p}},$$

for $p \in [1, \infty)$, and

$$\|(x_1, \dots)\|_\infty = \sup_{i \in \mathbb{N}} |x_i|.$$

3. It will also be useful to consider the following space. Let $(\Omega, \mathbb{P})$ be a probability space, and let

$$L^1(\Omega) = \{X : \Omega \to \mathbb{R} : X \text{ is measurable and } \int |X| d\mathbb{P} = \mathbb{E}(|X|) < \infty\}$$

equipped with the norm

$$\|X\|_1 = \int |X| d\mathbb{P} = \mathbb{E}(|X|).$$

4. There are many more complicated normed spaces, that one could consider. Consider, for example, M_n the space of $n \times m$ matrices, with $n < m$, and consider the nuclear norm:

$$\|M\|_n = \sum_{i=1}^n |\sigma_i(M)|,$$

where $\sigma_i(M)$ are the singular values of M. If, instead, we take an ℓ_p-norm of the singular values,

$$\|M\|_{p,s} = \|(\sigma_1(M), \dots, \sigma_n(M))\|_p,$$

we get the Shatten-norms on matrices.

Problem 2.8. Show that the nuclear norm is a norm. *Hint:* Use the fact that $\sum_{i=1}^n |\sigma_i(M)| = \sup_{U \in A} \text{trace}(MU)$, where A is the collection of $m \times n$ matrices with orthonormal rows, and trace(B) is the trace of a matrix B. Further, note that the supremum of a linear functions on a vector space is convex, and that this gives the triangle inequality. Asimilar expression can be written for other Shatten norms.

The $p = 2$ case of the theory above is quite special. It corresponds to Euclidean space, where associated to the norm we have a bi-linear inner product structure.

Definition 2.9. *An inner product space $(Y, \langle \cdot, \cdot \rangle)$ is a vector space Y together with an inner product $v, w \to \langle v, w \rangle \in \mathbb{R}$ for which the following hold.*

1. *Symmetry: $\langle v, w \rangle = \langle w, v \rangle$*
2. *Bilinearity: $\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$*
3. *$\langle v, v \rangle \in [0, \infty)$ for all $v \in Y$.*
4. *$\langle v, v \rangle = 0$ if and only if $v = 0$.*

Associated to an inner product, we have a norm $\|v\| = \sqrt{\langle v, v \rangle}$. This is a norm, and in proving so, one uses the Cauchy-Schwartz inequality:

$$|\langle v, w \rangle| \leq \|v\| \|w\|.$$

Every inner product space is isomorphic and isometric to ℓ_2^n or ℓ_2 for some $n < \infty$ - if Y is either finite dimensional or *separable.* An inner product for these spaces is given by the usual dot product:

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i y_i,$$

where the formula naturally extends to $n = \infty$. When we say inner product space, we mean either of these cases, and from now on, you can focus on the case ℓ_2^n. Below, we will also focus and write formulas for $n \in \mathbb{N}$ - but these calculations largely are identical for $n = \infty$.

On of the most important properties for us that distinguishes Euclidean space is the Parallelogram identity.

Lemma 2.10. *For all $a, b \in \ell_2^n$, we have*

$$\|a + b\|^2 + \|a - b\|^2 = 2(\|a\|^2 + \|b\|^2).$$

Proof. We have by bi-Linearity and symmetry

$$\begin{aligned}
\|a + b\|^2 + \|a - b\|^2 &= \langle a + b, a + b \rangle + \langle a - b, a - b \rangle \\
&= \langle a, a \rangle + 2\langle a, b \rangle + \langle b, b \rangle \\
&\quad + (\langle a, a \rangle - 2\langle a, b \rangle + \langle b, b \rangle) \\
&= 2\langle a, a \rangle + 2\langle b, b \rangle \\
&= 2(\|a\|^2 + \|b\|^2).
\end{aligned}$$

□

As a consequence of this, we get the following two geometric facts that concern midpoints. If X is a metric space, and $x, y \in X$, then a midpoint m between them, or for them, is a point for which $d(x, m) = d(y, m) = \frac{d(x,y)}{2}$.

Example 2.11. Let $X = \{0, 2, 3\}$, and equip X with the restricted metric from $\mathbb{R}$. Then the points $x = 0, y = 3$ do not have a midpoint in X.

Let $X = \{a, b, c, m\}$ and let $d(a, b) = d(b, c) = d(a, c) = 2$, and $d(z, m) = 1$ for all $z = a, b, c$. Them m is a midpoint for all the following pairs of points: (a, b), (b, c) and (a, c).

Let $X = \{0, 1\}^2$ be the Hamming square equipped with the Hamming metric. Then $x = (0, 0)$ and $y = (1, 1)$ have two mid-points $(1, 0)$ and $(0, 1)$. Higher dimensional Hamming cubes have even more midpoints.

Lemma 2.12. *Assume that $(Y, \langle\cdot\rangle)$ is an inner product space. Let $\epsilon > 0$. If $\|a\| = 1$ and $\|b\| = 1$ and $\|a - b\| \geq \epsilon > 0$, then*

$$\left\|\frac{a+b}{2}\right\| \leq \sqrt{1 - 4^{-1}\epsilon^2} \leq 1 - 2\epsilon^2.$$

Further, for every $a, b \in Y$ with $a \neq b$, there is a ***unique*** *midpoint m.*

Proof. From the parallelogram identity we get

$$\left\|\frac{a+b}{2}\right\|^2 + \left(\frac{\epsilon}{2}\right)^2 = \left\|\frac{a+b}{2}\right\|^2 + \left\|\frac{a-b}{2}\right\|^2 = \frac{\|a\|^2 + \|b\|^2}{2} = 1.$$

Solving from this, we get.

$$\left\|\frac{a+b}{2}\right\| \leq \sqrt{1 - 4^{-1}\epsilon^2} \leq 1 - 8^{-1}\epsilon^2.$$

We next prove uniqueness. By scale-invariance, we can assume that $\|a - b\| = 2$, and to simplify notation we also assume $a = 0$. It is clear that $m = \frac{b}{2}$ is a midpoint between a and b. Next, let m' be some potentially other midpoint. Then, we have $\|m\| = \|b - m\| = \|m'\| = \|b - m'\| = \frac{1}{2}\|b\| = 1$. If $\|m - m'\| = 0$, we are done. If not, then there is some $\epsilon > 0$ so that $\|m - m'\| > \epsilon$. By the first part of the proof, we get

$$\left\|\frac{m+m'}{2}\right\| \leq 1 - 2\epsilon^2 < 1,$$

and by the same argument applied to $b - m$ and $b - m'$,

$$\left\|b - \frac{m+m'}{2}\right\| = \left\|\frac{(b-m)+(b-m')}{2}\right\| \leq 1 - 2\epsilon^2 < 1,$$

Thus, we reach the following contradiction

$$2 = \|b\| \leq \left\|\frac{m+m'}{2}\right\| + \left\|b - \frac{m+m'}{2}\right\| < 2,$$

which implies that we must have $m = m'$ $\square$

Remark 2.13. For ℓ_p^n and ℓ_p-spaces the same uniqueness and an analogue of Lemma 3 holds for all $p \in (1,\infty)$. In the proof, parallelogram identity is replaced with so called *Clarkson's inequalities.*

In general, one can define the modulus of convexity of a normed space:

$$\omega(\delta) = 1 - \sup\{\|\frac{a+b}{2}\| : \|a\| = \|b\| = 1, \|a-b\| \geq \delta\}.$$

A normed space is said to be uniformly convex if $\omega(\delta) > 0$ for all $\delta > 0$. Using this, one can prove an analogue of the previous Lemma for uniformly convex spaces, and with the ϵ^2 term replaced with $\omega(\epsilon)$. By the previous paragraph, all ℓ_p spaces are uniformly convex when $p \in (1,\infty)$.

If $p = 1,\infty$, then mid-points are not unique, and the previous Lemma fails for ℓ_p^n and ℓ_p^∞.

Problem 2.14. Prove that if $p = 1,\infty$, then mid-points are not unique, and the previous Lemma fails for ℓ_p^n. Specifically, find an a, b with $\|a\| = \|b\| = 1$ and $\|a-b\| > 0$, but for which $\left\|\frac{a+b}{2}\right\| = 1$. Also, find a pair of points a, b with a non-unique mid-point. Does this hold for all pairs of points?

For general points $a, b, c, d \in \ell_2^n$ we have the following bound.

Lemma 2.15. *If $a, b, c, d \in \ell_2^n$, then*

$$\|a-c\|^2 + \|b-d\|^2 \leq \|a-b\|^2 + \|b-c\|^2 + \|c-d\|^2 + \|d-a\|^2$$

Proof. Group the terms

$$\begin{aligned}
&\|a-b\|^2 + \|b-c\|^2 + \|c-d\|^2 + \|d-a\|^2 - (\|a-c\|^2 + \|b-d\|^2) \\
&= \|a\|^2 + \|c\|^2 + 2\langle a, c\rangle - \langle a+c, b+d\rangle + \|b\|^2 + \|d\|^2 - 2\langle b, d\rangle \\
&= \|a+c\|^2 - 2\langle a+c, b+d\rangle + \|b+d\|^2 = \|a+c-b-d\|^2 \geq 0.
\end{aligned}$$

□

Problem 2.16. Show that equality holds in Lemma 2.15 if and only of a, b, c, d are the four corners of a parallelogram. Hint: Observe that the opposite sides $a-b$ and $c-d$ are equal, if the inequality is an equality.

2.3 Embeddings

We now put the two together. We will treat the target Y here as also a metric space, and use d for the metric on both X and Y.

Definition 2.17. *A mapping $f : X \to Y$ is called an isometric embedding, if*

$$d(f(x), f(y)) = d(x,y), \text{ for all } x, y \in X.$$

Let $L > 0$. A mapping $f : X \to Y$ between a metric space X and a metric space Y is ($L-$)Lipschitz, if

$$\frac{d(f(x), f(y))}{d(x,y)} \leq L, \text{ for all } x, y \in X, x \neq y.$$

The smallest L for which this inequality holds is denoted $\mathrm{LIP}\ (f)$.

Let $b > 0$. A ((b, L))-bi-Lipschitz embedding $f : X \to Y$ is a L-Lipschitz map, for which there exists a constant b for which also

$$b \leq \frac{d(f(x), f(y))}{d(x,y)} \leq L, \text{ for all } x, y \in X, x \neq y.$$

The largest constant b is also $\mathrm{LIP}\ (f^{-1})$, *where* $f^{-1} : \mathrm{Im}\ (f) \to X$.

We also say that f is (b, L)-biLipschitz or L-Lipschitz, if we wish

The distortion of a biLipschitz embedding $f : X \to Y$ is

$$D(f) = \frac{\mathrm{LIP}\ (f)}{\mathrm{LIP}\ (f^{-1})} = \inf \left\{ \frac{L}{b} : f \text{ is } (b, L) - \text{biLipschitz} \right\}.$$

It will often be useful to normalize either $b = 1$, when we say that f is *expanding* (since then $d(f(x), f(y)) \geq d(x,y)$ for all $x, y \in X$), or $L = 1$, when we say that f is *contracting* (since then $d(f(x), f(y)) \leq d(x,y)$ for all $x, y \in X$). In these cases, $D(f)$ is the smallest Lipschitz constant of an expanding biLipschitz map, or the reciprocal of the largest constant b for contracting biLipschitz maps f.

Isometric mappings have $D(f) = 1$. For all mappings $D(f) \geq 1$, and the distortion is a measurement of how far away f is from being an isometry. Already from Example 2.11 and Lemma 2.12, we see that the Hamming cube does not isometrically embed into Euclidean space, since Euclidean space has unique midpoints, which the Hamming cube does not have. We can actually say more.

