Massachusetts Institute of Technology Lecturer: Michel X. Goemans
**18.409: Topics in TCS: Embeddings of Finite Metric Spaces**
September 11, 2006 Scribe: David Charlton

# Lecture 2

Notation: We use the shorthand $[n] = \{1, \ldots, n\}$, and when $X$ is a set of points $\binom{X}{2}$ denotes all unordered pairs of distinct points.

## 1 $\ell_2$-embeddability

Recall from last time that $\ell_2$ can be isometrically embedded into $\ell_q$ for any $q \geq 1$. (In fact, though we will not prove it here, $\ell_p$ is isometrically embeddable into $\ell_q$ whenever $1 \leq q \leq p \leq 2$.) This lecture will focus on the converse question: under what circumstances can a finite metric space be isometrically embedded into $\ell_2$, and when it cannot, what is the distortion needed?

It is easy to see that not everything can be isometrically embedded into $\ell_2$: Consider the four points of the square, where where we give the edges distance 1, and diagonals distance 2 (this is the "graph metric" for the 4-cycle $C_4$, or the $\ell_1$ metric for the unit square). A little thought should convince you that it cannot be embedded into $\ell_2$ (the Euclidean norm would require each corner to coincide with the center of the opposing diagonal). How can we generalize this, that is, given an arbitrary set of points, how do we tell whether they are isometrically embeddable into $\ell_2$?

We address the question by looking at the properties of metrics that *can* be embedded into $\ell_2$. Specifically, let $X = [n]$ and $\delta : [n]^2 \to \mathbb{R}$ be such that there is an embedding $f : [n] \to \ell_2$ with $\delta(i,j) = \|f(i) - f(j)\|$.

We can assume without loss of generality that $f(n) = 0$. Then we have:

$$\begin{aligned}
\delta^2(i,j) &= \|f(i) - f(j)\|^2 && (1)\\
&= \|f(i)\|^2 - 2\langle f(i), f(j)\rangle + \|f(j)\|^2 && (2)\\
&= \delta^2(i,n) - 2\langle f(i), f(j)\rangle + \delta^2(j,n) && (3)
\end{aligned}$$

(where $\langle \cdot, \cdot \rangle$ is the usual Euclidean inner product).

Thus we have the identity, for all $i$ and $j$,

$$\langle f(i), f(j)\rangle = \frac{1}{2}(\delta^2(i,n) + \delta^2(j,n) - \delta^2(i,j)) \tag{4}$$

Now define the $(n-1)\times(n-1)$ symmetric matrix $M_\delta = (\frac{1}{2}(\delta^2(i,n)+\delta^2(j,n)-\delta^2(i,j)))_{1\leq i,j\leq n-1}$ (we only need $n-1$ rows and columns since $f(n) = 0$). The above gives us that $M_\delta$ is positive semidefinite. In addition, if $f$ maps to $\mathbb{R}^d$, then the rank of $M_\delta$ is at most $d$.

The converse of this is true:

**Theorem 1.** *A finite semi-metric $(X, \delta)$ is isometrically embeddable into $\ell_2^d$ if and only if the matrix $M_\delta$ as defined above is positive semidefinite of rank $\leq d$.*

*Proof.* If there is an isometric embedding, the above shows that $M_\delta$ satisfies the given properties. Conversely, if $M_\delta$ is positive semidefinite of rank $\leq d$, that implies there exists an $(n-1)\times d$ matrix $V$ such that $VV^T = M_\delta$. We define our embedding to be:

$$f(i) = \begin{cases} 0 & \text{if } i = n \\ \text{the } i\text{th row of } V & \text{otherwise} \end{cases} \tag{5}$$

Verifying that this $f$ gives an embedding with the correct distances is a routine calculation. □

Based on this theorem, we expect that the matrix corresponding to the graph metric of the square, above, should not be positive semidefinite. We see that this is the case: numbering the vertices from 1 to 4 going clockwise, we have

$$M_\delta = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 2 \\ -1 & 2 & 1 \end{pmatrix} \tag{6}$$

and this gives us:

$$\begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 2 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = -4 < 0 \tag{7}$$

so the matrix is indeed not positive semidefinite.

## 2 $\ell_2$-distortion of the Square

The previous section shows that checking isometric embeddability into $\ell_2$ is straightforward. But if we know something is not isometrically embeddable into $\ell_2$, what can we say about the distortion needed? In the example of the square, we see that the obvious embedding into the unit square in $\ell_2$ gives us distortion $\sqrt{2}$. Can we do better?

