Embedding Metrics into Geometric Spaces

Anupam Gupta
Carnegie Mellon University
Metric space \( M = (V, d) \)

- set \( V \) of points
- distances \( d(x, y) \)
- triangle inequality
  \[
  d(x, y) \leq d(x, z) + d(z, y)
  \]
We saw:

- embeddings into distributions over trees
- $\beta$-padded decompositions

Today:

- embeddings into geometric spaces ($\ell_p$ spaces)
- in particular, Euclidean space
Consider real space $\mathbb{R}^m$ with the $\ell^p$ metrics:

for $x, y$ in $\mathbb{R}^m$, and $1 \leq p < \infty$

$$|x - y|_p = \left( \sum_{i} |x_i - y_i|^p \right)^{1/p}$$

$$|x - y|_\infty = \max_{1 \leq i \leq m} |x_i - y_i|$$

Since we will deal with finite (n-point) metrics, we can (and will) give embeddings into finite dimensions.

In this lecture, we will ignore the dimensionality, just say $\ell^p$. 
For $1 \leq p \leq q \leq 2$:

\[ \ell_q \text{ embeds into } \ell_p \]

(Next lecture, we’ll see some ideas behind $\ell_2$ into $\ell_1$.)

For $2 \leq q \leq \infty$:

\[ \ell_2 \text{ embeds into } \ell_q \]

Everything embeds into $\ell_\infty$
Focus on

ℓ1: the “Manhattan” or “taxicab” metric

ℓ2: Euclidean space

ℓ∞: the “max”-norm

And introducing---

\( d(x, y) = \sum |x_i - y_i|^2 \) (ℓ2-square): the square of the Euclidean distance. (does not satisfy triangle inequality)
**Distortion.** Given metrics $M = (V, d)$ and $M' = (V', d')$, and a map $f : M \rightarrow M'$,

- the *expansion* of $f$ is $\max_{x,y \in V} \frac{d'(f(x), f(y))}{d(x, y)}$.
- the *contraction* of $f$ is $\max_{x,y \in V} \frac{d(x, y)}{d'(f(x), f(y))}$.
- the *distortion* of $f$ is

$$\text{expansion} \times \text{contraction} = \max_{x,y \in V} \frac{d'(f(x), f(y))}{d(x, y)} \times \max_{x,y \in V} \frac{d(x, y)}{d'(f(x), f(y))}.$$
We say that a metric \((V,d)\) “embeds in \(\ell^p\)” or “is in \(\ell^p\)” if it embeds isometrically (i.e., with distortion 1) into \(\ell^p\).

Hence, we are using \(\ell^p\) to denote both the metric space, and also the class of metrics that embed into \(\ell^p\).

(I’ll try to be careful when it’s ambiguous.)
Theorem: Every metric \((V,d)\) embeds into \(\ell_\infty\).

Consider the Frechét map \(F\) with \(|V|\) coordinates.

\[
f_v(x) = d(v, x) \quad \forall v \in V
\]

\[
F(x) = \bigoplus_v f_v(x)
\]
\[ \| F(x) - F(y) \|_0 = \max_{v \in V} | f_v(x) - f_v(y) | \]

\[ = \max_{v \in V} | d(v,x) - d(v,y) | \]

\( v = y \)

\[ | d(v,v) - d(v,y) | = | d(y,x) - 0 | = d(x,y) \]

\( \sum \)

\[ d(x,v) \leq d(y,v) + d(x,y) \]

\[ d(x,v) - d(y,v) \leq d(x,y) \]

\[ d(y,v) - d(x,v) \leq d(x,y) \text{ similarly} \]
how about ℓ2?
And \( \ell_1 \)?
Given a metric \((V,d)\)
how well does it embed into
\(\ell^p\) spaces?
**Upper Bound:**

\[(V/d) \xrightarrow{\log n} \ell_1 \]

\[\text{Bourgain} \]

\[\frac{\log n}{p} \xrightarrow{} \ell_p \]

**Lower Bound:**

\[O(\log n) \xrightarrow{} \ell_1, \ell_2 \quad [\text{LLR}]\]

\[\Omega(\log n/p) \xrightarrow{} \ell_p \quad [\text{M}]\]
tree metrics

\[
\begin{align*}
\text{min}(L_2) & \rightarrow l_2 \\
\text{logloss} & \rightarrow l_1 \\
\text{logloss} & \rightarrow l_2
\end{align*}
\]
planar metrics

planar graph \[ \sqrt{\log n} \rightarrow l_2 \]
\( \ell_2 \text{-squared metrics} \)
These were all “uniform” results.