Lemma 2.18. *Let $X = \{0,1\}^2$ be equipped with the Hamming metric. Then X does not isometrically embed to ℓ_2^n for any $n \in \mathbb{N}$, and $c_2(X) \geq \sqrt{2}$. Further, there exists a biLipschitz embedding $f : X \to \ell_2^2$ with $D(f) = \sqrt{2}$.*

Proof. Let f be a contracting $(b, 1)$-biLipschitz map. Let $a = f(0,0), b = f(0,1), c = f(1,0), d = f(1,1)$. We have by Lemma 2.15 that

$$\|a - c\|^2 + \|b - d\|^2 \leq \|a - b\|^2 + \|b - c\|^2 + \|c - d\|^2 + \|d - a\|^2.$$

We estimate the terms on the right by contraction and the terms on the left using the biLipschitz condition. We thus get

$$8b^2 \leq \|a - c\|^2 + \|b - d\|^2 \leq \|a - b\|^2 + \|b - c\|^2 + \|c - d\|^2 + \|d - a\|^2 \leq 4.$$

Solving from this $b \leq \sqrt{2}^{-1}$, and $D(f) \geq \sqrt{2}$. If we choose the "identity embeeding", $a = (0,0), b = (0,1), c = (1,0), d = (1,1)$, one can compute that f is 1-Lipschitz and $b = \sqrt{2}^{-1}$. (In that case, the inequality above becomes an equality.) $\square$

Motivated by the following, we look at the minimal possible distortion a metric space can be embedded into ℓ_p^n for some n.

$$c_p(X) = \inf\{D(f) : f : X \to \ell_p^n, n \in \mathbb{N} \cup \{\infty\}\}.$$

Using this notation, we can restate Lemma 2.18 as $c_2(\{0,1\}^2) = \sqrt{2}$.

Problem 2.19. Consider the second example from Example 2.11: $X = \{a, b, c, m\}$ and let $d(a, b) = d(b, c) = d(a, c) = 2$, and $d(z, m) = 1$ for all $z = a, b, c$. Determine $c_2(X)$. You can either do this directly, or use results from below.

Question 2.20. The big questions for us are the following.

1. How does $c_p(X)$ behave for a given X in terms of p?
2. Can we bound $c_p(X)$ in terms of the size of X?
3. Can $c_p(X)$ be infinite?

It should be noted at this juncture, that these questions are generally very difficult. Further, while much is known about $c_p(X)$, there are many questions which remain open. We will mention some below.

2.4 Lower bounds for $c_2(X)$

We start to look at the problem of how $c_2(X)$ behaves in terms of the size of X. We will work towards showing that $c_2(X) = O(\log(|X|)$. The following example shows a slightly weaker lower bound. (It is known that $c_2(X) = \Omega(\log(|X|)$ for some more involved metric spaces X.)

We will consider the Hamming cube $X = \{0,1\}^n$. It will be useful to consider X as a graph, where the edges consist of pairs $\{a, b\}$ with $d(a, b) = 1$ - i.e. all points which differ in only one component. It is direct to see that the graph metric is equal to the Hamming distance.

Lemma 2.21. *Let $X = \{0,1\}^n$ be equipped with the Hamming distance. Then, $c_2(X) = \sqrt{n}$.*

Proof. That $c_2(X) \leq \sqrt{n}$ follows by embedding X with the natural inclusion to ℓ_2.

First, we state a general inequality. Let $E \subset \{\{a, b\} : d(a, b) = 1\}$ be the collection of edges of X, and let $D = \{\{a, b\} : d(a, b) = n\}$ be the collection of all "long" diagonals, i.e. where each component is distinct – i.e. all pairs (a, b) where b is obtained by flipping each component of a. Let $x_a \in \ell_2^d$ be any vectors for $a \in X$. A generalization of Lemma 2.15 is the following.

$$\sum_{\{a,b\}\in D} \|x_a - x_b\|^2 \leq \sum_{\{a,b\}\in E} \|x_a - x_b\|^2. \tag{1}$$

If we choose $x_a = f(a)$, and if f is $(1, L)$-biLipschitz, we get

$$\begin{aligned} 2^{n-1}n^2 = \sum_{\{a,b\}\in D} (d(a,b))^2 &\leq \sum_{\{a,b\}\in D} \|f(a) - f(b)\|^2 \\ &\leq \sum_{\{a,b\}\in E} \|f(a) - f(b)\|^2 \leq \sum_{\{a,b\}\in E} L^2 = 2^{n-1}nL^2. \end{aligned}$$

Thus,

$$n \leq L^2,$$

and $L \geq \sqrt{n}$, which yields the desired distortion bound.

Thus, we are left to show (1). There are several proofs of this inequality – a Fourier analytic one, an induction proof, and one proving the positive definiteness of a given matrix. While there are advantages to each, for saving time, we shall do the induction proof.

Lets prove (1). We will prove it by induction on n. The base case $n = 2$ is the usual Parallelogram inequality. Now, suppose the claim has been shown for $n - 1$ and $n > 2$. Consider $X_{n-1}^0 = \{0,1\}^{n-1} \times \{0\}$ and $X_{n-1}^1 = \{0,1\}^{n-1} \times \{1\}$. These are two subsets of X_n differentiated by their last co-ordinate. Let also D_0, D_1, E_0, E_1 be the set of diagonals and edges of X_{n-1}^0 and X_{n-1}^1, which have been naturally identified with $\{0,1\}^{n-1}$. We have for $i = 0, 1$

$$\sum_{\{a,b\}\in D_i} \|x_a - x_b\|^2 \leq \sum_{(a,b)\in E_i} \|x_a - x_b\|^2. \tag{2}$$

Consider a pair of points $p, q \in \{0,1\}^{n-1}$ where q is obtained from p by flipping each bit. Next, $p0$ means the element in X obtained by adding the bit 0 as the last coordinate, and other notation is similar. Then, q is obtained from p by flipping each bit. Then, $\{p0, q0\} \in D_0$, $\{p1, q1\} \in D_1$ and $\{p0, q1\}, \{p1, q0\} \in D$. We get from Lemma 2.15, applied to the parallelogram $p0, p1, q1, q0$, that for all $p \in \{0,1\}^{n-1}$

$$\|x_{p0} - x_{q1}\|^2 + \|x_{p1} - x_{q0}\|^2 \leq \|x_{p0} - x_{q0}\|^2 + \|x_{p1} - x_{q1}\|^2 + \|x_{p0} - x_{p1}\|^2 + \|x_{q0} - x_{q1}\|^2.$$

Summing these over all p (and dividing by two, to avoid duplication) yields all the diagonals of X on the left hand side:

$$\sum_{\{a,b\}\in D} \|x_a - x_b\|^2 \leq \frac{1}{2} \sum_{p\in\{0,1\}^{n-1}} \|x_{p0} - x_{q0}\|^2 + \|x_{p1} - x_{q1}\|^2 + \|x_{p0} - x_{p1}\|^2 + \|x_{q0} - x_{q1}\|^2.$$

The first and second terms in the sum correspond to the diagonals D_0 and D_1, while the third and fourth term yield the same sum. Thus

$$\sum_{\{a,b\}\in D} \|x_a - x_b\|^2 \leq \sum_{\{a,b\}\in D_0} \|x_a - x_b\|^2 + \sum_{\{a,b\}\in D_1} \|x_a - x_b\|^2 + \sum_{p\in\{0,1\}^{n-1}} \|x_{p0} - x_{p1}\|^2.$$

Now, finally applying the induction hypothesis, we get

$$\sum_{\{a,b\}\in D} \|x_a - x_b\|^2 \leq \sum_{\{a,b\}\in E_0} \|x_a - x_b\|^2 + \sum_{\{a,b\}\in E_1} \|x_a - x_b\|^2 + \sum_{p\in\{0,1\}^{n-1}} \|x_{p0} - x_{p1}\|^2 = \sum_{\{a,b\}\in E} \|x_a - x_b\|^2,$$

where we recognized the final sum as the sum over all edges. $\square$

3 How? - Lecture 2

We will now look at some positive results on embeddings. The idea in both of these will be that Fréchet embeddings are "the only" way to embed. While there certainly have appeared many other embeddings, and these are not the only ones used, they are the only ones that are broadly applicable to the setting of all metric spaces. Thus, a good heuristic in constructing embeddings is: try distance embeddings. If it doesn't work, then likely something is causing problems.

3.1 Kuratowski embedding

We have observed that not all subsets embed to ℓ_2^n. However, the story with ℓ_∞ is quite different.

Theorem 3.1. *(Kuratowski embedding) We have*

$$c_\infty(X) = 1.$$

Indeed, every finite metric space X embeds isometrically to ℓ_∞^n for $n = |X|$.

Proof. The proof uses a very simple type of embedding, that we will also need in the second lecture.

Definition 3.2. *A Frechét embedding, or a distance embedding is an embedding $f : X \to \ell_p^k$ of the following form. Let $S_1, \ldots, S_k$ be subsets of X, and let*

$$f(x) = (d(x, S_1), \ldots, d(x, S_k)),$$

where $d(x, A) = \min\{d(x, a) : a \in A\}$ for a subset $A \subset X$.

Index the elements of $X = \{x_1, \ldots, x_n\}$, and let

$$f(x) = (d(x, x_1), \ldots, d(x, x_n)) \in \ell_\infty^n.$$

We have

$$d(f(x), f(y)) = \max_i |d(x, x_i) - d(y, x_i)|.$$

Now, we have by the reverse triangle inequality:

$$|d(x, x_i) - d(y, x_i)| \leq d(y, x)$$

for all $i = 1, \ldots, n$. If we choose i so that $x_i = x$, then

$$|d(x, x_i) - d(y, x_i)| = d(y, x).$$

Thus,

$$d(f(x), f(y)) = \max_i |d(x, x_i) - d(y, x_i)| = d(x, y),$$

and we see that f is an isometric embedding. Therefore, $D(f) = 1$, and $c_\infty(X) = 1$. $\square$

3.2 Bourgain embedding

Theorem 3.3. *We have*

$$c_2(X) = O(\log(|X|))$$

for all metric spaces X. In fact

$$c_p(X) = O(\log(|X|))$$

for all $p \in [1, \infty)$.

The proof of the theorem also shows that we only need $O(\log(|X|)^2)$ many coordinates. We will follow the original proof of Bourgain. The embedding will be of the form

$$f(x) = (d(x, S_{1,1}), \ldots, d(x, S_{N,M})),$$

for a given N, M, and sets $S_{i,j}$ for $i = 1, \ldots, k, j = 1, \ldots, M$.

In principle, it is possible to construct the sets S_i explicitly. However, this is quite tedious, and the involved randomized construction will be quite a bit simpler to analyse. We will construct the sets S_i via random sampling of the space. The idea is the following: we show that with positive probability the mapping f has the desired distortion. This is the power of the "probabilistic method", which is quite often used for combinatorial constructions and to prove existence of objects.

Let $k = \lceil \log_2(|X|) \rceil$, for which

$$2^{k-1} < |X| \leq 2^k.$$

Now,

$$S_{i,j} = \text{ random set where each } v \in S_{i,j} \text{ with probability } 2^{-i}.$$

and $j = 1, \ldots, M$, where $M = Ck$ for some constant C. The idea is to easily capture information about the space at *size* scales 2^{-i}. You can imagine X being partitioned into sets of size 2^i. Then $S_{i,j}$ are sets which intersect each element in the partition roughly once.

Proof of the Upper Lipschitz bound. Using Hölder's inequality, we get for $f : X \to \ell_p$ the following bound.

$$\begin{aligned} |f(x) - f(y)| &= \left(\sum_{i,j} |d(x, S_{i,j}) - d(y, S_{i,j})|^p \right)^{\frac{1}{p}} \\ &\leq \left(\sum_{i,j} d(x, y)^p \right)^{\frac{1}{p}} = (Ck)^{\frac{2}{p}}. \end{aligned}$$

□

Next, we reduce the problem to $p = 1$.

Enough to prove for $p = 1$. We have the reverse triangle inequality that

$$\begin{aligned}|f(x) - f(y)| &= \left(\sum_{i,j} (d(x, S_{i,j}) - d(y, S_{i,j})^p\right)^{\frac{1}{p}} \\ &\geq \sum_{i,j} |d(x, S_{i,j}) - d(y, S_{i,j})| k^{2\frac{1-p}{p}}\end{aligned}$$

Thus, if we can show that

$$\sum_{i,j} |d(x, S_{i,j}) - d(y, S_{i,j})| \geq \frac{k}{4},$$

we then get

$$\frac{L}{b} = \frac{(Ck)^{\frac{2}{p}}}{\frac{k}{4} k^{2\frac{1-p}{p}}} \sim k = O(\log(|X|)).$$

The claim has thus been shown. $\square$

Lower bound proof. First fix $x, y \in X$. We will compute the probability that we have the desired lower bound for x, y. What we want to do is estimate $|d(x, S_{i,j}) - d(y, S_{i,j})|$ from below. This involves showing that $S_{i,j}$ contains points close to x, but not close to y with some definite probability. For this proof, let $\overline{B}(p, r) = \{q : d(q, p) \leq r\}$.

