CS-67720 Metric Embedding Theory and Its Algorithmic Applications

Lecture 4

Lecturers: Nova Fandina and Yair Bartal *Fall 2018/19*

4.0.1 Dimension Reduction with ℓ_q-distortion

We continue the study of dimension reduction with average guarantees. Recall the theorem we proved:

Theorem 4.1. *Given any n-point $X \subset \ell_2^d$, any integer $k \geq 2$, and any $1 \leq q < k$, there is an embedding $f : X \to \ell_2^k$ such that*

$$\ell_q\text{-}dist(f) = 1 + O\left(\frac{q}{k-q}\right) + O\left(\frac{1}{\sqrt{k}}\right).$$

We note that it can be proved that the result is tight for all $q \geq \sqrt{k}$. The embedding in the theorem is the JL transform (implemented with Gaussian entries). The analysis we did worked for $q < k$. Applying some more technical considerations, we actually can prove that the JL transform, normalized by an appropriate factor, has bounded ℓ_q-distortion for any $q \geq k$:

Theorem 4.2. *Let $X \subset \ell_2^n$ be any n-point set. The following two statements hold:*

For any $k \geq 1$, there is an embedding $f : X \to \ell_2^k$ with ℓ_k-$dist(f) = O\left((\sqrt{\ln n})^{\frac{1}{k}}\right)$.

For any $k \geq 1$ and any $k < q$ there is an embedding $g : X \to \ell_2^k$ with ℓ_q-$dist(f) = O\left(n^{\left(\frac{1}{k}-\frac{1}{q}\right)}(\ln n)^{\frac{1}{4}}\right)$, where the constant of the growth rate depends on $1/(q-k)$.

These bounds are almost tight:

Theorem 4.3. *Any $f : E_n \to \ell_2^k$ has ℓ_k-$dist(f) = \Omega((\sqrt{\log n})^{1/k}/k^{1/4})$, and ℓ_q-$dist(f) = \Omega(n^{1/(2k)-1/(2q)})$, for any $q > k \geq 1$.*

We present a proof for the case of $q = k$. Before we get into the proof, we state several claims and define the objects we will use.

In the homework exercise, you will show the following claim:

Claim 4.4. *Assume that for any $q, k \geq 1$, for any non-expansive $g : E_n \to \ell_2^k$, it holds that ℓ_q-$dist(g) = \Omega(\alpha(n))$, for some function $\alpha(n)$. Then for any $h : E_n \to \ell_2^k$, it holds that ℓ_q-$dist(h) = \Omega(\sqrt{\alpha(n/3)})$.*

Thus, it is enough to prove that any non-expansive $f : E_n \to \ell_2^k$ has ℓ_k-$dist(f) = \Omega((\log n)^{1/k}/\sqrt{k})$. In our proof we use the *hierarchical separated tree (HST)* metric spaces, that were introduced in [1].

Definition 4.1. *For $s > 1$, consider a rooted tree T, with labels $\Delta(v) \geq 0$ assigned to each node as follows: the leaves of T have $\Delta(v) = 0$, and for any two nodes $u \neq v$, if u is a child of v, then $\Delta(u) \leq \Delta(v)/s$. Such a tree induces a metric on the set of its leaves: for all $u \neq v$ leaves, define $d_T(u, v) = \Delta(lca_T(u, v))$, where lca states for the least common ancestor. We call this metric space an s-HST metric.*

For $t \geq 1$ and $\delta > 1$, let

$$= \left\{T \mid T \text{ is a 2-HST on } n \text{ leaves, with degree at most } 2^t\text{, and with } \Delta(root(T)) = \delta\right\}.$$

The proof of the theorem follows from the following claim, by the norm relation between $\|\|_2$ and $\|\|_\infty$:

Claim 4.5. *Any non-expansive embedding $g : E_n \to \ell_\infty^k$ has ℓ_k-dist$(g) = \Omega((\log n)^{1/k})$.*

The proof consists of two steps that are formulated in the following two lemmas. First, we show that any non-expansive embedding g of E_n into ℓ_∞^k can be "improved" by embedding its image set into 2-H_k^1. Namely, there exists an embedding $h : g(E_n) \to$ 2-H_k^1 such that $h \circ g$ is a non-expansive, and ℓ_k-*dist*$(g) \geq \ell_k$-*dist*$(h \circ g)$. Second, we show that any non-expansive embedding $F : E_n \to$ 2-H_k^1 has $(\ell_k\text{-}dist(F))^k = \Omega(\log n)$.