What about the algorithmic question:

**Given a metric (X,d), what is the smallest distortion $D$ possible for embedding this metric?**

$\ell_\infty$: trivial

$\ell_1$: NP-hard

*open question: $o(\log n)$ approximation for this problem.*
optimal embeddings into $\ell_2$

$$\bar{0} = \bar{x}_1 \bar{x}_2 \cdots \bar{x}_n$$

$$d(i,j)^2 = \|x_i - x_j\|_2^2 = \|x_i\|^2 + \|x_j\|^2 - 2 \langle x_i, x_j \rangle$$

$$= d(i,i)^2 + d(j,j)^2 - 2 \langle x_i, x_j \rangle$$

$$\langle x_i, x_j \rangle = \frac{1}{2} (d(i,i)^2 + d(j,j)^2 - d(i,j)^2)$$

$$A_{ij} = \langle x_i, x_j \rangle$$
• general metric $\to \ell_1$ with distortion $O(\log n)$

• general metric $\to \ell_2$ with distortion $O(\log^{3/2} n)$

• general metric $\to \ell_2$ with distortion $O(\log n)$

• exists a metric $\to \ell_2$ requires distortion $\Omega(\log n)$
Use a new dimension for each edge.
\( \ell_1 \) forms a convex cone

\[ d \in \ell_1, \quad d' \in \ell_1 \]

\[ d(x, y) = d(x, y) + d'(x, y) \]

\[ \bar{d} \in \ell_1 \]

closed under addition.

\[ d \in \ell_1, \quad \alpha d \in \ell_1 \quad \forall \alpha \geq 0 \]

\[ d_1, d_2 \in \ell_1 \]

\[ \alpha d_1 + \beta d_2 \in \ell_1 \]
\[(v, \mu \subseteq M) \xrightarrow{O(\log n)} \text{distributions} \xrightarrow{1} \ell_1\]

\[d(x, y) = \sum_{t} p_t \cdot d_t(x, y) \in \ell_1\]
general metric → $\ell_1$ with distortion $O(\log n)$

• general metric → $\ell_2$ with distortion $O(\log^{3/2} n)$

• general metric → $\ell_2$ with distortion $O(\log n)$

• exists a metric → $\ell_2$ requires distortion $\Omega(\log n)$
Give a map $F$ that has $\log \Delta$ groups of coordinates one for each “distance-scale”

$$F = \bigoplus_k f_k$$

where $f_k$ handles distances $\approx 2^k$

1. Each of the maps $f_k$ does not expand any distances

$\Rightarrow$ $F$ has expansion $\sqrt{\log \Delta}$

$$\|F(x) - F(y)\|_2^2 = \sum_{k=0}^{\log \Delta} \|f_k(x) - f_k(y)\|_2^2 \leq \sum_k d(x, y)^2$$

$$\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$$

$$\|F(x) - f_k(y)\| \leq \sqrt{\log \Delta} \cdot d(x, y)$$
Give a map $F$ that has $\log \Delta$ groups of coordinates one for each “distance-scale”

$$F = \bigoplus_k f_k$$

where $f_k$ handles distances $\approx 2^k$

1. Each of the maps $f_k$ does not expand any distances
   $$\Rightarrow F \text{ has expansion } \sqrt{\log \Delta}$$

2. if $d(x,y)$ in $(2^k, 2^{k+1}]$, then
   $$| f_k(x) - f_k(y) |_2 \geq 2^k/O(\log n)$$
   $$\Rightarrow | F(x) - F(y) |_2 \geq d(x,y)/O(\log n)$$
   $$\Rightarrow O(\log n (\log \Delta)^2)$$
Given a metric and a value $k$, want map $f_k$ such that

1. the map $f_k$ does not expand any distances
2. if $d(x,y)$ in $(2^k, 2^{k+1}]$, then $|f_k(x) - f_k(y)|_2 \geq 2^k/O(\log n)$