Choose $r_t = d(x, y)/4$, and let t be the largest integer so that

$$|\overline{B}(x, r_t)| \geq 2^t, \text{ and } |\overline{B}(y, r_t)| \geq 2^t$$

Define for $l = 0, \ldots, t-1$ the radii

$$r_l = \min\{r : |\overline{B}(x, r)| \geq 2^l, |\overline{B}(y, r)| \geq 2^l\}.$$

With this choice, $r_0 = 0$, r_l is an increasing sequence, and

$$\sum_{l=0}^{t-1} r_{l+1} - r_l = r_t - r_0 = \frac{d(x, y)}{4}.$$

Consider the good event $G_{i,j}$ of when

$$S_{i,j} \cap \overline{B}(x, r_i) \neq \emptyset \text{ and } S_{i,j} \cap B(y, r_{i+1}) = \emptyset.$$

This is good, because, when $G_{i,j}$ happens, then

$$|d(x, S_{i,j}) - d(y, S_{i,j})| \geq r_{i+1} - r_i.$$

The good news is that the good event has positive probability.

Lemma 3.4. *There is a constant $\delta > 0$ (independent of x, y) so that for $i = 1, \dots, t-1$ and any $j = 1, \dots, Ck$ we have*

$$\mathbb{P}(G_{i,j}) \geq \delta.$$

Proof. The balls $\overline{B}(x, r_i)$ and $B(y, r_{i+1})$ are disjoint, and thus, whether $A_{i,j}$ chooses a point in them are independent event, thus

$$\begin{aligned}\mathbb{P}(G_{i,j}) &= \mathbb{P}(S_{i,j} \cap \overline{B}(x, r_i) \neq \emptyset)\mathbb{P}(S_{i,j} \cap B(y, r_{i+1}) = \emptyset) \\ &\geq \left(1 - (1 - 2^{-i})^{2^i}\right)\left((1 - 2^{-i})^{2^{i+1}}\right).\end{aligned}$$

Now,

$$1 - x \leq e^x$$

for all x, since e^x is convex, and

$$1 - \frac{x}{2} \geq 1 - x + \frac{1}{2}x^2 \geq e^{-x}$$

for all $x \in [0, 1]$, by using the alternation of the Taylor series. Thus

$$\left(1 - (1 - 2^{-i})^{2^i}\right)\left((1 - 2^{-i})^{2^{i+1}}\right) \geq \left(1 - e^{-1}\right) e^{-2^{1-i}2^{i+1}} = e^{-2} - e^{-3} = \delta.$$

□

We can now fix $C = \lceil 100\delta^{-1} \rceil$. Next, let G_i be the event that for k many indices $j = 1, \dots, Ck$ the event $G_{i,j}$ happens. We expect at least $\delta Ck \geq 100k$ events to happen. Let B_i be the event that G_i does not happen, that is, that less than k events $G_{i,j}$ occur. By Lemma 1.1 with $\epsilon = 1/2$, we get

$$\mathbb{P}(B_j) \leq e^{-C\delta k/8} \leq e^{-10k} \leq n^{-10}.$$

Each of these is indepenent, and we can compute

$$\mathbb{P}(G_i) \geq 1 - n^{-10}.$$

The probability that all G_i occur, is $1 - kn^{-10}$. When all G_i occur, we get

$$\sum_{i,j} |d(x, S_{i,j}) - d(y, S_{i,j})| \geq \sum_{i=0}^{t-1} k(r_{i+1} - r_i) \geq k\frac{d(x, y)}{4}. \tag{3}$$

Call G_{xy} the event that all G_i occurs. Now, this good event, G_{xy} has probability at least $1 - kn^{-10}$. The probability that for some x, y we have that $G_{x,y}$ fails is at most

$$kn^{-10}n^2 = kn^{-8} < 1,$$

and thus with positive probability all G_{xy} succeed. This means, that with positive probability, f satisfies the bound (3) for all $x, y \in X$. This was the desired bound. □

4 How? Continued: Assouad embedding

4.1 Example

Let us start with a simple example. Consider the metric space

$$[0, 1], d(x, y) = |x - y|^\epsilon,$$

for some $\epsilon \in (0, 1)$. This is called the snow-flake curve.

Let $f_k(x) = (\sin(2^k x)2^{-k}, \cos(2^k x)2^{-k}) \in \mathbb{R}^2$, and let

$$F(x) = (f_0(x), 2^{(1-\epsilon)}f_1(x), 2^{2(1-\epsilon)}f_2(x), \cdots, f_k(x)2^{k(1-\epsilon)}, \cdots).$$

We claim that this is a biLipschitz embedding.

Lemma 4.1. *The map F is bi-Lipschitz from $([0, 1], |\cdot|^\epsilon)$ to ℓ_2.*

Proof. First, we prove the Lipschitz bound. We have

$$|\sin(x) - \sin(y)| \leq |x - y|, |\cos(x) - \cos(y)| \leq |x - y| \tag{4}$$

Let $x, y \in [0, 1]$. Choose k so that $2^{-k-1} < |x - y| \leq 2^{-k}$. For $i \leq k$, we have by (4)

$$|f_i(x) - f_i(y)| \leq 2|x - y|.$$

For $k > i$, we have for all $x \in [0, 1]$ that

$$|f_k(x)| \leq 22^{-k}.$$

Thus,

$$\begin{aligned}
|F(x) - F(y)| &= \sqrt{\sum_{i=0}^{\infty} \left(2^{(1-\epsilon)i}|f_i(x) - f_i(x)|\right)^2} \\
&= \sqrt{\sum_{i=0}^{k} \left(2^{(1-\epsilon)i}|f_i(x) - f_i(x)|\right)^2 + \sum_{i=k+1}^{\infty} \left(2^{(1-\epsilon)i}|f_i(x) - f_i(x)|\right)^2} \\
&\leq \sqrt{2\sum_{i=0}^{k} 2^{2(1-\epsilon)i}|x - y|^2 + \sum_{i=k+1}^{\infty} 2^4 2^{-2i\epsilon}} \\
&\leq \sqrt{\frac{2}{2^{2(1-\epsilon)} - 1}2^{2(k+1)(1-\epsilon)}|x - y|^2 + \frac{2^4}{1 - 2^{-\epsilon}}2^{-2(k+1)\epsilon}} \\
&\lesssim 2^{-k\epsilon} \lesssim |x - y|^\epsilon.
\end{aligned}$$

This gives the upper Lipschitz bound.

Next, we work on the lower bound. The idea is that the k'th term takes care of the points x, y with distance to each other roughly 2^{-k}.

$$\begin{aligned}
|F(x) - F(y)| &\geq \sqrt{|f_k(x) - f_k(y)|^2} \\
&= 2^{-k\epsilon}\sqrt{|\sin(2^k x) - \sin(2^k y)|^2 + |\cos(2^k x) - \cos(2^k y)|^2} \\
&= 2^{-k\epsilon}\sqrt{2 - 2\cos(2^k(x-y))} \\
&\gtrsim |x-y|^\epsilon,
\end{aligned}$$

since $2^k(x-y) \in [\frac{1}{2}, 1]$, and thus $\cos(2^k(x-y)) < \cos(1/2) < 1$. □

The idea of the previous embedding is that each component f_k is an embedding for pairs of points x, y with distance comparable to 2^{-k}. We then simply combine these embeddings by adding them together. The same proof idea is true for the general Assouad embedding theorem, which we shall focus on next.

4.2 Metric doubling

The lower bound from biLipschitz embeddings is loosely connected to the problem of packing balls in Euclidean space. The problem of optimally packing balls is difficult, but we can give the following very weak bound.

Lemma 4.2. *Let $r < R$ and let $S \subset B(0,R) \subset \ell_2^n$ be a set of points for which each $s, t \in S$ with $s \neq t$ we have $d(s,t) \geq r$. We have*

$$|S| \leq \frac{4^n R^n}{r^n}.$$

On the other hand, there exists a set $S \subset B(0,R)$ so that for each $s, t \in S$ with $s \neq t$ we have $d(s,t) \geq r$, and for which

$$|S| \geq \frac{R^n}{r^n}.$$

Proof. For each $s \in S$ consider the ball $B_s = B(s, r/2)$. These balls are disjoint. Their number can be estimated using a volume argument.

$$\begin{aligned}
\omega_n 2^n R^n = |B(0,2R)| &\geq \left|\bigcup_{s\in S} B_s\right| \\
&\geq \sum_{s\in S} |B_s| = |S|\omega_n r^n 2^{-n}.
\end{aligned}$$

This gives the desired estimate by dividing both sides by the extra constants.

The lower bound can be done in many ways. Let S be a maximal subset of $B(0,R)$ with the stated property. That is, one can not add any more points to S from $B(0,R)$ without violating the distance separation condition. Given this, we must have

$$B(0,R) \subset \bigcup_{s\in S} B_s. \tag{5}$$

Indeed, if this subset relation were to fail, we could find some $t \in B(0,R)$ with $t \notin B_s$ for any $s \in S$. Thus, by definition of balls, $d(s,t) \geq r$ for all $s \in S$. Consequently, $S \cup \{t\}$ would be a larger set, which contradicts maximality.

A minor point to address is the existence of such a maximal set. In this case, we can argue it algorithmically. Let $S_0 = \emptyset$. Proceed recursively and set $i = 0$. While the set S_i is not maximal, we add a point t so that $d(t,s) \geq r$ for all $s \in S_i$. That is, $S_{i+1} = S_i \cup \{t\}$, and we increment i. We can only repeat this for finitely many steps, since each time the set becomes bigger and the first part of the proof states a size bound for such sets.

Now, given the maximal set S which is output by this algorithm, we compute volumes again. By (5) and sub-additivity of volumes, we get

$$\omega_n R^n = |B(0,R)| \leq \sum_{s\in S} |B_s| \leq |S|\omega_n r^n.$$

This yields the lower bound for the size of S. $\square$

We can use this to derive a distortion bound for embedding some graphs in Euclidean space. Let K_n be the complete graph equipped with the path metric. That is, $d(v,w) = 1$ for each distinct $v, w \in K_n$.

Lemma 4.3. *We have $D(f) \geq n^{\frac{1}{d}}4^{-1}$ for all biLipschitz mappings $f : K_n \to \mathbb{R}^d$. Also, $D(f) = 1$ if $d \geq n-1$.*

Proof. Suppose that f is distance expanding, and that $f(v_0) = 0$ for some $v_0 \in K_n$. Then, let $S = \{f(x) : x \in K_n\}$. If L is the Lipschitz constant of f, then $S \subset B(0,L)$, and by Lemma 4.2 applied with $r = 1$ and $R = L$, we get

$$n = |S| \leq 4^d L^d.$$

Thus, $L \geq n^{\frac{1}{d}}4^{-1}$. This yields the lower bound for the distortion.

When $d \geq n-1$, there exists an isometric embedding. Indeed, enumerate the vertices of K_n as $v_1, \ldots, v_n$. let first $g : K_n \to \mathbb{R}^n$ be given by $g(v_i) = \frac{1}{\sqrt{2}}e_i$. The image of g is contained in an affine subspace of dimension $n-1$, and the isometric embedding into $\mathbb{R}^{n-1}$ is obtained by mapping that subspace isometrically to $\mathbb{R}^{n-1}$. $\square$

Problem 4.4. Show that the bound for distortion is essentially sharp by giving a comparable upper bound.

Definition 4.5. *A metric space (X,d) is said to be N-metric doubling, if for every $x \in X$ and every $r > 0$, there exist points $x_1, \dots, x_N \in X$ for which*

$$B(x,r) \subset \bigcup_{i=1}^{N} B(x_i, r/2).$$

The constant $D = \lceil \log_2 N \rceil$ is called the doubling dimension of the metric space.

Problem 4.6. Show, using the proof of Lemma 4.2, that $\mathbb{R}^n$ is 8^n-doubling.