Lemma 4.6. *For a non-expansive $g : E_n \to \ell_\infty^k$ there is $h : g(E_n) \to$ 2-H_k^1 s.t. ℓ_k-dist$(g) \geq \ell_k$-dist$(h \circ g)$.*

Proof. Since g is non-expansive, w.l.o.g. $g(E_n) \subset [0,1]^k$. Note that this cube has edge and diagonal lengths 1 (in ℓ_∞-norm). Divide this cube into 2^k sub-cubes with edge length 1/2, and diagonal 1/2, denote these sub-cubes by $C_{1/2}^{(j)}$, $1 \leq j \leq 2^k$. We build an embedding $h : g(E_n) \to T$, $T \in$ 2-H_k^1, recursively as follows. The root of the tree T is defined to be $r(T)$, with $\Delta(r(T)) = 1$. For each cube $C_{1/2}^{(j)}$ recursively build the sub-trees T_j, with the leaves being the points in $g(E_n) \cap C_{1/2}^{(j)}$. Define$\Delta(r(T_j)) = 1/2$. Set the (at most 2^k) children of $r(T)$ to be T_j. If for some j, $g(E_n) \cap C_{1/2}^{(j)} = \emptyset$, then T_j is an empty tree. If $|g(E_n) \cap C_{1/2}^{(j)}| = 1$, then the tree T_j contains only the root with label 0.

Note that, trivially, the embedding $h \circ g$ is non-expansive. Also, by the construction, for any $u \neq v \in g(E_n)$, $d_T(u,v) \geq \|u - v\|_\infty$, implying that ℓ_k-*dist*$(g) \geq \ell_k$-*dist*$(h \circ g)$, which completes the proof. $\square$

To complete the proof of the theorem, it remains to prove the following lemma:

Lemma 4.7. *For any $k \geq 1$, for any non-expansive embedding $F : E_n \to$ 2-H_k^1, ℓ_k-dist$(F) = \Omega((\log n)^{1/k})$.*

Proof. For a non-expansive embedding $F : E_n \to$ 2-H_k^1, let $T \in$ 2-H_k^1 be its image tree. Note that the ℓ_k-distortion of the embedding is solely defined by topology of T. Define the *k-weight* of T by $w_k(T) = (\ell_k\text{-}dist(F))^k \cdot \binom{n}{2}$. For all $k, n \geq 1$, $\delta \leq 1$ define $S_k(n,\delta) = \min\{w_k(T) | T \in H_k^\delta\}$, $S_k(0,\delta) = 0$. Using the notations, we have to prove that $S_k(n,1) \geq \Omega(n^2 \log n/k)$, for all $k \geq 1$, which we show by induction on n. We will prove for any $k \geq 1$, for all $n \geq 2$, that $S_k(n,1) \geq n^2 \log n/(4k)$. For $n = 2$, $S_k(2,1) = 1 \geq \frac{1}{k}$, for all $k \geq 1$. Assume the claim holds for all $n' < n$. Define the set of constraints $D = \{0 \leq n_i \leq n,\ 1 \leq i \leq 2^k,\ \sum_i n_i = n\}$. Then it holds that $S_k(n,1) = \min_D\{\sum_i S_k(n_i, 1/2) + \sum_{i \neq j} n_i n_j\}$. Note that for all n and k, by the definition $S_k(n,1/2) = 2^k S_k(n,1)$. Substituting this and applying the induction assumption on each $S_k(n_i,1)$, we arrive at $S_k(n,1) \geq \min_D\{\frac{1}{4k} \cdot 2^k \sum_i n_i^2 \log n_i + \sum_{i \neq j} n_i n_j\}$. We can write $\sum_{i \neq j} n_i n_j = \frac{1}{2}((\sum_i n_i)^2 - \sum_i n_i^2) = \frac{n^2}{2} - \frac{1}{2}\sum_i n_i^2$, therefore $S_k(n,1) \geq \min_D\{\sum_i n_i^2(\frac{2^k}{4k}\log n_i - \frac{1}{2}) + \frac{1}{2}n^2\}$. Consider the real function $f(x) = x^2(\frac{2^k}{4k}\log x - \frac{1}{2})$, on $x \geq 2$. It is convex for all $k \geq 1$. Therefore, by Jensen's inequality the minimum value of the above minimization program is obtained on $n_i = n/2^k$, for all i. Therefore, we have $S_k(n,1) \geq \frac{n^2 \log(\frac{n}{2^k})}{4k} + (\frac{1}{2} - \frac{1}{2^{k+1}})n^2 = \frac{n^2 \log n}{4k} + (\frac{1}{2} - \frac{1}{4} - \frac{1}{2^{k+1}})n^2 \geq \frac{n^2 \log n}{4k}$ for all $k \geq 1$, as required $\square$