Let’s use padded decompositions again!
A metric \((V,d)\) admits \(\beta\)-padded decompositions, if for every \(\delta > 0\), we can output a random partition

\[ V = V_1 \cup V_2 \cup \ldots \cup V_p \]

1. each \(V_j\) has diameter \(\leq \delta\)
2. \(\Pr[B(x,\rho) \text{ split }] \leq \frac{\rho}{\delta} \cdot \beta\)
Repeat \( t = 1 \ldots L = O(\log n) \) times

1. take a random partition with \( \delta = 2^k \)
2. color each “cluster” red or blue u.a.r.
3. if \( x \) in red \( V_j \), set \( f_{kt} = 0 \) else (\( x \) in blue \( V_j \)) set \( f_{kt} = d(x, V \setminus V_j) \)
4. set \( f_k = 1/L \times \bigoplus_t f_{kt} \)

Claim 1: \( f_{kt} \) does not expand distances (hence neither does \( f_k \))
Repeat $t = 1 \ldots L = O(\log n)$ times

1. take a random partition with $\delta = 2^k$
2. color each “cluster” red or blue u.a.r.
3. if $x$ in red $V_j$, set $f_{kt} = 0$ else (x in blue $V_j$) set $f_{kt} = d(x, V \setminus V_j)$
4. set $f_k = 1/L \times \bigoplus_t f_{kt}$

Claim 2: if $d(x,y) > 2^k$ then $|f_k(x) - f_k(y)| \geq 2^k/O(\log n)$
Given a metric and a value $k$, now we have map $f_k$ such that

1. the map $f_k$ does not expand any distances
2. if $d(x,y) \in (2^k, 2^{k+1}]$, then $|f_k(x) - f_k(y)|_2 \geq 2^k/O(\log n)$

Hence we get the $O(\log n \times \sqrt{\log \Delta})$ embedding into $\ell_2$. If you have a $\beta$-padded decomposition the idea really gives $O(\beta \sqrt{\log \Delta})$ embedding into $\ell_2$. [Rao]
✓ general metric $\rightarrow \ell_1$ with distortion $O(\log n)$

✓ general metric $\rightarrow \ell_2$ with distortion $O(\log^{3/2} n)$

• general metric $\rightarrow \ell_2$ with distortion $O(\log n)$

• (exists a metric $\rightarrow \ell_2$ requires distortion $\Omega(\log n)$)
Big picture:
suppose we have non-negative values $a_{ij}$ and $b_{ij}$
and
\[ R(d) = \frac{\sum a_{ij} d_{ij}^2}{\sum b_{ij} d_{ij}^2} \]
we will show metrics $d$ for which we can set these $a$’s, $b$’s such that $R(d)$ is $1/(n \log^2 n)$
but $R(d’) for any $d’$ in $\ell^2$ is at least $1/n$.
Hence must need $\Omega(\log n)$ distortion for $d \rightarrow \ell^2$. 
to start off, some background

1. (edge)-expander graphs.
2. constant-degree expander graphs.
3. logarithmic diameter
4. most points are log distance apart
5. spectral properties of expander graphs

\[
\left( \frac{n^2}{2} \right)(1 - o(1)) \\
\leq o(n) \\
\left( \frac{n}{\log n} \right) < \log n \cdot c
\]
✓ general metric → $\ell_1$ with distortion $O(\log n)$

✓ general metric → $\ell_2$ with distortion $O(\log^{3/2} n)$

• general metric → $\ell_2$ with distortion $O(\log n)$

✓ exists a metric → $\ell_2$ requires distortion $\Omega(\log n)$
Bourgain/Matousek theorem

**Theorem:** General metric $\rightarrow \ell^p$ with distortion $O((\log n)/p)$.

Different kind of scale-based embeddings.
proof idea
✓ general metric $\to$ $\ell_1$ with distortion $O(\log n)$

✓ general metric $\to$ $\ell_2$ with distortion $O(\log^{3/2} n)$

✓ general metric $\to$ $\ell_2$ with distortion $O(\log n)$

✓ exists a metric $\to$ $\ell_2$ requires distortion $\Omega(\log n)$
app: sparsest cut
better: sparsest cut via SDPs
the ARV structure lemma
l2-squared metrics into l2