Problem 4.7. Show, that if X is metric N-doubling, then for every $x \in X$, and $r > 0$, and any $k \in \mathbb{N}$, there exist N^k many points $x_1, \dots, x_{N^k} \in X$ for which

$$B(x,r) \subset \bigcup_{i=1}^{N^k} B(x_i, r/2^k).$$

Lemma 4.8. *Let $f : X \to \mathbb{R}^n$ has distortion $D(f)$. Then X is D-doubling with $D \leq D(f)^n 32^n$.*

Proof. Assume that f is normalized so that

$$d(x,y) \leq |f(x) - f(y)| \leq D(f) d(x,y)$$

for all $x, y \in X$.

Let $B(x,R) \subset X$ be any ball. Consider the image

$$f(B(x,R)) \subset B(f(x), D(f)R).$$

Cover $B(f(x), D(f)R)$ by the smallest number of balls of radius $R/4$, $B(y_i, R/8)$, $i \in I$. We have $|I| \leq D(f)^n 32^n$. Let $J \subset I$ be the set of $i \in I$ so that $B(y_i, R/8) \cap f(B(x,R)) \neq \emptyset$. For each $j \in J$ choose $z_j \in B(y_j, R/8) \cap f(B(x,R))$. Since $z_j \in f(B(x,R))$, there exists an $x_j \in B(x,R)$ so that $f(x_j) = z_j$.

Now, we have

$$f^{-1}(B(z_j, R/4)) \subset B(x_j, R/2).$$

Indeed, if $d(y, x_j) \geq R/2$, then $d(f(y), d(x_j)) > R/2$, and thus

$$d(f(y), f(z_j)) > d(f(y), d(x_j)) - d(f(z-j), d(x_j)) \geq \frac{R}{4}.$$

Since

$$f(B(x,R)) \subset B(f(x), D(f)R) \subset \bigcup_{i \in I} B(y_i, R/8) \subset \bigcup_{i \in I} B(z_i, R/4),$$

we get

$$B(x,R) \subset \bigcup_{i \in J} f^{-1}(B(z_i, R/4)) \subset \bigcup_{i \in J} B(x_j, R/2).$$

Since $|J| \leq |I| \leq D(f)^n 32^n$ we obtain the desired doubling bound. □

4.3 Embedding result

Theorem 4.9. *If X is metric D doubling and* $\operatorname{diam}(X) < 1$, *then* (X, d^ϵ) *biLipschitz embeds to ℓ_2 for all $\epsilon \in (0,1)$.*

Remark 4.10. We shall see that Assouad embedding is a sharp statement. Indeed, we will construct a sequence of metric spaces X_n which are all D-doubling, but for which $c_2(X_n) \to \infty$.

Question 4.11. A very famous open problem of Lang and Plaut asks the following. If X is metric doubling, and if $c_2(X) < \infty$, then does there exist a dimension N so that X biLipschitz embededs into ℓ_2^N and N is controlled by the doubling constant D.

Proof of Theorem 4.9. We shall construct mappings $f_k : X \to \mathbb{R}^n$ which satisfy the following, for some constants δ, C.

1. If $d(x,y) \in [2^{-k-1}, 2^k]$, then $|f_k(x) - f_k(y)| \geq \delta d(x,y)$.
2. $\sup_{x \in X} |f_k(x)| \leq C 2^{-k}$.
3. The mapping f_k is C-Lipschitz.

These will be constructed in Lemma 4.19 after the proof.

If we succeed, then the Theorem is shown with the embedding

$$F(x) = (f_0, 2^{(1-\epsilon)} f_1, 2^{2(1-\epsilon)} f_2, 2^{3(1-\epsilon)} f_3, \dots).$$

The upper Lipschitz bound is computed exactly the same as before in our example:

$$\begin{aligned}
|F(x) - F(y)| &= \sqrt{\sum_{i=0}^{\infty} \left(2^{(1-\epsilon)i} |f_i(x) - f_i(x)|\right)^2} \\
&= \sqrt{\sum_{i=0}^{k} \left(2^{(1-\epsilon)i} |f_i(x) - f_i(x)|\right)^2 + \sum_{i=k+1}^{\infty} \left(2^{(1-\epsilon)i} |f_i(x) - f_i(x)|\right)^2} \\
&\leq \sqrt{C^2 \sum_{i=0}^{k} 2^{2(1-\epsilon)i} |x-y|^2 + \sum_{i=k+1}^{\infty} C^2 2^{-2i\epsilon}} \\
&\leq C \sqrt{\frac{2}{2^{2(1-\epsilon)} - 1} 2^{2(k+1)(1-\epsilon)} |x-y|^2 + \frac{2^4}{1 - 2^{-\epsilon}} 2^{-2(k+1)\epsilon}} \\
&\lesssim 2^{-k\epsilon} \lesssim |x-y|^\epsilon.
\end{aligned}$$

The lower bound follows also quite directly. If $x, y \in X$, we have $d(x,y) \in [2^{-k-1}, 2^{-k}]$ for some $k \in \mathbb{N}$ and

$$|F(x) - F(y)| \geq 2^{k(1-\epsilon)} |f_k(x) - f_k(y)| \geq \delta 2^{-k\epsilon} \sim \delta d(x,y)^\epsilon.$$

$\square$

The mapping $F : X \to \ell_2$ is a biLipschitz embedding of the snowflaked metric space (X, d^ϵ). In the original metric, it is a biHölder embedding. We say that F is ϵ-Hölder, if there is a constant $C > 0$ for which

$$|F(x) - F(y)| \leq Cd(x, y)^\epsilon.$$

We say We say that F is ϵ-biHölder, if there exist constants, $0 < b < C$ for which

$$bd(x, y)^\epsilon \leq |F(x) - F(y)| \leq Cd(x, y)^\epsilon.$$

Remark 4.12. We constructed an embedding with infinitely many coordinates. It is only a bit more technical to construct an embedding to a finite dimensional space $\mathbb{R}^N$. This can be obtained as follows. Choose $k_0 \in \mathbb{N}$ and define

$$\begin{aligned} G(x) &= (f_0 + 2^{(1-\epsilon)k_0} f_{k_0} \cdots, f_1 + 2^{(1-\epsilon)(k_0+1)} f_{k_0+1}, \cdots, f_{k_0-1} + 2^{(1-\epsilon)(2k_0-1)} f_{k_0+k_0-1}) \\ &= (\sum_{i=0}^{\infty} 2^{(1-\epsilon)(0+k_0 i)} f_{0+ik_0}, \sum_{i=0}^{\infty} 2^{(1-\epsilon)(1+k_0 i)} f_{1+ik_0}, \\ &\quad \sum_{i=0}^{\infty} 2^{(1-\epsilon)(2+k_0 i)} f_{2+ik_0}, \cdots, \sum_{i=0}^{\infty} 2^{(1-\epsilon)(k_0-1+k_0 i)} f_{k_0-1+ik_0}). \end{aligned}$$

Here, the idea is to “recycle” coordinates. That is, instead of bringing a new coordinate for every f_k, we cycle through coordinates with a period of k_0. This gives a mapping to $\mathbb{R}^{Nk_0}$. The upper Lipschitz bound is proven in the same way as before. The lower Lipschitz bound is where things are a bit more technical. Indeed, for every $x, y \in X$, you choose k so that $d(x, y) \in [2^{-k-1}, 2^k]$, and then let $m = 0, \ldots, k_0 = -1, n \in \mathbb{N}$ be such that $k = m + nk_0$. We then get

$$
\begin{aligned}
|G(x)-G(y)| &\geq \left|\sum_{i=0}^{\infty} 2^{(1-\epsilon)(m+k_0 i)} f_{m+ik_0}(x) - f_{m+ik_0}(y)\right| \\
&\geq \Bigg|2^{(1-\epsilon)k}(f_k(x)-f_k(y)) + \sum_{i=0}^{n-1} 2^{(1-\epsilon)(m+k_0 i)} f_{m+ik_0}(x) - f_{m+ik_0}(y) \\
&\quad + \sum_{i=n+1}^{\infty} 2^{(1-\epsilon)(m+k_0 i)} f_{m+ik_0}(x) - f_{m+ik_0}(y)\Bigg| \\
&\geq |2^{(1-\epsilon)k}(f_k(x)-f_k(y))| - \sum_{i=0}^{n-1} 2^{(1-\epsilon)(m+k_0 i)} |f_{m+ik_0}(x) - f_{m+ik_0}(y)| \\
&\quad - \sum_{i=n+1}^{\infty} 2^{(1-\epsilon)(m+k_0 i)} |f_{m+ik_0}(x) - f_{m+ik_0}(y)| \\
&\geq \frac{\delta}{2} d(x,y)^{\epsilon} - C\sum_{i=0}^{n-1} 2^{(1-\epsilon)(m+k_0 i)} 2^{-k} - C\sum_{i=n+1}^{\infty} 2^{(1-\epsilon)(m+k_0 i)} 2^{-(m+k_0 i)} \\
&\geq \frac{\delta}{2} d(x,y)^{\epsilon} - \frac{C}{2^{1-\epsilon}-1} 2^{(1-\epsilon)(k-k_0+1)} 2^{-k} - \frac{C}{1-2^{-\epsilon}} 2^{-\epsilon(k+k_0)} \\
&\geq \frac{\delta}{2} d(x,y)^{\epsilon} - \left(\frac{C2^{-k_0(1-\epsilon)-2+\epsilon}}{2^{1-\epsilon}-1} + \frac{C2^{-\epsilon k_0 - 1}}{1-2^{-\epsilon}}\right) d(x,y)^{\epsilon}.
\end{aligned}
$$

By choosing k_0 so that the factor of the second term is less than $\delta/4$, we get the desired lower bound.

We construct the mappings F_k in the next subsection.

Problem 4.13. The proof we gave, if done carefully, yields that for all $\epsilon \in (1/2, 1)$ we have that

$$c_2(X, d^{\epsilon}) = O(\epsilon^{-\frac{1}{2}}).$$

This embedding may need infinitely many coordinates.

There is an open problem to determine if there is a constant M, which is independent of ϵ, and only depends on the doubling constant, so that X can be embedded already in $\mathbb{R}^M$ with distortion comparable to $O(\epsilon^{-\frac{1}{2}})$.

It is known, that one can do $O(\epsilon^{-(1+\delta)})$ for any $\delta > 0$. Also, recently Terence Tao showed that for one of the most difficult examples, the Heisenberg group, the conjecture is correct. See the following for some interesting work.

1. Tao's paper on "Embedding the Heisenberg group into a bounded dimensional Euclidean space with optimal distortion": `https://arxiv.org/abs/1811.09223`

2. Seung-Yeon Roo's follow up "Embedding snowflakes of Carnot groups into bounded dimensional Euclidean spaces with optimal distortion": `https://arxiv.org/abs/2004.07441`

3. Naor and Neiman "Assouad's embedding theorem with dimension independent of the snowflaking": `https://web.math.princeton.edu/~naor/homepage%20files/assouad-N(K).pdf`

4. Deterministic version of the previous theorem by Guy David and Marie Snipes: A constructive proof of the Assouad embedding theorem with bounds on the dimension `https://hal.science/hal-00751548/document`

It had been considered by Naor and Neiman that this problem could be used to disprove the Lang-Plaut problem. However, presently this seems (to the lecturer) a bit unlikely to work. Indeed, Seung-Yeon Roo poses the previous problem in her paper as a conjecture.

4.4 Hierarchical decomposition and mappings F_k

We want to describe a space by behaviors at different scales. This is achieved by using the notion of an ϵ-net.

Definition 4.14. *A set N is ϵ separated if for all $x, y \in N$ we have $d(x,y) \leq \epsilon$.*
An ϵ-net N_ϵ is a maximal ϵ separated set.

Problem 4.15. If X is N-metric doubling, and if S is a $2^{-k}r$ separated set in $B(x,r)$, then $|S| \leq N^{k+1}$. Hint: cover $B(x,r)$ by balls of radius $2^{-k-1}r$. Each such ball can contain at most one point of S – Why?

A maximal set is one that can not be made any bigger.