4.1 Multidimensional Scaling

We continue the study of embedding into low dimensional normed spaces with the Multidimensional scaling problem. This problem was studied since early 60's in the various practical communities (biology, physics, psychology and more) [4, 2, 3]. Given an n point metric space (X, d_X) and an integer $k \geq 1$ we seek to find an optimal embedding $f : X \to \ell_2^k$, while considering different measures of the (average) quality. In additional to the ℓ_q-distortion, the following measures are of particular interest in practice.

For a given $f : X \to Y$, denote by $d_{ij} = d_X(x_i, x_j)$, and by $\hat{d}_{ij} = d_Y(f(x_i), f(x_j))$. Given any $q \geq 1$,

$$Stress_q(f) = \left(\frac{\sum_{1 \le i \ne j \le n} \left| d_{ij} - \hat{d}_{ij} \right|^q}{\sum_{1 \le i \ne j \le n} (d_{ij})^q} \right)^{1/q}, \quad Stress^*{}_q(f) = \left(\frac{\sum_{1 \le i \ne j \le n} \left| d_{ij} - \hat{d}_{ij} \right|^q}{\sum_{1 \le i \ne j \le n} (\hat{d}_{ij})^q} \right)^{1/q}.$$

$$Energy_q(f) = \left(\frac{1}{\binom{n}{2}} \sum_{1 \le i \ne j \le n} \left(\frac{\left| d_{ij} - \hat{d}_{ij} \right|}{d_{ij}} \right)^q \right)^{1/q}, \quad REM_q(f) = \left(\frac{1}{\binom{n}{2}} \sum_{1 \le i \ne j \le n} \left(\frac{|d_{ij} - \hat{d}_{ij}|}{\min\{d_{ij}, \hat{d}_{ij}\}} \right)^q \right)^{1/q}.$$

Currently, there are no exact solutions to the general MDS problem. On the other hand, there are lots of heuristic algorithms that are frequently used in the practical community. Perhaps, one of the most famous heuristics, is so called classical MDS algorithm, based on the SVD/PCA algorithms.

Given any n-point $X \subset \ell_2^d$, PCA finds the k-dimensional subspace that contains the maximal variance of X. Basically, PCA finds a projection f onto k-dimensional subspace with the minimal value of $\sum \|x_i - f(x_i)\|_2^2$, which can be shown to result in the minimal value of $\sum(\|x_i - x_j\|_2^2 - \|f(x_i) - f(x_j)\|_2^2)$. Applying PCA on E_n, for $n > 2k$ would result in roughly half of the points to be mapped to the same point, implying ℓ_q-distortion, $Energy_q$, REM_q to be infinite, and $Stress_q \sim 1 - k/n$.

Therefore, an interesting question would be to find an approximation solution to the MDS problem. It can be shown, that all the above measures are bounded by REM_q, and in addition that REM_q itself is strongly related to the ℓ_q-distortion measure. Using similar considerations as in analyzing ℓ_q-*dist* of the JL transform, we can obtain the following result.

Claim 4.8. *Given any n-point subset $X \subset \ell_2^d$, an integer $k \ge 1$, and $1 \le q \le k-1$, the JL transform $f : X \to \ell_2^k$ is such that with constant probability the following hold*

$$Stress_q(f), Stress_q^*(f), Energy_q(f), REM_q(f) = O\left(\sqrt{q/k}\right).$$

The bounds are tight for any $k > q \ge 2$.

In this lecture we will show that analyzing the JL transform for these notions (and using some properties of the composition of embeddings under these notions), as well as using techniques from a convex optimization theory, we can approximate the optimal solution of the MDS problem, to the small additive constant factor. Let us formally state the result.

