Lecture 6

Lecturers: Yair Bartal and Nova Fandina *Fall 2018/19*
For comments contact fandina@gmail.com

6.1 Range Preserving Embeddings and Its Applications

We continue to study dimension reduction in ℓ_p, for any $p \neq 2$, with "reasonable" accuracy[1]. In the context of nearest neighbor search, reasonable means one that can be used to speed up a solution for ANN problem, defined on ℓ_p spaces. It turns out, that for this sake, the demand on preserving all pairwise distances can be much relaxed, which leads us to the notion of a range preserving embedding.

Definition 6.1. *Let (X, d_x), (Y, d_y) be metric spaces and $[a, b]$ be a range of reals, $a < b$. An embedding $f : X \to Y$ is an $[a, b]$-range preserving with distortion α, if there is $c > 0$, such that $\forall\, x, y \in X$:*

1. *if $a \leq d_x(x, y) \leq b$, then $d_x(x, y) \leq c \cdot d_y(f(x), f(y)) \leq \alpha \cdot d_x(x, y)$,*
2. *if $d_x(x, y) > b$, then $c \cdot d_y(f(x), f(y)) \geq b$,*
3. *if $d_x(x, y) < a$, then $c \cdot d_y(f(x), f(y)) \leq \alpha \cdot a$.*

Definition 6.2. *Let $(X, d_x), (Y, d_y)$ be metric spaces and $R > 1$. We say that X enables R-range embedding into Y with distortion α, if $\forall u > 0$ there is $[u, uR]$-range preserving embedding into Y, with distortion α.*

As we hinted before, range preserving embeddings are useful for solving the NNS problem, i.e., one may use them to reduce time complexity of the preprocessing and query algorithms, by reducing the dimension of the instance. An embedding should keep its guarantees on the distortion, with some constant probability, when a new query point is considered. Also, it should be possible to embed a new point without an information on the other points in the space, such embeddings called *oblivious*. Note that, as we observed before, the JL embedding is oblivious.

Next, we show that the range preserving dimension reduction guarantees can be used to speed up a solution to the approximate NNS problem.

Claim 6.1. *Let $P \subset \ell_p^d$, $|P| = n$, $d' << d$, $\alpha \geq 1$, $\beta \geq 1$, $R > \alpha\beta$, $1 \leq p, t \leq \infty$. Assume that P enables R-range preserving oblivious embedding into $\ell_t^{d'}$, with distortion α. Assume that this embedding keeps its distortion guarantees with a constant probability, for all points in $P \cup v$, for any $v \in \mathbb{R}^d$. If there is an algorithm for $(\beta, \alpha r)$-ARNN in $\ell_t^{d'}$, with pre-processing time $T(n, d')$ and query time $Q(d')$, then there is a randomized algorithm for (R, r)-ARNN for P, in ℓ_p^d, that with constant probability correctly answers a query, and with the following parameters: Pre-processing time: $T(n, d')+$ time to embed the point set P. Query time: $Q(d')+$ time to embed one point.*

Proof. Since P enables R-range preserving, it holds that for $u = r$ there exists an embedding $f : P \to \ell_t^{d'}$ with distortion α, which preserves range $[r, rR]$. Let $q \in \mathbb{R}^d$ be a query point, and let $z \in P$ be a nearest neighbor to q. There are two cases:

[1] Recall, that for $p = 1, \infty$ there is no dimension reduction, and for $p \neq 1, 2, \infty$ the picture is even less clear.

- (case 1): $d_p(q,z) \leq r$. Then, $c \cdot d_t(f(q), f(P)) \leq c \cdot d_t(f(q), f(z)) \leq \alpha \cdot r$, with constant probability (by the definition of the range preserving embedding). Therefore, $(\beta, \alpha \cdot r)$-$ARNN$ on point $f(q)$ will find a point w', such that $c \cdot d_t(f(q), w') \leq \beta \cdot \alpha \cdot r$. Then, $d_p(q,w) \leq R \cdot r$, where $w = f^{-1}(w')$, since otherwise, by the definition of R-range preserving, $c \cdot d_t(f(q), w') \geq R \cdot r$, which is a contradiction, as $R > \alpha\beta$.

- (case 2): $d_p(q,z) > r$. The $(\beta, \alpha r) - ARNN$ always results either NONE or point w' with guarantees as in the previous case. The both results are good for us in this case (the same analysis).