Lemma 4.16. *If N_ϵ is an ϵ-net, then*

$$X \subset \bigcup_{n \in N_\epsilon} B(n, \epsilon).$$

Proof. If $X \setminus \bigcup_{n \in N_\epsilon} B(n,\epsilon) \neq \emptyset$, then choose $x \in X \setminus \bigcup_{n \in N_\epsilon} B(n,\epsilon)$.. We have $d(x,n) \geq \epsilon$ for all $n \in N_\epsilon$, and thus $N \cup \{x\}$ is a larger ϵ separated set. This contradicts the maximality of N_ϵ. □

An ϵ-net can be found by appealing to the so called Zorn's lemma. However, there is also an algorithmic way. If X is compact, we can find N_ϵ by adding points repeatedly to an ϵ separated set, until the condition in the previous Lemma is satisfied.

Now, if $\operatorname{diam}(X) \leq 1$, then we can construct, recursively, a sequence of ϵ-nets for $\epsilon = 2^{-k}$. First, we construct N_0 a 1-net for X, and then $N_1 \supset N_0$, which is a 2^{-1} net, and recursively, N_k which is a 2^{-k} net. By doing the procedure recursively, we get the nested relationship:

$$N_0 \subset N_1 \subset N_2 \cdots .$$

Further, we have a sequence of covers of X by balls

$$X \subset \bigcup_{n \in N_k} B(n, 2^{-k}).$$

We think of these coverings as a hierarchical decomposition of X at the scales 2^{-k}. This is slightly inaccurate, since the pieces of our decomposition, the balls $B(n, 2^{-k})$ overlap each other.

Our idea next is to define f_k as a distance embedding, but this requires first doing some grouping. We will divide N_k into finitely many sets that are well separated.

Lemma 4.17. *There is a constant M so that we can divide*

$$N_k = N_k^1 \cup N_k^2 \cdots \cup N_k^M$$

where each N_k^i is 2^{5-k} separated.

Proof. Let G be a graph which is constructed as follows. Its vertices are N_k and it has an edge $\{n, m\} \in E$ if and only if $d(n, m) < 162^{-k}$.

An M-coloring of G is a mapping $\chi : V \to \{1, \ldots, M\}$, so that no edge $\{n, m\} \in E$ is monochromatic, i.e. $\chi(n) \neq \chi(m)$. If we have such a coloring, we set

$$N_k^i = \{n \in N_k : \chi(n) = i\}.$$

In general, it is hard to decide if an M-coloring exists for a graph G. (For example, it was a long-standing open problem to decide if all planar graphs were 4-colorable. This was only shown by Appel and Hanken with a computer assisted proof.) However, a simple bound exists, which uses the degree of the graph.

The degree of a graph G is the largest number of edges that meet at a vertex, that is

$$\max_{n \in V} |\{m \in V : \{n, m\} \in E\}|.$$

The graph G that we constructed has degree at most N^6, where N is the doubling constant from Definition 4.5. Indeed, by Problem 4.15 we have that any 2^{-k} -separated set within $B(n, 2^{5-k})$ can have at most N^6 many elements. Thus

$$|\{m \in V : \{n, m\} \in E\}| = |\{m \in N_k : m \in B(n, 2^{5-k}\}| \leq N^6.$$

Now, choose $M = N^5 + 1$. Thus, G has degree at most M. We claim that G is $M-$colorable. This can be achieved with the following iterative algorithm, where we simply assign an available color to each added vertex.

1. List elements of V in any order.

2. While there are uncolored vertices, take the next vertex V in our list.

3. Choose $\chi(v)$ to be a color distinct from all of its neighbors whose color has already been assigned.

4. Repeat until all vertices colored.

The choice of $\chi(v)$ is possible since every v has at most $M-1$ neighbors. Thus, these have at most $M-1$ distinct colors. There is thus at least one color always left for the added vertex. $\square$

Problem 4.18. Show that the algorithm in the previous proof is not optimal. That is, find some graph G with say degree M, but which can be colored by 2 colors. (Or, just less than $M+1$ colors.)

Using these, we can construct the local embeddings looking for.

Lemma 4.19. *If X is doubling and $\operatorname{diam}(X) \leq 1$, then for every $k \in \mathbb{N}$ there exist mappings $f_k : X \to \mathbb{R}^n$ which satisfy the following, for some constants δ, C.*

1. *If $d(x,y) \in [2^{-k-1}, 2^k]$, then $|f_k(x) - f_k(y)| \geq \delta d(x,y)$.*
2. *$\sup_{x\in X} |f_k(x)| \leq C2^{-k}$.*
3. *The mapping f_k is C-Lipschitz.*

Proof. Let N_{k+3} be the 2^{-k-3} net, and let

$$N_{k+1} = \bigcup_{i=1}^{M} N_{k+3}^i$$

be the decomposition from Lemma 4.17. Let

$$f_k(x) = (\min(d(x, N_{k+3}^1), 2^{-k}), \min(d(x, N_{k+3}^2), 2^{-k}), \cdots, \min(d(x, N_{k+3}^M), 2^{-k}).$$

Then, f_k is $\sqrt{M}$-Lipschitz, and $\sup_{x\in X} |f_k(x)| \leq \sqrt{M}2^{-k}$. We are left to show the first property of the Lemma.

Let $x, y \in X$ with $d(x,y) \in [2^{-k-1}, 2^k]$. Choose $n_x \in N_{k+3}$ with $d(n_k, x) < 2^{-k-3}$. We have $n_x \in N_{k+3}^i$ for some i. We also get

$$d(y, n_x) \geq d(y,x) - d(x, n_x) \geq 2^{-k-1} - 2^{-k-3} \geq 2^{-k-2},$$

$$d(y, n_x) \leq d(y,x) + d(x, n_x) \leq 2^{-k} + 2^{-k-3} \leq 2^{1-k},$$

We now show that n_x is the closest point in N_{k+3}^i to both x and y. For every $m \in N_{k+2}^i$, we have $d(n_x, m) \geq 2^5 2^{-k-3} = 2^{2-k}$. We get the following estimates from the triangle inequality:

$$d(m,x) \geq d(m, n_x) - d(n_x, x) \geq 2^{2-k} - 2^{-k-3} \geq d(n_x, x),$$

and

$$d(m,y) \geq d(m, n_x) - d(n_x, y) \geq 2^{2-k} - 2^{1-k} \geq 2^{1-k} \geq d(y, n_x).$$

These imply that $d(x, N_{k+2}^i) = d(x, n_x)$ and $d(y, N_{k+2}^i) = d(y, n_x)$. Thus,

$$\begin{aligned} |f_k(x) - f_k(y)| &\geq |d(x, N_{k+2}^i) - d(y, N_{k+2}^i)| \\ &= |d(y, n_x) - d(x, n_x)| \\ &\geq d(y, n_k) - d(x, n_k) \\ &\geq 2^{-k-2} - 2^{-k-3} \geq 2^{-2} d(x,y). \end{aligned}$$

$\square$

4.5 Assouad dimension

Related to the previous discussion there is the notion of Assouad dimension.

Definition 4.20. *A metric space (X, d) is said to have Assouad dimension at most α if there exists a constant C so that for all $R > r > 0$ every ball $B(x, R)$ can be covered by at most $CR^{\alpha}r^{-\alpha}$ balls of radius r. That is, there exist $x_1, \ldots, x_N \in X$, with $N \leq CR^{\alpha}r^{-\alpha}$ for which*

$$B(x, R) \subset \bigcup_{i=1}^{N} B(x_i, r).$$

The infimum of such α is called the Assouad dimension of X, and is denoted $\dim_A(X)$.

Problem 4.21. Show that $\mathbb{R}^d$ has Assouad dimension d. Hint: Use the volume argument from Lemma 4.2. Also, choose S to be a maximal set in $B(x, R)$ where each pair of points in S has distance at least r. Then $B(x, R) \subset \bigcup_{i=1}^{N} B(x_i, r)$, since otherwise one could add another point.

Doubling and Assouad dimension are some of the simplest ways of measuring dimension.

5 Why? - Lecture 3

We have seen now some answers to what metric embeddings are, and seen some examples of how to construct them. What remains is to understand why we would care. Besides a deep connection between the geometry of the metric space and the embeddability of the space, there are many algorithmic questions, which embeddings help to answer. An important role is played by embeddings into ℓ_1.

5.1 Problem of interest

We consider a graph $G = (V, E)$. A set $S \subset V$ is called a *cut*, since it is thought of as cutting the points inside S from the points outside of it. Let $\partial S \subset E$ be the set of edges connecting S to its complement: $\partial S = \{\{x, y\} \in E : x \in S, y \notin S\}$. We consider the sparsest cut problem:

$$\min\{\frac{|\partial S|}{\min\{|S|, |V \setminus S|\}} : S \subset V\}.$$

This problem is NP complete to solve. However, we can see the problem as an optimization problem on semi-metrics. Let X be a set, and let $S \subset X$ be any set. A cut semi-metric is given by

$$d_S(x, y) = |1_S(x) - 1_S(y)|.$$

This is called a semi-metric, since it satisfies all properties of the distance function except that $d_S(x, y) = 0$ whenever x, y lie on the same side of the cut.

When X is finite, the collection of all subsets E of X will be called Cut (X) and is thought of as the space of cuts of X. Now, we can express the sparsest cut problem by observing the following two facts. First

$$|\partial S| = \sum_{\{x,y\} \in E} d_S(x, y)$$

and second, since either $|S| \geq n/2$ or $|V \setminus S| \geq n/2$, we get

$$\frac{|S||V \setminus S|}{|V|} \leq \min(|S|, |V \setminus S|) \leq \frac{2|S||V \setminus S|}{|V|}.$$

We also have

$$|S||V \setminus S| = \sum_{x,y \in V} d_S(x, y).$$

This latter expression is much easier to work with, and gives a constant approximation for the minimum of the sparsest cut problem. We thus focus on it next.

Now, the problem has been reduced to the following problem:

$$\min\{\frac{\sum_{\{x,y\} \in E} d_S(x, y)}{\sum_{x,y \in V} d_S(x, y)} : S \subset V\}.$$

This problem is a difficult *integer linear optimization problem.*

An equivalent way to write this as a linear program is introducing a parameter $\lambda > 0$ and writing this as

$$\inf \sum_{\{x,y\} \in E} \lambda d_S(x, y)$$

subject to

$$\sum_{x,y \in V} \lambda d_S(x, y) \leq 1.$$

A linear relaxation is obtained by replacing $\lambda d_S(x, y)$ by any (semi-)metric $d(x, y)$:

$$\inf \sum_{\{x,y\} \in E} d(x, y)$$

subject to

$$\sum_{x,y \in V} d(x, y) = 1,$$

$$d(x, y) \geq 0, d(x, z) \leq d(x, y) + d(y, z).$$

Let $\mathcal{D}(X)$ be the space of metrics on X. Since any semimetric can be approximated by a metric, we will not belabour this small difference between semi and true metrics. The relaxed problem gives the same optimum when optimized over either spaces.

We have thus relaxed to the linear optimization problem on the convex cone $\mathcal{D}(X)$, which is defined by a polynomial number of linear constraints. Our question is.

Question 5.1. How much do we lose in this approximation.

Further, as we will see, we may consider the problem for any subset of semimetrics $\mathcal{D}'(X) \subset \mathcal{D}(X)$, as long as $\mathcal{D}'(X)$ includes all cut semi-metrics. We shall consider three classes of metrics:

1. ℓ_1-metrics $\mathcal{D}_1(X)$, for which the relaxed problem has the same optimum as the non-relaxed problem. This will be Theorem 5.8.

2. All metrics $\mathcal{D}(X)$. We get a $\log(n)$ approximation via Bourgain's embedding theorem. This will be Theorem 5.10.

3. Briefly we will consider metrics of negative type $\mathcal{D}_{1/2}(X)$. This will yield a $(\log(n))^{1+\delta}$-approximation for any $\delta > 0$. This will be Theorem 5.14.

While the relaxation to $\mathcal{D}_1(X)$ preserves the optimal value, the relaxations into $\mathcal{D}(X)$ and $\mathcal{D}_{1/2}(X)$ lose something. How much do they lose? Since $\mathcal{D}_1(X)$ metrics can be described by embeddings, the loss will be seen to be precisely the distortion of an embedding.

Why would we accept this loss? The optimization over $\mathcal{D}_1(X)$ is infeasible – indeed NP complete. However, the optimization over the cones $\mathcal{D}(X)$ and $\mathcal{D}_{1/2}(X)$ is feasible. In the first case it can be done with linear programming, and in the second case it can be done with so called semi-definite programming. For this reason, these are also called the linear- and semi-definite relaxations of the sparsest cut problems.