We address the question by examining an arbitrary embedding $f : [4] \to \ell_2$. We saw that the matrix $M_\delta$ is not positive semidefinite, but inspired by equation 1 we can define $M_f = (\frac{1}{2}(\|f(i) - f(n)\|^2 + \|f(j) - f(n)\|^2 - \|f(i) - f(j)\|^2)_{1 \le i,j \le n-1}$, and know that this matrix will be positive semidefinite for *any* embedding into $\ell_2$. So for any $x \in \mathbb{R}^n$ ($\mathbb{R}^3$ in our case), $xM_f x^T \ge 0$. Setting $x = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$ and expanding this inequality through a trivial but tedious set of (omitted) computations, we get:

$$\|f(1) - f(2)\|^2 + \|f(2) - f(3)\|^2 + \|(f(3) - f(4)\|^2 + \|f(1) - f(4)\|^2 \\ \ge \|f(1) - f(3)\|^2 + \|f(2) - f(4)\|^2 \tag{8}$$

If the distortion of the embedding is $\alpha$, that means $\delta(i,j) \le \|f(i) - f(j)\| \le \alpha\delta(i,j)$ (ignoring the constant factor $r$ since we can change the scale arbitrarily), and we have that the left hand side above is $\le 4\alpha^2$, while the right hand side is at least 8 by our definition of the metric, so we conclude $\alpha \ge \sqrt{2}$, showing that the obvious embedding really is the best possible.

**Remark 1.** *The computation needed for equation (8) is not specific to the particular metric $\delta$, or to the particular points involved: it is a statement about any four points under the $\ell_2$ norm, and so we can also use this inequality in proofs about arbitrary embeddings. This will be useful in the next section.*

## 3 $\ell_2$-distortion of the Hamming Cube

Recall that the Hamming cube $C_m$ of dimension $m$ is the space of $n = 2^m$ points associated with binary strings of length $m$, where the distance between two points is the number of coordinates (bits) in which they differ (this distance is called the *Hamming distance*, and is identical to the

$\ell_1$ metric on the $m$-dimensional hypercube). The example we worked through so far, the square, is the Hamming cube in dimension 2. In this section we will generalize what we have proven to arbitrary dimension.

**Theorem 2.** *(Enflo, 1969) The distortion needed to embed $C_m$ into $\ell_2$ is exactly $\sqrt{m} = \sqrt{\log_2(n)}$.*

**Remark 2.** *(Open question) Instead of looking at the distortion needed to embed the Hamming cube into $\ell_2$, we could check the distortion needed to embed anything that's embeddable into $\ell_1$ into $\ell_2$. That is, in the notation of the previous lecture, what is the value of $D_n(\ell_1, \ell_2)$? This is not known as a function of $n$. It is known that it is at least $\sqrt{\log_2 n}$ (this follows directly from Theorem 2). It is believed that this bound is tight, but no proof is known. The best known upper bound (which is very recent) is $D_n(\ell_1, \ell_2) = O(\sqrt{\log n} \log \log n)$ (due to Arora, Lee, Naor 2005). Before that the best known bound was $O(\log n)$, and as we will see later anything is embeddable into $\ell_2$ with $O(\log n)$ distortion.*

We now proceed with the proof of Theorem 2.

*Proof.* It is obvious that an embedding with this distortion exists, since we can just embed into the $m$-dimensional unit cube as we did in the case $m = 2$. So we need only show that this is the best possible.

Let $X = C_m$ be the $m$-dimensional Hamming cube, so that $|X| = n = 2^m$. We will prove a distance inequality similar to what we did with the square. Let $E$ be the edges of $C_m$ (that is, the vertex pairs at distance 1), and let $F$ be the long diagonals (the vertex pairs at distance $m$). We have the following lemma:

**Lemma 3.** *For any $f : X \to \ell_2$:*

$$\sum_{(i,j)\in E} \|f(i) - f(j)\|^2 \geq \sum_{(i,j)\in F} \|f(i) - f(j)\|^2 \tag{9}$$

*Proof.* By induction: we showed the case $m = 2$ in the previous section, so we assume the claim holds for $m - 1$ and need to show it remains true for $m$.

$C_m$ consists of two copies of $C_{m-1}$ connected by edges of length 1. We know equation (9) holds within each copy. Let $E_{m-1}$ be the edges in the smaller copies, $F_{m-1}$ be the long diagonals in the smaller copies, and $T$ be the set of edges going between the two copies.

Now, in particular, note that each element of $F$ is also one of the long diagonals of a rectangle whose sides are two elements of $F_{m-1}$ (one from each sub-copy) and two elements of $T$, *and* that these rectangles cover *every* element of $F_{m-1}$ and $T$. This allows us to apply equation (8), and we get:

$$\begin{aligned}
&\sum_{(i,j)\in F} \|f(i) - f(j)\|^2 \\
&\leq \sum_{(i,j)\in F_{m-1}} \|f(i) - f(j)\|^2 + \sum_{(i,j)\in T} \|f(i) - f(j)\|^2 \\
&\leq \sum_{(i,j)\in E_{m-1}} \|f(i) - f(j)\|^2 + \sum_{(i,j)\in T} \|f(i) - f(j)\|^2 && \text{(by the I.H.)} \\
&= \sum_{(i,j)\in E} \|f(i) - f(j)\|^2 && \text{(since } E = E_{m-1} \cup T\text{)}
\end{aligned}$$

so we are done. $\square$

Now, let $f : C_m \to \ell_2$ be given, and suppose that the distortion is $\leq \alpha$. Then (after scaling as appropriate) we have for $(i,j) \in E$, $\|f(i) - f(j)\|_2^2 \leq \alpha^2$, while for $(i,j) \in F$), $\|f(i) - f(j)\|_2^2 \geq m^2$. Since $|E| = m2^{m-1}$ and $|F| = 2^{m-1}$, the lemma implies $\alpha^2 m 2^{m-1} \geq m^2 2^{m-1}$, and we conclude $\alpha \geq \sqrt{m} = \sqrt{\log n}$, as was to be shown. $\square$