□

We note that the above claim is true for any metric spaces. The meaning of the above claim is that for $\alpha \sim 1 + \epsilon$, and $\beta \sim 1 + \epsilon$, we can solve $(1+3\epsilon)$-ARNN in ℓ_p^d, (potentially) more efficiently, via range preserving dimension reduction (given that good dimension reductions exist). Note that we do need range preservation property for any u (i.e. for any interval $[u, uR]$), since we reduced the ANN problem to the number of range ANN problems (in the above prove we needed it for $u = r$, and R corresponded to the quality of approximation of the ANN).

Note that we didn't use the fact that embedding preserves the distances in the range $[r, Rr]$ up to a factor α, we only used the "separation" property.

Thus, we focus on providing such dimension reductions, with good guarantees for dimension/distortion. We start with the case of $p = 1$.

6.2 Range Preserving Dimension Reduction in ℓ_1

The main result is the following theorem:

Theorem 6.2 ([2]). *Let $R > 1$ and $0 < \epsilon \leq 1/2$. Any n-point $X \subset \ell_1$ enables R-range embedding into a d-dimensional hypercube, $d = O(\frac{R \log n}{\epsilon^3})$, with distortion $1 + \epsilon$.*

Proof. The proof consists of two main steps: An embedding of $X \subset \ell_1$ into a hypercube (of high dimension), with distortion $(1+\epsilon)$, Lemma 6.3; An R-range dimension reduction on a hypercube, with distortion $(1+\epsilon)$, Lemma 6.4. □

Lemma 6.3. *Let $X \subset \ell_1^d$ be a set with aspect ratio Φ. Then, for any $0 < \epsilon \leq \frac{1}{2}$, there is an embedding of X into a hypercube of dimension $d' = O\left(\frac{\Phi d^2}{\epsilon}\right)$, with distortion $(1+\epsilon)$.*

Proof. If all the coordinates of all the vectors in X are integers, then there is an isometric embedding of X into the hypercube: Take the concatenation of unary representations of the coordinates. We extend this idea to a non-integer coordinates, in which case we lose in distortion.

W.l.o.g. one of the points in X is 0. Therefore, the absolute value of a coordinate of each vector in X is at most d_{max}. Therefore, we can translate the set X in such a way that all the coordinates of all the points are positive (in the direction of (d_{max})), without changing the original distances. Let $\delta = \frac{\epsilon d_{min}}{d}$. We build an embedding from X into a hypercube as follows: given any $x = (x_1, x_2, \ldots x_d) \in X$, for each $1 \leq i \leq d$ we match $\acute{x_i} = \delta\lceil\frac{x_i}{\delta}\rceil$. This already incurs a small loss in distortion, but the main problem is that the values of this embedding are not integers. Therefore, we embed each $\acute{x_i}$ into $\tilde{x}_i = \lceil\frac{x_i}{\delta}\rceil$. The embedding is then defined by concatenation of the unary representations of all the $\tilde{x}_i$. Namely, we divide (discretize) the real line into buckets of length δ, and each coordinate x_i is mapped to the index of the bucket it belongs to. The value of δ is carefully chosen so that the loss in precision is small. Note that the mapping $\acute{x_i} \to \tilde{x}_i$ is an isometry (up to δ).

Let $x, y \in \ell_1^d$ be any vectors, then, for all $1 \leq i \leq d$:

$$|y_i - x_i| - \delta \leq |\acute{y_i} - \acute{x_i}| = \delta \left| \left(\left\lceil \frac{y_i}{\delta} \right\rceil - \left\lceil \frac{x_i}{\delta} \right\rceil \right) \right| \leq |y_i - x_i| + \delta.$$

Therefore, it holds that $\|f(y)-f(x)\|_1 \geq \sum_{i=1}^{d}|y_i-x_i| - d\delta = \|y-x\|_1 - \epsilon d_{min} \geq (1-\epsilon)\|y-x\|_1$, and $\|f(y)-f(x)\|_1 \leq \sum_{i=1}^{d}|y_i-x_i| + d\delta = \|y-x\|_1 + \epsilon d_{min} \leq (1+\epsilon)\|y-x\|_1$.