5.2 Cut metrics and ℓ_1

The first class $\mathcal{D}_1(X)$ of ℓ_1 metrics, which we shall define next. While a cut metric d_E is not a distance, certain metrics, namely ℓ_1 metrics, can be expressed as linear combinations of such metrics.

Definition 5.2. *A metric d on a set X is called an ℓ_1^m-metric, if there exists an isometry $f : X \to \ell_1^m$ so that $d(x, y) = \|f(x) - f(y)\|$.*

The collection of ℓ_1-metrics on X is denoted $\mathcal{D}_1(X)$.

If $a < b$, then we can write

$$|a - b| = \int_{\mathbb{R}} |1_{A_t}(a) - 1_{A_t}(b)| dt,$$

where $A_t = \{y : y > t\}$, and $1_{A_t}(x)$ is the characteristic function of A_t, that is $1_{A_t}(x) = 1$ if $x \in A_t$ and 0 otherwise. This "trivial" observation underlies the following cut-cone decomposition.

Lemma 5.3. *If d is an ℓ_1^m-metric on a finite set X, then there exists an isometry $g : X \to \ell_1^{2^{|X|}}$, and we can write*

$$d(x, y) = \sum_{E \in \text{ Cut } (X)} c_E d_E(x, y).$$

Proof. Let $f : X \to \ell_1^m$ be the isometry given by the definition. For each (i,t), with $i = 1, \dots, m$ and $t \in \mathbb{R}$, let

$$E_{(i,t)} = \{x \in X : f(x)_i > t\}.$$

These are cuts in X. Now, let

$$c_E = \sum_{i=1}^{m} |\{t \in \mathbb{R} : E_{(i,t)} = E\}|.$$

Let $A_t = \{z \in \mathbb{R} : z > t\}$. By the observation stated before the lemma, we get

$$\begin{aligned} \|f(x) - f(y)\| &= \sum_{i=1}^{m} |f(x)_i - f(y)_i| \\ &= \sum_{i=1}^{m} \int_{\mathbb{R}} |1_{A_t}(f(x)_i) - 1_{A_t}(f(y)_i)| dt. \end{aligned}$$

Now, $1_{A_t}(f(x)_i) = 1_{E_{(i,t)}}(x)$, and thus we get

$$\|f(x) - f(y)\| = \sum_{i=1}^{m} \int_{\mathbb{R}} |1_{E_{(i,t)}}(x) - 1_{E_{(i,t)}}(y)| dt.$$

Now, $E_{(i,t)}$ is always some element in the (finite) collection of cuts $\text{Cut}\ (X)$. The integral in $\mathbb{R}$ can be split into pieces where $E_{(i,t)} = E$ for any $E \in \text{Cut}\ (X)$. The measure of the set of t where the cut E is used is given by $|\{t \in \mathbb{R} : E_{(i,t)} = E\}|$. Using this, we get

$$\begin{aligned} \|f(x) - f(y)\| &= \sum_{i=1}^{m} \int_{\mathbb{R}} |1_{E_{(i,t)}}(x) - 1_{E_{(i,t)}}(y)| dt \\ &= \sum_{i=1}^{m} \sum_{E \in \text{ Cut }(X)} |\{t \in \mathbb{R} : E_{(i,t)} = E\}| |1_E(x) - 1_E(y)| \\ &= \sum_{E \in \text{ Cut }(X)} c_E |1_E(x) - 1_E(y)|. \end{aligned}$$

Next, let $g : X \to \ell_1(2^{|X|})$ be constructed as follows. We index the coordinates of $\mathbb{R}^{2^{|X|}}$ by the cuts $E \in \text{Cut}\ (X)$, and we set

$$g(x)_E = c_E 1_E(x).$$

Then, for all $x, y \in X$,

$$\|g(x) - g(y)\| = \sum_{E \in \text{ Cut }(X)} c_E |1_E(x) - 1_E(y)| = d(x,y).$$

$\square$

The same proof works if X embeds into $L^1(\Omega)$ for some probability space $(\Omega, \Sigma, \P)$.

Lemma 5.4. *If (X, d) be a finite set which embeds isometrically to $L^1(\Omega)$, then there exists an isometry $g : X \to \ell_1^{2^{|X|}}$, and we can write*

$$d(x, y) = \sum_{E \in \text{ Cut } (X)} c_E d_E(x, y).$$

Proof. Let $f : X \to L^1(\Omega)$ be the isometry given by the definition. For each (ω, t), $\omega \in \Omega$, with $\omega \in S^{n-1}$ and $t \in \mathbb{R}$, let

$$E_{(\omega,t)} = \{x \in X : f[x](\omega) > t\}.$$

These are cuts in X. Now, let

$$c_E = \int |\{t \in \mathbb{R} : E_{(\omega,t)} = E\}| d\mathbb{P}_\omega$$

Let $A_t = \{z \in \mathbb{R} : z > t\}$. By the observation stated before the lemma, we get

$$\begin{aligned} \|f(x) - f(y)\| &= \int |f[x](\omega) - f[y](\omega)| d\mathbb{P}_\omega \\ &= \int \int_{\mathbb{R}} |1_{A_t}(f[x](\omega)) - 1_{A_t}(f[y](\omega)| dt d\mathbb{P}_\omega. \end{aligned}$$

Now, $1_{A_t}(f[x](\omega)) = 1_{E_{(\omega,t)}}(x)$, and thus we get

$$\|f(x) - f(y)\| = \int \int_{\mathbb{R}} |1_{E_{(\omega,t)}}(x) - 1_{E_{(\omega,t)}}(y)| dt d\mathbb{P}_\omega$$

Now, $E_{(\omega,t)}$ is always some element in the (finite) collection of cuts $\text{ Cut } (X)$. The integral in $\mathbb{R}$ can be split into pieces where $E_{(\omega,t)} = E$ for any $E \in \text{ Cut } (X)$. The measure of the set of t where the cut E is used is given by $|\{t \in \mathbb{R} : E_{(\omega,t)} = E\}|$. Using this, we get

$$\begin{aligned} \|f(x) - f(y)\| &= \int \int_{\mathbb{R}} |1_{E_{(\omega,t)}}(x) - 1_{E_{(\omega,t)}}(y)| dt d\mathbb{P}_\omega \\ & \int \sum_{E \in \text{ Cut } (X)} |\{t \in \mathbb{R} : E_{(\omega,t)} = E\}| |1_E(x) - 1_E(y)| d\P_\omega \\ &= \sum_{E \in \text{ Cut } (X)} c_E |1_E(x) - 1_E(y)|. \end{aligned}$$

The rest of the proof is identical. ∎

Corollary 5.5. *If (X, d) is an isometric subset of ℓ_2^n, then d is an ℓ_1 metric.*

Proof. Consider m i.i.d. normal random variables X_i, $i = 1, \ldots, m$ with distribution $N(0,1)$. Let Ω be the probability space of X_i (which can be taken as $\mathbb{R}^m$), and let $\mathbb{P}$ be the corresponding probability measure. Now, we map

$$F : \ell_2^n \to L^1(\Omega)$$

with

$$F(v_1, \ldots, v_n) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n} v_i X_i.$$

We have

$$\|F(v_1, \ldots, v_n)\|_{L^1} = \frac{1}{\sqrt{2\pi}} \int \| \sum_{i=1}^{n} X_i \| d\mathbb{P} = \frac{1}{\sqrt{2\pi}} \mathbb{E}(| \sum_{i=1}^{n} X_i |).$$

We have that $\sum_{i=1}^n X_i$ is a normal random variable with mean 0 and variance $\sum_{i=1}^n v_i^2$. Thus, the standard deviation of it is $\sqrt{\sum_{i=1}^n v_i^2}$. The expectation of the absolute value of such a random variable is $\sqrt{2\pi}$ times the standard deviation. Thus

$$\|F(v_1, \ldots, v_n)\|_{L^1} = \frac{1}{\sqrt{2\pi}} \mathbb{E}(| \sum_{i=1}^{n} X_i |) = \sqrt{\sum_{i=1}^{n} v_i^2}.$$

Thus, F is an isometry of ℓ_2^m to $L^1(\Omega)$. Consequently, X is also isometric to a subset of $L^1(\Omega)$. □

Remark 5.6. In the previous lemma, we can also take $n \to \infty$ and embed ℓ_2 to $L^1(\Omega)$. This involves the Kolmogorov three series theorem. Indeed, if $v = (v_1, v_2, \ldots,) \in \ell_2$, then

$$\lim_{n \to \infty} \sum_{i=1}^{n} v_i X_i = \sum_{i=1}^{\infty} v_i X_i$$

converges almost surely to a Gaussian normal variable with mean 0 and variance $\sum_{i=1}^\infty v_i^2$. The rest of the proof is identical.

This implies that

$$c_1(X) \leq c_2(X),$$

since whenever X embeds into ℓ_2 with distortion D, then we can compose this with the previous isometry to obtain an embedding in ℓ_1 with distortion D.

Next, we need a simple inequality for real numbers.

Lemma 5.7. *Let $a_1, \ldots, a_n > 0$ and $b_1, \ldots, b_n > 0$. Then*

$$\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \geq \min_{i=1} \frac{a_i}{b_i}.$$

Proof. Let i_0 be such that
$$\frac{a_{i_0}}{b_{i_0}} = \min_{i=1} \frac{a_i}{b_i}.$$
Then for all $i = 1, \ldots, n$ we have
$$\frac{a_{i_0}}{b_{i_0}} b_i \leq \frac{a_i}{b_i} b_i \leq a_i.$$
Summing these estimates over i, we get
$$(\sum_{i=1}^{n} b_i) \min_{i=1} \frac{a_i}{b_i} \leq \sum_{i=1}^{n} a_i.$$
From these the claim follows. □

Recall that $\mathcal{D}_1(X)$ is the space of ℓ_1-metrics on X.

Theorem 5.8. *The following two optimization problems have the same minimum:*
$$\inf \sum_{\{x,y\} \in E} \lambda d_S(x, y)$$
subject to
$$\sum_{x,y \in V} \lambda d_S(x, y) = 1,$$
and
$$\inf \sum_{\{x,y\} \in E} d(x, y)$$
subject to $d \in \mathcal{D}_1(X)$
$$\sum_{x,y \in V} d(x, y) = 1,$$

Proof. Let $M_{\text{ Cut } (X)}, M_{\ell_1}$ be the minima of the first and second problems.

For every $\lambda > 0$ and any $S \in$ Cut (X), we have $\lambda d_S \in \mathcal{D}_1(X)$. Thus
$$M_{\ell_1} \leq M_{\text{ Cut } (X)}.$$

Now, let $d \in \mathcal{D}_1(X)$ be such that $\sum_{x,y \in V} d(x, y) = 1$. Choose the metric d so that
$$\sum_{\{x,y\} \in E} d_S(x, y) = M_{\ell_1}.$$
(This minimization problem has a true minimum by an argument using compactness. However, if you don't like that, you can always add an $\epsilon > 0$ to the right hand side.)