# 4 $\ell_2$-distortion of Arbitrary Metrics

With both the square and the Hamming cube, we lower-bounded the distortion by looking at some inequality on Euclidean metrics, and combining it with the definition of distortion. Can we always do this? That is, does every lower-bound on distortion follow from an inequality on Euclidean metrics? Fortunately, the answer is yes, so let's formalize this:

**Theorem 4.** *(Minimum distortion of $\ell_2$ embeddings) If a finite metric space $(X, \delta)$ cannot be embedded into $\ell_2$ with distortion $D$, then there exists a proof of the following form:*
*There are weights $\eta$ and $\phi$ from $\binom{X}{2}$ to $[0, \infty)$ such that for all $f : X \to \ell_2$:*

$$\sum_{(i,j)\in\binom{X}{2}} \eta(i,j)\|f(i) - f(j)\|^2 \geq \sum_{(i,j)\in\binom{X}{2}} \phi(i,j)\|f(i) - f(j)\|^2 \tag{10}$$

$$\sum_{(i,j)\in\binom{X}{2}} D^2\eta(i,j)\delta^2(i,j) < \sum_{(i,j)\in\binom{X}{2}} \phi(i,j)\delta^2(i,j) \tag{11}$$

These weights $\eta$ and $\phi$, when they exist, "prove" that $X$ cannot be embedded with distortion $D$ by giving explicitly a property posessed by all embeddings, but not by any metric within distortion $D$ of the given $\delta$. In our earlier theorems, we implicitly used $\eta$ and $\phi$ giving weight one to the edges and the long diagonals, respectively.

*Proof.* Define the following sets:

$$K = \{(x_{ij}) \in \mathbb{R}^{\binom{X}{2}} \mid (\exists r > 0), (\forall i,j), r\delta^2(i,j) \leq x_{ij} \leq rD^2\delta^2(i,j)\} \tag{12}$$

$$\mathfrak{L}_2 = \{(\|f(i) - f(j)\|_2^2)_{(i,j)\in\binom{X}{2}} | f : X \to \ell_2\} \tag{13}$$

Intuitively, $K$ consists of all functions on pairs of elements of $X$ (not necessarily metrics) with distortion $\leq D$ compared to $\delta$, while $\mathfrak{L}_2$ consists of all vectors of distances on $X$ obtained by embedding into $\ell_2$.

$K$ is a cone: it is easy to see that it is closed under multiplication by positives and under addition by scaling $r$ as needed. Also note that $0 \notin K$, since $r$ must be positive.

$\mathfrak{L}_2$ is also a cone: If we have $u, v \in \mathfrak{L}_2$ then $u+v$ is also in it, because we can take their respective embeddings using $p, q$ dimensions and take their sum as the pairwise embedding into the direct sum $\mathbb{R}^{p+q}$, so that squaring componentwise gives the sum of the norms of the two embeddings. Closure under multiplication by positives is obvious. (Note that this is just the set $\mathrm{NOR}_2$ from the previous lecture).

**Remark 3.** *What we call a cone is called a "convex cone" by some authors. Cones are not always defined to be closed under addition. In this class, we will always require this.*

Since by assumption there is no embedding into $\ell_2$ with distortion $\leq D$, we have $K \cap \mathfrak{L}_2 = \emptyset$. Thus, we can separate the two cones by a hyperplane through the origin. Thus, there must exist $a \in \mathbb{R}^N$ with $\langle a, x \rangle > 0$ for all $x \in K$ (since $0 \notin K$) and $\langle a, x \rangle \leq 0$ for all $x \in \mathfrak{L}_2$.

We define our weight functions as follows:

$$\eta(u,v) = \begin{cases} -a(u,v) & \text{if } a(u,v) \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$\phi(u,v) = \begin{cases} a(u,v) & \text{if } a(u,v \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Plugging these definitions back into the statement of the theorem, we get exactly the inequalities we need.

□

A similar statement applies for all $\ell_p$, $1 \leq p$, by modifying the theorem to use $\mathfrak{L}_p$, similar to the theorem proved last lecture.

Note that this proof is just existential: if the metric space is not embeddable with distortion $D$, it implies there must exist a proof of this form (and similarly with $\ell_p$), but says nothing about how to find it or even how to check whether (10) is satisfied. The difference is that in $\ell_2$, we actually *do* know efficient ways to find such a proof (up to $D + \epsilon$, in time $\log(1/\epsilon)$) via semidefinite programming, whereas for $p < 2$ we do not know such efficient techniques. In some cases (such as $p = 1$), the problem is NP-complete.