Next, we estimate the dimension of the image hypercube. Each coordinate is bounded by $2d_{max}$ (recall that we have moved the set X), therefore the maximal length of unary representation of some coordinate is $\lceil \frac{2d_{max}}{\delta} \rceil$, implying $d' \leq d\lceil \frac{2d_{max}}{\delta} \rceil \leq \frac{4d \cdot d_{max}}{\delta} = \frac{4d^2 \cdot d_{max}}{\epsilon d_{min}} = \frac{4d^2\Phi}{\epsilon}$.

□

Lemma 6.4 (Range Preserving Dimension Reduction in Hypercube)**.** *Let X be an n-point set on the hypercube of dimension d, let $R > 1$ and $0 < \epsilon < \frac{1}{2}$. Then X enables R-range preserving embedding into hypercube of dimension $d' = O(R\frac{\log n}{\epsilon^3})$ with distortion $1+\epsilon$.*

Proof. Let $u > 0$. We show that there exists a $[u, uR]$-range preserving embedding of n points on hypercube $f : \{0,1\}^d \to \{0,1\}^{d'}$ with distortion $1+\epsilon$, where $d' = O(R\frac{\log n}{\epsilon^3})$. We show the existence of such an embedding via probabilistic argument. Namely, we randomly embed the set into hypercube and show that with positive probability this embedding is $[u, uR]$-range preserving with distortion $1+\epsilon$.

Let $p = \frac{\epsilon}{4uR}$ be a probability parameter. Choose independently at random d' vectors of dimension d, $r^{(j)} = (r_1^{(j)}, r_2^{(j)}, \ldots r_d^{(j)})$, where $r_k^{(j)} = 1$ with probability p, and $r_k^{(j)} = 0$ with probability $1-p$, for all k.

Given any $z \in \{0,1\}^d$ define:

$$f(z) = ((< z, r^{(1)} >) \mod 2, (< z, r^{(2)} >) \mod 2, \ldots (< z, r^{(d')} >) \mod 2).$$

Denote $f_i(z) = (< z, r^{(i)} >) \mod 2$. As f is an additive map (under $\mod 2$ operation) it is enough to show that for any z on the hypercube of dimension d with probability of at least $(1-1/n^2)$ it holds that for some constant $c > 0$:

if $u \leq \|z\|_1 \leq uR$ then $(1-\epsilon)\|z\|_1 \leq c\|f(z)\|_1 \leq (1+\epsilon)\|z\|_1$;

if $\|z\|_1 \leq u$ then $c\|f(z)\|_1 \leq (1+\epsilon)u$;

if $\|z\|_1 \geq uR$ then $c\|f(x)\|_1 \geq (1-\epsilon)uR$.

The existence of the range preserving embedding of the whole set will follow by the union bound argument. (Note that the above requirements are just a scaling version of our original definition). If so, let $z \in \{0,1\}^d$ be any vector, denote $\|z\|_1 = l$. We have to estimate the following probabilities:

1. for $u \leq \|z\|_1 \leq uR$
$$Pr[\|f(z)\|_1 \geq (1+\epsilon)cl],$$
$$Pr[\|f(z)\|_1 \leq (1-\epsilon)cl].$$
2. for $\|z\|_1 \leq u$
$$Pr[\|f(z)\|_1 \geq (1+\epsilon)cu].$$
3. for $\|z\|_1 \geq uR$
$$Pr[\|f(z)\|_1 \leq (1-\epsilon)cuR].$$

We will use the following version of Chernoff bound: Let $X_1, \ldots, X_n$ be an independent random variables taking values in $\{0,1\}$, and let X denote their sum and $\mu = E[X]$. Then for any $0 < \epsilon < 1$: $Pr[X \geq (1+\epsilon)\mu] \leq e^{-\frac{\epsilon^2\mu}{3}}$, and $Pr[X \leq (1-\epsilon)\mu] \leq e^{-\frac{\epsilon^2\mu}{2}}$.