Theorem 4.9. *Given an n point metric space (X, d_X), an integer $k \ge 3$, $2 \le q \le k-1$, and given one of the above measures, denoted by $\mathbb{M}_q$, denote $OPT_{\mathbb{M}_q} = \inf_{f:X \to \ell_2^k}\{\mathbb{M}_q(f)\}$. There is a randomized, polynomial time embedding $F : X \to \ell_2^k$ such that with high probability*

1. *ℓ_q-dist$(F) = \left(1 + O\left(\frac{1}{\sqrt{k}} + \frac{q}{k-q}\right)\right) \cdot OPT_{\ell_q\text{-}dist}$.*

2. *$\mathbb{M}_q(F) = O\left(OPT_{\mathbb{M}_q}\right) + O\left(\sqrt{\frac{q}{k}}\right)$, for $\mathbb{M}_q \in \{S\text{-}Energy_q, Energy_q, Stress_q, Stress^*{}_q\}$.*

In this lecture we will show the (part of) first item of the above theorem. It should be noted that rest of the items are obtained in a similar way. In what follows, we focus on the ℓ_q-distortion measure.

The general idea is to perform two steps: first, find an embedding $f : X \to \ell_2^n$ that optimizes ℓ_q-distortion. Second, apply JL to reduce the dimension into k dimensions. The embedding $F : X \to \ell_2^k$ is then defined to be the composition of the two embeddings. We have to show that we can find an optimal embedding into Euclidean space (in polynomial time), and that the ℓ_q-distortion of the composition of two embeddings is somehow related to the ℓ_q-distortion of the embeddings being composed.

Convex Optimization. Optimization plays central role in practice and theory of CS. Many important problems in real life are optimization problems in their nature. There is a whole branch in math/CS that develops efficient solutions to various types of optimization problems (if possible).

The convex optimization is a problem of minimizing a convex function over a convex set. It is a whole class of various special case convex problems: linear programming, semidefinite-programming and more. For each case there are efficient algorithms designed for a specific type of the problem. For a general convex optimization problem there are also a variety of efficient algorithms that approximate the optimal solution to any level of precision.

4.1.1 Computing the optimal embedding into ℓ_2

We present two rsults: optimal embedding with the w.c. distortion (SDP), and optimal embedding with the optimal ℓ_q-distortion (CP).

We start with few notations and lemmas.

Definition 4.2. *Let A be an $n \times n$ symmetric real matrix. Then A is said to be positive semidefinite, denoted by $A \succeq 0$, if for all $x \in \mathbb{R}^n$, $x^T Ax \geq 0$. Also, A is PSD iff all the eigenvalues of A are non-negative.*

The following standard linear-algebraic fact provides an equivalent definition of the PSD matrix.

Lemma 4.10. *$A_{n\times n}$ is PSD iff $A = B^T B$ for some real $n \times n$ matrix.*

Note that the set of all the PSD $n \times n$ matrices form a closed convex cone. Next we prove the lemma that characterizes the squared Euclidean metrics.

Lemma 4.11. *Let $V = \{0, 1, \ldots, n\}$ be an $(n + 1)$-point set, and let $\mathbf{z} = (z_{ij})_{0\leq i\neq j\leq n}$ be a vector of real numbers, such that $z_{ij} = z_{ji}$ and $z_{ii} = 0$. Then there exists an embedding $f : V \to \ell_2$ such that $z_{ij} = \|f(i) - f(j)\|_2^2$, for all $i \neq j$, iff the $n \times n$ matrix defined by $G = (g_{ij})_{i,j=1}^n$, $g_{ij} := \frac{1}{2}(z_{0i} + z_{0j} - z_{ij})$, $i, j = 1, 2, \ldots n$ is positive semidefinite.*

Proof. $\Rightarrow$ Assume there exists $f : X \to \ell_2^d$ s.t. $z_{ij} = \|f(i) - f(j)\|_2^2$. Denote the vectors of the image of f by $p_0, p_1, \ldots, p_n$. Then we get for all $0 \leq i, j \leq n$,

$$z_{ij} = \|p_i - p_j\|_2^2 .$$

Denote $v_i = p_i - p_0$. Then, we get for all $1 \leq i, j \leq n$,

$$g_{ij} = \frac{1}{2}(z_{0i} + z_{0j} - z_{ij}) = \frac{1}{2}\left(\|v_i\|_2^2 + \|v_j\|_2^2 - \|v_i - v_j\|_2^2\right) =< v_i, v_j >,$$

since $\|v_i - v_j\|_2^2 = \|v_i\|_2^2 + \|v_j\|_2^2 - 2 < v_u, v_j >$. Thus we obtain that G is a Gram matrix of the set of n vectors $\{v_i\}$, as required.