Then, by Lemma 5.3, there are numbers c_S for $S \in \text{ Cut } (X)$, for which

$$d(x,y) = \sum_{S \in \text{ Cut } (X)} c_S d_S(x,y).$$

Let $\lambda_S = \sum_{x,y \in V} c_S d_S(x,y)$. We have

$$\sum_{S \in \text{ Cut } (S)} \lambda S = 1.$$

Thus,

$$M_{\ell_1} = \sum_{\{x,y\} \in E} d(x,y) = \frac{\sum_{S \in \text{ Cut } (S)} \sum_{\{x,y\} \in E} c_S d_S(x,y)}{\sum_{S \in \text{ Cut } (S)} \lambda_S}.$$

Let now $a_S = \sum_{\{x,y\} \in E} c_S d_S(x,y)$ and $b_S = \lambda_S$. By Lemma 5.7, we have some $S \in \text{ Cut } (S)$ for which

$$\frac{\sum_{\{x,y\} \in E} c_S d_S(x,y)}{\lambda_S} \leq M_{\ell_1}.$$

Now, let $\lambda = c_S \lambda_S^{-1}$. Then, we get

$$\sum_{x,y \in V} \lambda d_S(x,y) = 1,$$

and

$$\sum_{\{x,y\} \in E} \lambda d_S(x,y) = \frac{\sum_{\{x,y\} \in E} c_S d_S(x,y)}{\sum_{S \in \text{ Cut } (S)} \lambda_S} \leq M_{\ell_1}.$$

Thus, we found a cut which we can plug into the first optimization problem. Thus, we get

$$M_{\text{ Cut } (X)} \leq M_{\ell_1}.$$

□

5.3 General metrics

Theorem 5.9. *Let $\mathcal{D}_s$ be some cone of metrics on X, so that*

$$\mathcal{D}_1(X) \subset \mathcal{D}_s(X).$$

Then, let $M_{\text{ Cut } (X)}$ be the optimum of the sparsest cut problem in Theorem 5.8, and let $M_{\mathcal{D}_s}$ be the optimum of the following problem

$$\inf \sum_{\{x,y\} \in E} d(x,y)$$

subject to $d \in \mathcal{D}_s$
$$\sum_{x,y\in V} d(x,y) = 1,$$
Then, we have for
$$D_s = \sup_{d\in\mathcal{D}_s} c_1(X,d)$$
that
$$M_{\mathcal{D}_s} \le M_{\text{ Cut } (X)} \le D_s M_{\text{ Cut } (X)}.$$

Proof. First, since $\mathcal{D}_1(X) \subset \mathcal{D}_s$, we get from Theorem 5.8 that $M_{\text{ Cut } (X)} = M_{\ell_1} \ge M_{\mathcal{D}_s}$.

Next, suppose that $d \in \mathcal{D}_s$ is the metric for which
$$M_{\mathcal{D}_s} = \inf \sum_{\{x,y\}\in E} d(x,y)$$
and
$$\sum_{x,y\in V} d(x,y) = 1.$$

Now, by the definition of the distortion, each metric $d \in \mathcal{D}_s$ admits an $f : X \to \ell_1$ for which
$$d(x,y) \le \|f(x) - f(y)\| \le D_s d(x,y).$$
Let $d_f(x,y) = \|f(x) - f(y)\|$. This is an ℓ_1-metric. And
$$\sum_{\{x,y\}\in E} d_f(x,y) \le \sum_{\{x,y\}\in E} D_s d(x,y) = M_{\mathcal{D}_f}.$$
with
$$\sum_{x,y\in V} d_f(x,y) \ge \sum_{x,y\in V} d(x,y) = 1,$$
Let $\lambda = \sum_{x,y\in V} d_f(x,y) \ge 1$. Then $d_{new} = \lambda^{-1} d_f$ is still a ℓ_1-metric, for which
$$\sum_{\{x,y\}\in E} d_{new}(x,y) \le D_s M_{\mathcal{D}_f}.$$
with
$$\sum_{x,y\in V} d_{new}(x,y) = 1.$$
Thus,
$$M_{\ell_1} \le \sum_{\{x,y\}\in E} d_{new}(x,y) \le M_{\mathcal{D}_f},$$
as desired, since M_{ℓ_1} is the optimum for the ℓ_1-metric problem. $\square$

Now, using this we get the following.

Theorem 5.10. *Let $\mathcal{D}(X)$ be the class of all metrics and let $M_{\mathcal{D}(X)}$ be the solution of the relaxed problem*

$$\inf \sum_{\{x,y\}\in E} d(x,y)$$

subject to $d \in \mathcal{D}(X)$

$$\sum_{x,y\in V} d(x,y) = 1.$$

Then

$$M_{\mathcal{D}(X)} \leq M_{\text{ Cut }(X)} \leq O(\log(|X|)M_{\mathcal{D}(X)}.$$

Proof. This follows from Theorem 5.9 together with Bourgain, which states that if $d \in \mathcal{D}(X)$, then (X,d) embeds with distortion $O(\log(|X|))$ into ℓ_1^m for some m. This means, $\sup_{d\in\mathcal{D}(X)} c_1(X,d) \leq O(\log(|X|))$, and the claim follows. □

The main simplicity here is that $\mathcal{D}(X)$ is actually a convex cone defined by the following linear constraints:

$$d(x,y) \geq 0, d(x,z) \leq d(x,y) + d(y,z).$$

(This defines the cone of semi-metrics, but as stated earlier, this does not alter the minima.)

5.4 Goemans-Linial relaxation

While $\mathcal{D}(X)$ is convenient, a slightly more complicated class of metrics can be considered.

Definition 5.11. *A metric (X,d) is said to be of negative type, if $(X,\sqrt{d})$ embeds isometrically to ℓ_2. The collection of metrics of negative type are denoted $\mathcal{D}_{1/2}(X)$.*

In other words, d is a metric of negative type, if

$$d(x,y) = \|f(x) - f(y)\|^2$$

for some map $f : X \to \ell_2^n$. Given such a mapping f, we can write our previous optimization problem as

$$\inf \sum_{\{x,y\}\in E} \|f(x) - f(y)\|^2$$

subject to $d \in \mathcal{D}_s$

$$\sum_{x,y\in V} \|f(x) - f(y)\|^2 = 1.$$

This problem is *tractable* since it is expressed as a so called *semidefinite programming* problem. Explaining this in detail would take us too far on a side track. However, what suffices for our present purposes is to note that semidefinite programming problems are a subset of convex optimization problems which contains all linear problems. Further, they can be solved quite efficiently computationally. In fact, theoretically, they can be solved in polynomial time.

To apply Theorem 5.9 to these metrics, we need two facts.

Lemma 5.12. $\mathcal{D}_1(X) \subset \mathcal{D}_{1/2}(X)$.

Proof. By the cut-cone decomposition, Lemma 5.3, we have that all metrics $d \in \mathcal{D}_1(X)$ can be expressed as

$$d(x,y) = \sum_{S \in \text{ Cut } (X)} c_S d_S(x,y).$$

Now, if $S \in \text{ Cut } (X)$, then we set $f_S(x) = \sqrt{c_S}$ if $x \in S$ and $f_S(x) = 0$ if $x \notin X$. Then,

$$c_S d_S(x,y) = |f_S(x) - f_S(y)|^2.$$

Next, let $F(x)$ be a mapping to $\mathbb{R}^{\text{ Cut } (X)}$, where $F(x)_S = f_S(x)$. [1] Then

$$\begin{aligned} d(x,y) &= \sum_{S \in \text{ Cut } (X)} c_S d_S(x,y) \\ &= \sum_{S \in \text{ Cut } (X)} |f_S(x) - f_S(y)|^2 \\ &= \sqrt{\sum_{S \in \text{ Cut } (X)} |f_S(x) - f_S(y)|^2}^2 \\ &= \|F(x) - F(y)\|^2. \end{aligned}$$

Thus, $d \in \mathcal{D}(X)$. $\square$

We also need an embedding result for negative type metrics. This is much more recent than Bourgain, and quite involved. We do not have the chance to cover it here, but see Arora, Lee and Naor's paper, `https://www.ams.org/journals/jams/2008-21-01/S0894-0347-07-00573-5/S0894-0347-07-00573-5.pdf`, for a proof. The proof uses a distance embedding, but the proof, analysis and argument are more delicate.

Theorem 5.13. *If (X,d) is a metric space with d a metric of negative type, then*

$$c_1(X,d) \leq c_2(X,d) \leq O(\sqrt{\log(|X|)} \log(\log(|X|))).$$

In fact, by a seminal result of Cheeger, Kleiner and Naor[2], which was later sharpened by Naor and Young[3], that $c_1(X,d) \geq O(\sqrt{\log(|X|)}$ for some metric spaces X. This shows that the bound given above is essentially optimal. This involved a lot of heavy geometric measure theory, and the study of Carnot groups. It was one of the big reasons that Analysis on Metric spaces, Subriemannian geometry and embedding problems got increasing attention in the past decade. It was also a beautiful example of how pure math and geometry met a problem in thoeretical computer science.

---

[1] Here $\mathbb{R}^{\text{ Cut } (X)}$ denotes the Euclidean space of dimension $|\text{ Cut } (X)|$, where the components are identified with $S \in \text{ Cut } (X)$. The components of $x \in \mathbb{R}^{\text{ Cut } (X)}$ is denoted x_S, for $S \in \text{ Cut } (X)$. Note that in general, A^B is the collection of all functions $f : B \to A$, and this convention here corresponds to this usage.

[2] See `https://arxiv.org/abs/0910.2026`.

[3] See `https://arxiv.org/abs/1701.00620`.

Theorem 5.14. *Let $\mathcal{D}_{1/2}(X)$ be the class of all metrics of negative type and let $M_{\mathcal{D}_{1/2}(X)}$ be the solution of the relaxed problem*

$$\inf \sum_{\{x,y\}\in E} d(x,y)$$

subject to $d \in \mathcal{D}_{1/2}(X)$

$$\sum_{x,y\in V} d(x,y) = 1.$$

Then

$$M_{\mathcal{D}_{1/2}(X)} \leq M_{\text{ Cut } (X)} \leq O(\sqrt{\log(|X|)}\log\log(|X|)M_{\mathcal{D}_{1/2}(X)}.$$

Proof. This follows from Theorem 5.9 together with Theorem 5.13, which states that if $d \in \mathcal{D}_{1/2}(X)$, then (X,d) embeds with distortion $O(\sqrt{\log(|X|)}\log\log(|X|))$ into ℓ_1^m for some m. This means, $\sup_{d\in\mathcal{D}_{1/2}(X)} c_1(X,d) \leq O(\sqrt{\log(|X|)}\log\log(|X|))$, and the claim follows. □

These results answered negatively and fully (or, at least up to a $\log\log$) factor the Goemans-Linial conjecture from theoretical computer science. This conjectured that the ratio of $M_{\mathcal{D}_{1/2}(X)}$ and $M_{\text{ Cut } (X)}$ could be taken to be $O(1)$.

6 Important examples

We will end, with as much time as is permitted, with analysing a few specific graphs of particular interest. These often serve as examples or counter examples for embedding problems.

6.1 Laakso diamond graph

A great reference for the topics of this subsection is the paper by Lang and Plaut, "Bilipschitz Embeddings of Metric Spaces into Space Forms"[4].

Let $G_0 = (V_0, E_0)$ be the graph with two vertices $V_0 = \{0,1\}$ and one edge E_0. We construct a new graph $G_1 = (V_1, E_0)$ from this, with vertices by adding four vertices for each edge in G_0:

$$V_1 = V_0 \cup \{e_1, e_2, e_3, e_4 : e \in E_0\}$$

Then, we set,

$$E_1 = \{\{x, e_1\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_3, e_4\}, \{e_3, e_4\}, \{e_4, y\} : e = \{x,y\} \in E_0\}.$$

We can repeat this process:

$$V_n = V_{n-1} \cup \{e_1, e_2, e_3, e_4 : e \in E_{n-1}\}$$

[4]See `https://link.springer.com/article/10.1023/A:1012093209450`.

Then, we set, $I_e = \{x, e_1\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_3, e_4\}, \{e_3, e_4\}, \{e_4, y\}$ and

$$E_n = \bigcup_{e \in E_{n-1}} I_e = \{\{x, e_1\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_3, e_4\}, \{e_3, e_4\}, \{e_4, y\} : e = \{x, y\} \in E_{n-1}\}.$$

We also define a mappings: $F : V_n \cup E_n \to V_{n-1} \cup E_{n-1}$ by sentting

$$F(v) = v, \text{ for } v \in V_{n-1}$$

$$F(v) = e, \text{ for } v = e_i \text{ for some } e \in E_{n-1}$$

and

$$F(f) = e, \text{ for } f \in I_e \text{ for some } e \in E_{n-1}.$$

Let also, $F \circ \cdots F = F^m$, where we have composed F m times.

This constructs a sequence of graphs G_n. Equip G_n with the path metric d.

Theorem 6.1. *There are constants δ, C, D so that for all $n \in \mathbb{N}$ the graphs G_n are all D-doubling, and*

$$\delta\sqrt{n} \leq c_2(G_n, d) \leq C\sqrt{n}.$$

Proof. First, we take care of some **preliminaries.** It will be useful to rescale the metrics. Let $d_n = 4^{-n}d$. By induction, one shows that if $v, w \in V_m$, then for all $n > m$ we have $d_n(v, w) = d_m(v, w)$. Further, it is clear that $d_m(v, w) \geq 4^{-m}$, with equality only if v, w are connected by a single edge in G_m. We also have that the diameter of G_n in the metric d_n is at most 1 for all $n \in \mathbb{N}$.