Thus we want to develop estimations on $E[\|f(z)\|_1] = \sum_{i=1}^{d'} Pr[f_i(z) = 1]$. For any $1 \leq i \leq d'$ we compute the probability as follows. Let A_i denote the event that $(f_i(z) = 1)$. Note that the randomized procedure for computing $f_i(z)$ is equivalent to the following randomized process: First, for each coordinate $r^{(i)}$, independently, set its value to be 1 with probability $2p$, and 0 with probability $1-2p$; Secondly, for all the coordinates with value 1 obtained in the previous step, set the value of this coordinate to be 1 with probability $\frac{1}{2}$, and 0 with probability $1/2$. The probability to get at least one 1 at the places of 1's of the vector z, after the first step of the process, equals $1-(1-2p)^l$ (recall that l is the number of 1's of z). From the set of 1's we obtained in the first step, with probability 1/2 we get an odd number of 1's remained after the second step. Therefore, $Pr[A_i] = (1/2)(1-(1-2p)^l)$, implying $E[\|f(z)\|_1] = d'(1/2)(1-(1-2p)^l)$. We consider three cases, for each we provide several bounds on $E[\|f(z)\|_1]$ that will be useful for the specific case. For all the cases we use the following estimation $\forall x > 0, l > 0, l \in \mathbb{N}, 1 - lx \leq (1-x)^l \leq 1 - lx + \frac{1}{2}l^2x^2$.

(1) $E[\|f(z)\|_1] = d'/2(1-(1-2p)^l \leq d'pl = d'\frac{\epsilon}{4uR}l$.
(2) $E[\|f(z)\|_1] = d'/2(1-(1-2p)^l \geq d'(lp - l^2p^2) = d'pl(1-pl) = d'\frac{\epsilon}{4uR}l(1-\frac{\epsilon}{4uR}l) \geq^{l \leq uR} \geq \frac{d'\epsilon l}{4uR}(1-\frac{\epsilon}{4})$.
(3) $E[\|f(z)\|_1] \geq \frac{d'\epsilon l}{4uR}(1-\frac{\epsilon}{4}) \geq \frac{d'\epsilon l}{4uR}\frac{1}{2} \geq^{l \geq u} \geq \frac{d'\epsilon}{8R}$.

Define $c = d'\frac{\epsilon}{4uR}$.

Case of $u \leq l \leq uR$.
$Pr[\|f(z)\|_1 \geq (1+\epsilon)cl] \leq^{(1)} Pr[\|f(z)\|_1 \geq (1+\epsilon)\cdot E[\|f(z)\|_1]] \leq e^{-\frac{\epsilon^2}{3}\cdot E[\|f(z)\|_1]} \leq^{(3)} e^{-\frac{\epsilon^3 d'}{24R}} \leq \frac{1}{\acute{c}n^2}$, where the last inequality holds for taking $\acute{d} = O(\frac{R\log n}{\epsilon^3})$). For another direction, note that

$$Pr[\|f(z)\|_1 \leq (1-\epsilon)cl] \leq Pr[\|f(z)\|_1 \leq \left(1-\frac{\epsilon}{2}\right) E[\|f(z)\|_1]],$$

since from (2) we have $E[\|f(z)\|_1] \geq \frac{ld'\epsilon}{4uR}(1-\frac{\epsilon}{4})$, and for any $\epsilon < 1$ it holds that $\left(1-\frac{\epsilon}{4}\right)\left(1-\frac{\epsilon}{2}\right) \geq 1-\epsilon$. Therefore,

$$Pr[\|f(z)\|_1 \leq (1-\epsilon)cl] \leq Pr[\|f(z)\|_1 \leq \left(1-\frac{\epsilon}{2}\right) E[\|f(z)\|_1]] \leq e^{-\frac{\epsilon^2}{8}\cdot E[\|f(z)\|_1]} \leq^{(3)} e^{-\frac{d'\epsilon^3}{64R}}.$$

For $d' = O(\frac{R\log n}{\epsilon^3})$ (with slightly bigger constant) we obtain that probability is bounded by $\frac{1}{c'n^2}$.

Case of $l \geq uR$.
Recall that, $E[\|f(z)\|_1] = (1/2)d'(1-(1-2p)^l)$. Note that $(1/2)(1-(1-2p)^l)$ is a monotonically increasing (as a function of l), then the minimum value on the range $[uR, d]$ attained on $l = uR$. From (2), we have that for any $l \geq 1$:

$$(d'/2)(1-(1-2p)^l \geq d'\frac{\epsilon}{4uR}l\left(1-\frac{\epsilon}{4uR}l\right).$$