$\Leftarrow$ Assume G is a PSD matrix. Then there is some matrix $B_{n\times n}$ such that $G = B^T B$. Denote by $v_1, v_2, \ldots, v_n$ the columns of B. Then we define $f : X \to \ell_2^n$ by $f(i) = v_i$, and $f(0) = 0$. Then, we have to show that $z_{ij} = \|v_i - v_j\|_2^2$. Note that $g_{ij} =< v_i, v_j >= \frac{1}{2}(z_{0i} + z_{0j} - z_{ij})$, and $g_{ii} =< v_i, v_i >= z_{0i} - \frac{1}{2}z_{ii} = z_{0i}$, which completes the proof. □

The SDP (semidefinite programming) is a problem of optimizing a linear function of some real variables over a set that is specified by a a linear inequalities and by the requirement that the matrix of the variables is a PSD. That is, the optimization domain is an intersection of a convex cone (the set of PSD matrices) with affine hyperplanes. This object is called a *spectrahedron*.

Thus, we can formulate the following SDP, which computes the smallest w.c. distortion of embedding into ℓ_2^n. Given an $(n + 1)$-point metric space (X, d_X), define the following variables: z_{ij} describes the squared distance of a pair $x_i, x_j \in X$ after an embedding $f : X \to \ell_2^n$ has been applied, D- describes the squared expansion of the embedding f. Following SDP program finds an embedding $f : X \to \ell_2^n$ with minimal distortion.

$$\begin{array}{lll} \text{minimize} & D & \\ \text{subject to} & d_X(x_i,x_j)^2 \le z_{ij} \le D \cdot d_X(x_i,x_j)^2, & i \ne j = 0,\dots,n-1, \\ & g_{ij} = \frac{1}{2}(z_{0i} + z_{0j} - z_{ij}),\ G = (g_{ij}) \succeq 0 & i \ne j = 0,\dots,n-1, \\ & z_{ij} \ge 0 & i \ne j = 0,\dots,n-1. \end{array}$$

Next we show how to compute optimal embedding for the ℓ_q-distortion measure. The idea is to use the above program, and optimize the required function. Next is the program itself.

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{0\le i<j\le n-1} \left(\max\left\{\frac{z_{ij}}{(d_{ij})^2}, \frac{(d_{ij})^2}{z_{ij}}\right\}\right)^{q/2} & \\ \text{subject to} & g_{ij} = \frac{1}{2}(z_{0i} + z_{0j} - z_{ij}),\ G[i,j] = (g_{ij}) \succeq 0 & i \ne j = 0,\dots,n-1, \\ & z_{ij} \ge 0 & i \ne j = 0,\dots,n-1. \end{array}$$

The above optimization program is an instance of the general convex program (note that the objective and the constraints are all convex functions).

4.1.2 Approximating optimal embedding into ℓ_2^k

Let (X, d_X) be an n-point metric space. In what follows we describe a general paradigm for approximating an optimal embedding for a given quality measure: MDS or ℓ_q-distortion. To approximate the optimal embedding into k dimensions, we perform the following two steps: first, apply a convex optimization problem to get an optimal embedding into n dimensions, next reduce the dimension to k, by applying the JL transform. Denote this embedding by $F : X \to \ell_2^k$. To obtain the final result, we develop several bounds on composition of two embeddings. The following property can be shown: Let (X, d_X) be an n-point metric space, and let $f : X \to \ell_2^d$ be any embedding. Denote by $g : \ell_2^d \to \ell_2^k$ the JL transform. Consider a randomized embedding $h = g \circ f$. Then, for a given $q \ge 1$ it holds that

$$E\left[\ell_q\text{-}dist(g \circ f)\right] \le \ell_q\text{-}dist(f) \cdot \left(E\left[(\ell_q\text{-}dist(g))^q\right]\right)^{1/q}.$$

Since $\ell_q\text{-}dist(f) = OPT^n \le OPT^k$, we obtain that $\ell_q\text{-}dist(F) \le \left(1 + O\left(\frac{1}{\sqrt{k}} + O\left(\frac{q}{k-q}\right)\right)\right) \cdot OPT^k$.

References

[1] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 184–193, 1996.

[2] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling (Monographs on Statistics and Applied Probability)*. Chapman and Hall/CRC, 2nd edition, 2000.

[3] W. J Heiser. Multidimensional scaling with least absolute residuals. In *In H. H. Bock (Ed.) Classification and related methods*, pages 455–462. Amsterdam: NorthHolland, 1988a.

[4] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.