1 Better Distortion with Distribution

There is a well known lowerbound for the distortion of embedding a metric space (X,d) into ℓ_1.

Theorem 1.1 *For any metric space (X,d) on n points, one has*

$$(X,d) \xhookrightarrow{\Omega(\log n)} \ell_1.$$

For ℓ_2 the lowerbound is still $\Omega(\log n)$ [1].

Recall that we want to find a $(O(k),(1+\varepsilon)c)$-outlier embedding into ℓ_2 for any metric space (X,d) which admits a (k,c)-outlier embedding into ℓ_2. If we can do this deterministically, we actually find an embedding of the outlier points into ℓ_2 with distortion $O(k)$, which contradicts the lowerbound. However, maybe we can do $O(k)$ via embedding into some distribution of ℓ_2 metrics.

Let (X,d) be a finite metric space and let $\mathcal{Y} = \{(Y_1,d_1),\dots(Y_h,d_h)\}$ be a set of metric spaces. Let π be a distribution of embeddings into $\mathcal{Y}$. The original metric space (X,d) embeds into π with distortion D if there is an $r>0$ such that for all $x,y\in X$,

$$r \le \frac{\mathrm{E}_{i\leftarrow\pi}[d_i(\alpha_i(x),\alpha_i(y))]}{d(x,y)} \le Dr.$$

SODA23 paper also embeds (X,d) into distribution. We call this kind of embeddings stochastic embedding.

Example: Random Trees Consider the problem of embedding some finite metric into a tree metric. We can get an $O(n)$ lowerbound via the unit edge length cycle C_n. However, if embedding into distortions is allowed, we can do $O(\log n)$.

Theorem 1.2 (Bartal) *Let (X,d) be a metric space on n points with diameter Δ, let $\mathcal{D}T$ be the set of tree metrics that dominate d, there is a distribution π on $\mathcal{D}T$ such that (X,d) embeds into π with distortion $O(\log n)$.*

2 Stochastic Embedding into ℓ_2

We first ignore the outlier condition and see if stochastic embeddings break the $\Omega(\log n)$ lowerbound.

Theorem 2.1 (Bourgain) *For any metric space (X,d) and for any p, there is an embedding of (X,d) into $\ell_p^{O(\log^2 n)}$ with distortion $O(\log n)$.*

Bourgain develops an algorithm that finds a desired embedding with probability at least 1/2.[2] For the ℓ_2 case, the embedding has the following bounds:

Expansion $\|f(x)-f(y)\|_2 \le O(\log n)d(x,y)$

Contraction $\|f(x)-f(y)\|_2 \ge \frac{d(x,y)}{O(1)}$

[1] https://web.stanford.edu/class/cs369m/cs369mlecture1.pdf
[2] https://home.ttic.edu/~harry/teaching/pdf/lecture3.pdf

The contraction bound is almost tight. Let K be the dimension of the target space. For the expansion bound, we have

$$\begin{aligned}
\|f(x)-f(y)\|_2 &= \left(\sum_{i=1}^{K} |f_i(x)-f_i(y)|^2\right)^{1/2} \\
&\le \left(\sum_{i=1}^{K} d(x,y)^2\right)^{1/2} \\
&= \sqrt{K} d(x,y) \\
&= O(\log n) d(x,y)
\end{aligned}$$

One thing we can try is to tighten the second line. Recall that for each dimension i a random subset $S_i \subseteq X$ is selected and the value of $f_i(x)$ is $\min_{s \in S_i} d(x,s)$. We want to show that for any fixed $x, y \in X$ and any dimension i the event that distance $|f_i(x)-f_i(y)|^2$ is much smaller than $d(x,y)^2$ happends with high probability.