Therefore,

$$E[\|f(z)\|_1] \geq d'\frac{\epsilon l}{4uR}\left(1-\frac{\epsilon}{4uR}l\right),$$

and thus for $l = uR$,

$$E[\|f(z)\|_1] \geq \frac{d'\epsilon}{4}\left(1-\frac{\epsilon}{4}\right).$$

Therefore, since $(1-\frac{\epsilon}{2})E[\|f(z)\|_1] \geq (1-\epsilon)(\frac{\epsilon}{4uR}d')uR$,

$$Pr[\|f(z)\|_1 \leq (1-\epsilon)cuR] \leq Pr[\|f(z)\|_1 \leq \left(1-\frac{\epsilon}{2}\right) E[\|f(z)\|_1]] \leq^{Chernoff} \leq e^{-\frac{\epsilon^2 E[\|f(z)\|_1]}{2}} \leq e^{-\frac{\epsilon^3 d'}{16}},$$

taking $d' = O(\frac{R\log n}{\epsilon^3})$, results in the bound of $\frac{1}{c'n^R}$.
Case of $l \leq u$.
From the estimations of (1) we obtain: $E[\|f(z)\|_1] \leq d'\frac{\epsilon}{4uR}l \leq d'\frac{\epsilon}{4R} = cu$. Therefore, $Pr[\|f(z)\|_1 \geq (1+\epsilon)cu] \leq Pr[\|f(z)\|_1 \geq (1+\epsilon)E[\|f(z)\|_1]$, by applying similar considerations as before, we conclude the proof.

$\square$

6.3 Range Preserving Dimension Reduction in ℓ_p

6.3.1 Dimension reductions for $1 \le p \le 2$

There are (non-trivial) techniques that provide range preserving dimension reduction with the following guarantees.

Theorem 6.5 ([1])**.** *For all $1 \le q \le p \le 2$, for all n-point set $S \subset \ell_p^d$, and for any range parameter $R > 1$, there exists an R-range preserving embedding $f : S \to \ell_q^k$, with distortion $1 + \epsilon$, such that*

$for\ \ q < p,\ \ k = O\left(\frac{poly(R)\log n}{poly(\epsilon)}\right),$
$for\ \ 1 < q = p,\ \ k = O\left(poly(R)\log n \cdot exp\left(\frac{1}{\epsilon}\right)\right),$
$for\ \ 1 = q < p,\ \ k = O\left(\frac{R\log n}{poly(\epsilon)}\right).$

We briefly discuss the methods that imply such results. The main tool is the stable distributions, particularly p-stable distributions.

Definition 6.3. *Let p be any positive real number, the continuous distribution G_p over real numbers called a p-stable distribution, if for any i.i.d. $g_1, g_2, \ldots g_m \sim G_p$, and any real numbers $v_1, v_2, \ldots v_m$, the random variable $\sum_{i=1}^m v_i g_i \sim \|v\|_p \cdot g$, where $g \sim G_p$.*

We have seen that the standard normal distribution $N(0,1)$ is 2-stable. It is known that p-stable distributions exist for any $0 < p \le 2$, and do not exist for any $p > 2$. Let us list several properties of such distributions.

1. The p.d.f. function of G_p is given by: $\forall\, x \in \mathbb{R},\ \ h(x) = \frac{1}{\pi}\int_0^\infty \frac{cos(tx)}{e^{t^p}}dt$. Note that the distribution is symmetric around 0.

2. It is known that $\frac{c_p}{1+x^{p+1}} \le h(x) \le \frac{c'_p}{1+x^{p+1}}$, where c_p, c'_p are constants that depend only on p. We can notice that the tail of this distribution is heavier than the tail of 2-stable distribution.

3. Let's try to imitate the JL transform embedding in our setting, namely embedding $\ell_p \to \ell_q$. For any $v \in \mathbb{R}^d$, let $g_1, g_2, \ldots g_d \sim G_p$, similarly to the JL transform, we define one coordinate of an embedding by $< v, g >$. Let c be a normalized factor. Then,

$$\frac{1}{c} \cdot E\left[\left|\sum_i v_i g_i\right|^q\right] = \frac{1}{c} \cdot E\left[\|v\|_p^q\,|g|^q\right] = \|v\|_p^q,$$

for taking $c = E\left[|g|^q\right]$. We should show that $E\left[|g|^q\right]$ is finite. So let us just compute it:

$$E\left[|g|^q\right] = \int_{-\infty}^{+\infty} |x|^q h(x)\,dx = 2\int_0^\infty x^q h(x)\,dx \approx \int_0^1 x^q \frac{1}{1+x^{p+1}}\,dx + \int_1^\infty x^q \frac{1}{1+x^{p+1}}\,dx$$
$$\approx \int_0^1 x^q\,dx + \int_1^\infty x^{q-p-1}\,dx.$$

Hence, we see that we should require $q < p$, since otherwise the second integral doesn't converge. Performing simple math, we get $E\left[|g|^q\right] \approx \frac{1}{p-q}$, for any $1 \le q < p \le 2$. Thus, we will embed from ℓ_p into ℓ_q, for $q < p$. If we want to perform an actual dimension reduction (i.e. from ℓ_p into ℓ_p, for $p > 1$), we can embed into ℓ_q, for $q < p$, which is close enough to p, such that the loss in the dimension/distortion will be arbitrarily small.