Doubling bound: Let $n > m \in \mathbb{N}$. For each $e \in G_m$, let $I_{n,m}(e) \subset V_n$ be the set of $v \in V_n$ for which $F^{n-m}(v) = e$ or $F^{n-m}(v) \in e$. By induction, each set $I_{n,m}$ has diameter equal to 4^{-m}. Now, let $v \in V_n$ and $r > 0$. If $r \geq 2$, then $B(v, r) = B(v, r/2)$ and we need a single ball of half the radius to cover the ball.

Fix $m \in \mathbb{N}$ so that $24^{-m-1} < r \leq 24^{-m}$. Let

$$I = \{e \in E_{m+1} : I_{n,m}(e) \cap B(v, r) \neq \emptyset\}.$$

First, we establish a size bound for I. For each $e \in I$ choose an end point of e, say v_e. Also, choose an end point v_0 for $F^{n-m}(v)$ (unless, $F^{n-m}(v)$ is already a vertex, in which case $F^{n-m}(v) = v_0$). Then,

$$d_m(v_0, v_e) = d_n(v_0, v_e) \leq 24^{-m} + r \leq 164^{-m-1}.$$

The graphs G_n, by induction, have degree at most 3, and thus there are at most 2^{16} possible values of v_e. Further, each v_e can come from at most 3 edges. Thus

$$|I| \leq 3 \cdot 2^{16}.$$

Now,

$$B(x, r) \subset \bigcup_{e \in I} I_{n,m}(e) \subset \bigcup_{v_e, e \in I} B(v_e, 4^{-m-1}) \subset \bigcup_{v_e, e \in I} B(v_e, r/2).$$

Embedding upper bound: Since G_n is doubling, we can use Lemma 4.19 to find mappings $f_0, \dots, f_{2k}$ as in that lemma. We claim that

$$F(x) = (f_0(x), f_1(x), f_2(x), \cdots, f_{2n}(x))$$

is our desired embedding. For every $x, y \in V_n$ we have $d(x,y) > 2^{-2n-1}$ or that $d(x,y) = 0$ and $x = y$. Thus, if we choose $k \in [1, 2n]$ so that

$$2^{-2n-1} < d(x,y) \leq 2^{-2n},$$

then we have

$$|F(x) - F(y)| \geq \delta |f_k(x) - f_k(x)| \geq \delta d(x,y).$$

We also have the upper bound:

$$\begin{aligned}
|F(x) - F(y)| &\leq \sqrt{\sum_{k=0}^{2n} |f_k(x) - f_k(y)|^2} \\
&\leq \sqrt{\sum_{k=0}^{2n} C^2 d(x,y)^2} \\
&\leq C\sqrt{2n} d(x,y).
\end{aligned}$$

Lower bound: The lower bound uses the parallelogram inequality. The idea is to exploit the mid-point relationship between e_2, e_3 for each $e \in E$.

Let $f_n : V_n \to \ell_2$ be an expanding $(1, L_n)$ bi-Lipschitz map. Let $f_m = f_n|_{V_m}$. We claim that for some constant $c > 0$ we have

$$D(f_{n+1})^2 \geq D(f_n)^2 + \frac{1}{2}.$$

The metric on G_n is a path metric, so the worst possible stretching always occurs on edges. That is, for all n there always exists a pair of vertices $x_1, x_3 \in V_n$ so that $\{x_0, x_2\} \in E_n$ and for which

$$D(f_n) = \frac{|f_n(x_0) - f_n(x_2)|}{d_n(x_0, x_3)}.$$

Now, let x_1, x_3 be the two midpoints of x_0, x_2 in V_{n+1}. We get from the Parallellogram inequality that

$$\begin{aligned}
\|f_{n+1}(x_1) - f_{n+1}(x_3)\|^2 + \|f_{n+1}(x_0) - f_{n+1}(x_2)\|^2 &\leq \|f_{n+1}(x_0) - f_{n+1}(x_1)\|^2 + \|f_{n+1}(x_1) - f_{n+1}(x_2)\|^2 \\
&\quad + \|f_{n+1}(x_2) - f_{n+1}(x_3)\|^2 + \|f_{n+1}(x_3) - f_0(x_0)\|^2.
\end{aligned}$$

We get

$$\begin{aligned}
4^{-2n}/4 + D_n^2 4^{-2n} &\leq \|f_{n+1}(x_1) - f_{n+1}(x_3)\|^2 + \|f_{n+1}(x_0) - f_{n+1}(x_2)\|^2 \\
&\leq \|f_{n+1}(x_0) - f_{n+1}(x_1)\|^2 + \|f_{n+1}(x_1) - f_{n+1}(x_2)\|^2 \\
&\qquad + \|f_{n+1}(x_2) - f_{n+1}(x_3)\|^2 + \|f_{n+1}(x_3) - f_0(x_0)\|^2 \\
&\leq 4D_{n+1}^2 4^{-2n-1} = D_{n+1}^2 4^{-2n}.
\end{aligned}$$

Dividing both sides by 4^{-2n} we get

$$D(f_{n+1})^2 \geq D(f_n)^2 + \frac{1}{4}.$$

Now, from iteraring this, we get

$$D(f_{n+1})^2 \geq D(f_0)^2 + \frac{n}{4},$$

and thus $D(f_{n+1}) \geq \sqrt{n/4}$. □

6.2 Trees

Some good references for this section are:

1. Bourgain, The metrical interpretation of superreflexivity in banach spaces, `https://link.springer.com/article/10.1007/BF02766125`.
2. Matoušek, Jiří, On embedding trees into uniformly convex Banach spaces, `https://link.springer.com/article/10.1007/BF02785579`.
3. David, Eriksson-Bique and Vellis, Bi-Lipschitz embeddings of quasiconformal trees, `https://arxiv.org/abs/2106.13007`.
4. A. Gupta, R. Krauthgamer and J. R. Lee, Bounded geometries, fractals, and low-distortion embeddings, `https://ieeexplore.ieee.org/document/1238226`.

Another case, where can say more about embeddability is the case of trees.

Definition 6.2. *A Graph G is a tree, if it is connected and it does not contain any cycles. Equivalently, there is a unique non-repeating edge path $e_1, \ldots, e_n$ between x and y.*

A metric space X is a geodesic tree, if for every x, y there exists a unique path $\gamma : [0,1] \to X$ which is parametrized by constant speed, and $\gamma(0) = x, \gamma(1) = y$.

A leaf of a tree G is a vertex v whose removal leaves the tree with one component.

In both cases, there is a *unique* arc, whose image is denoted $[x, y]$. (In the graph case, it is helpful to think of the arc as a collection of edges and vertices.)

The principal reason trees are easier to embed is the following simple fact.

Lemma 6.3. *If $p \in V$ is a vertex of a tree, then removal of p, and all the edges associated to it results in a graph G_p which is a union of finitely many disjoint trees. The number of components is equal to the degree of p.*

Proof. Let q_i be the neighbors of p. Once we remove p, there is no path connecting q_i to each other. Thus, they lie in different components. On the other hand, every point t can be connected to p with a unique path, and this must go through one of the p_i. Thus, the number of components is exactly the degree, and each p_i lies in a corresponding component. The removal of edges and points did not result in the creation of any cycles, and thus the components are still trees (since they are also connected). □

This simple structure of trees means they are very amenable to *divide and conquer* approaches to embeddings. Much more so than general metric spaces, where dividing the space into smaller pieces may require using complex, and usually higher dimensional, separating sets.

A standard lemma is the following.

Lemma 6.4. *If G is a tree with at most n leaves, then G can be covered by at most n^2 simple paths.*

Proof. If we take the union of all arcs between two leaves we obtain all of the graph. □

The significance of this is the following.

Lemma 6.5. *If X is a metric space which is a union of two sets $X = A \cup B$, and if $f : X \to \mathbb{R}^n$ is a Lipschitz mapping for which $f|_A$ is a biLipchitz embedding and $g : X \to \mathbb{R}^m$ is a Lipschitz mapping for which $g|_B$ is a biLipschitz embedding, then $F : (f, g, d(A, \cdot)) : X \to \mathbb{R}^n$ is a biLipschitz embedding of X. Here, we im*

This lemma allows us to effectively embed trees by dividing into paths. I only give a sketch of the following proof.

Theorem 6.6. *If T is a doubling and geodesic tree, then it bi-Lipschitz embeds into ℓ_2.*

Proof. We give a simple sketch of an embedding. Let diam$(G) \leq 1$ for simplicity, and let $N_{\frac{1}{2}}$ be a 2^{-1}-net in V. Let $T_{2^{-1}}$ be a unique smallest subtree which contains $N_{2^{-1}}$. The tree $T_{2^{-1}}$ has at most

First, there is an embedding $\tilde{f}_1 : T_{1/2} \to \mathbb{R}^N$ which is biLipschitz, since by doubling $N_{1/2}$ is finite. Next, consider $G \setminus T_{2^{-1}}$, where we also remove all the edges adjacent to points in $T^{2^{-1}}$. Each component is a doubling tree with fewer vertices. Extend $\tilde{f}_1$ to be constant on all the edges in these components, and get $f_1 : G \to \mathbb{R}^N$. Now, for each components repeat the process, but with 2^{-2} replacing 2^{-1}. This gives f_2 for each subtree. But here is the beauty: Each subtree attaches to $T_{1/2}$ by a single point. By choosing these mappings so that the $f_2(x) = 0$ for all these points, and mapping each component (of which there are finitely many) we can combine all the mappings f_2 into a single mapping $f_2 : X \to \mathbb{R}^{NM}$, where M is the number of components. This process is repeated, and we get mappings f_k. The desired mapping is

$$F(x) = (f_1, f_2, f_3, f_4, f_5, f_6, \cdots).$$

Now,

$$\|F(x) - F(y)\| = \sqrt{\sum_{i=1}^{\infty} (f_i(x) - f_i(y))^2},$$

and we can analyse the differences in the sum as follows. If $[x, y]$ is the unique arc between $[x, y]$, then $f_i(x) - f_i(y) = 0$ unless the arc $[x, y]$ intersects some tree constructed at the i'th stage. For each i, the arc can intersect at most two such trees, and we can see that

$|f_i(x) - f_i(y)|$ is comparable to the length within that tree, and is bounded by $O(2^{-i})$. Summing the differences, we get that $\|F(x) - F(y)\| \leq O(d(x,y))$. The lower bound is obtained by finding near-by net points, and using finitely many components to see that the majority of the arc $[x,y]$ is covered by finitely many of the edges, and all edges are mapped perpendicularly.

$\square$

6.3 Open problem on trees

Definition 6.7. *A metric tree T is one where for each x, y can be connected by a unique compact connected set $[x,y]$, which is homeomoprhic to $[0,1]$. Denote*

A metric tree is said to be bounded turning, if there is a constant $C \geq 1$ for which $\mathrm{diam}([x,y]) \leq Cd(x,y)$ *for all $x, y \in T$.*

In our work with David and Vellis, we showed that a doubling and bounded turning tree T biLipschitz embeds. However, the following version was left open.

Problem 6.8. Assume that T is a bounded turning tree, and $K \subset T$ is a compact subset, where K is doubling. Then does K embed biLipschitz to a Euclidean space?

If T is geodesic, then the problem is true, and it follows from a "filling algorithm", whereby we can find another tree T_d which is doubling, geodesic and to which K embeds biLipschitz. Similarly, one could ask the following.

Problem 6.9. Assume that T is a bounded turning tree, and $K \subset T$ is a compact subset, where K is doubling. Then does K embed biLipschitz to a bounded turning tree T_d, which is doubling.

The algorithm for this filling construction is found in the following paper: Anupam Gupta and Kunal Talwar, Making doubling metrics geodesic, `https://link.springer.com/article/10.1007/s00453-010-9397-x`. Unfortunately, the divide-and-conquer analysis that they perform doesn't seem to directly work for bounded turning trees. We made some attempts at generalizing this algorithm, but were not able to prove that it works. I think this is a nice problem, so feel free to contact me, if you want to work on it – or let me know, if you solve it!