As we have mentioned before, the p-stable distributions have heavy tails and thus the simulation of the JL transform and its analysis will not suffice for the dimension reduction. On the other hand, these distributions do preserve the original norm of any vector, in expectation. So the idea is to use a stronger measure concentration inequalities, for which the embedding should be changed accordingly.

Hoeffding's Inequality. Let $X_1, \ldots, X_k$ be independent real-valued random variables, and assume $|X_i| \leq s$, with probability one. Let $\hat{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$, then for any $z > 0$:

$$Pr[|\hat{X} - E[\hat{X}]| \geq z] \leq 2\exp\left(-\frac{2kz^2}{s^2}\right).$$

Thus, we change the embedding such that each coordinate becomes to be a bounded variable.

Description of The Embedding. For any n-point $S \subset \ell_p^d$, any $1 \leq q < p \leq 2$, we describe the construction of one coordinate of the embedding. Pick i.i.d. $g_1, \ldots, g_d \sim G_p$, and pick uniformly and independently at random an angle ϕ from $[0, 2\pi]$. Fix a (threshold) parameter $s > 0$. Given any $v \in \ell_p^d$:

$$F_s(v) = \frac{s}{2(C_q)^{1/q}} \sin\left(\phi + \frac{2}{s}\sum_{i=1}^{m} v_i g_i\right),$$

where C_q is a normalization constant that we will choose later.

Note that $|F_s(v)| \leq s$, as we wanted. In addition, note that the Hoeffding's inequality depends on s, and this dependence provides different guarantees for different scales. This s parameter will be chosen according to the ranges of the original distances, i.e. $s(R, \epsilon, p, q)$. This 'explains' what role s plays.

It remains to do some computations, in order to see that the embedding we defined preserves the original distances in expectation. Given any $v \neq w \in S$:

$$|F_s(v) - F_s(w)|^q = \frac{s^q}{2^q(C_q)}\left|\sin\left(\phi + \frac{2}{s}\sum v_i g_i\right) - \sin\left(\phi + \frac{2}{s}\sum w_i g_i\right)\right|^q =$$
$$= \frac{s^q}{(C_q)}\left|\sin\left(\frac{1}{s}\sum (v_i - w_i)g_i\right) \cdot \cos\left(\phi + \frac{1}{s}\sum (v_i + w_i)g_i\right)\right|^q.$$

Therefore

$$E_{\phi,g_i}\left[|F_s(v) - F_s(w)|^q\right] = \frac{s^q}{(C_q)} E_{\phi,g_i}\left[\left|\sin\left(\frac{1}{s}\sum (v_i - w_i)g_i\right) \cdot \cos\left(\phi + \frac{1}{s}\sum (v_i + w_i)g_i\right)\right|^q\right] =$$
$$= \frac{s^q}{(C_q)} E_{\phi,g_i}\left[\left|\sin\left(\frac{1}{s}\sum (v_i - w_i)g_i \cdot \cos(\phi)\right)\right|^q\right] = \frac{s^q E_\phi\left[|\cos\phi|^q\right]}{(C_q)} E_{g_i}\left[\left|\sin\left(\frac{\|v - w\|_p}{s} g\right)\right|^q\right].$$

Define $C_q = E_\phi\left[|\cos\phi|^q\right]$, and note the dependence of the original distance and the new distance. For small scale distances $\sin x \sim x$, which gives us the required bounds for a particular range.

References

[1] Yair Bartal and Lee-Ad Gottlieb. Dimension reduction techniques for l_p (1<p<2), with applications. In *32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA*, pages 16:1–16:15, 2016.

[2] Rafail Ostrovsky and Yuval Rabani. Polynomial-time approximation schemes for geometric min-sum median clustering. *J. ACM*, 49(2):139–156, March 2002.