NNS lower bounds via metric expansion for l_∞ and EMD

Michael Kapralov[1⋆] and Rina Panigrahy[2]

[1] Stanford iCME, Stanford, CA, kapralov@stanford.edu
[2] MSR Silicon Valley, Mountain View, CA, rina@microsoft.com

Abstract. We give new lower bounds for randomized NNS data structures in the cell probe model based on *robust* metric expansion for two metric spaces: l_∞ and Earth Mover Distance (EMD) in high dimensions. In particular, our results imply stronger non-embedability for these metric spaces into l_1. The main components of our approach are a strengthening of the isoperimetric inequality for the distribution on l_∞ introduced by Andoni et al [FOCS'08] and a robust isoperimetric inequality for EMD on quotients of the boolean hypercube.

1 Introduction

In the Nearest Neighbor Problem we are given a data set of n points $x_1, ..., x_n$ lying in a metric space V. The goal is to preprocess the data set into a data structure such that when given a query point $y \in V$, it is possible to recover the data set point which is closest to y by querying the data structure at most t times. The goal is to keep both the querying time t and the data structure space m as small as possible. Nearest Neighbor Search is a fundamental problem in data structures with numerous applications to web algorithms, computational biology, information retrieval, machine learning, etc. As such it has been researched extensively.

Natural metric spaces include the spaces $\Re^d$ equipped with the ℓ_1 or ℓ_2 distance that have been extensively studied in terms of upper and lower bounds. But other metrics such as ℓ_∞, edit distance and earth mover distance may be more appropriate in some settings [3, 9]. Naturally, the time space tradeoff of known solutions crucially depend upon the underlying metric space. The known upper bounds exhibit the 'curse of dimensionality': for d dimensional spaces either the space or time complexity is exponential in d – thus encouraging research on approximate solutions. In the c-approximate nearest neighbor version, one returns a neighbor that is at most distance c times that to the nearest neighbor [10], [12], [9], [2] –for example there is an algorithm to obtain a c-approximate near neighbor in time $\tilde{O}(1)$ and space $n^{1+O(1/c)}$ using locality sensitive hashing in the l_1 metric; for the l_2 metric the space drops to $n^{1+O(1/c^2)}$ [2]. For the ℓ_∞ metric Indyk [9] shows how to compute a $O(\log_{1/\epsilon} \log d)$-approximate NNS using space $n^{\Omega(1/\epsilon)}$; most of these algorithms are *randomized*, while the algorithm of Indyk [9] is deterministic. Our lower bounds for ℓ_∞ show that the space/approximation tradeoff in [9] is essentially optimal even if randomization is allowed.

⋆ Research supported by NSF grant 0904325. Part of the work was done while the author was an intern at Microsoft Research Silicon Valley.

There is a substantial body of work on lower bounds covering various metric spaces and parameter settings many of which assume the algorithm to be deterministic. Most previous papers are concerned with the Hamming distance over the d-dimensional hypercube. The cases of exact or deterministic algorithms were handled in a series of papers[7], [6],[13], [5]. These lower bounds hold for any polynomial space. In contrast the known upper bounds are both approximate and randomized, and with polynomial space can retrieve the output with one query. Chakrabarti and Regev [8] allow for both randomization and approximation, with polynomial space and show a tight bound for the *nearest* neighbor problem. Patrascu and Thorup[18] showed lower bounds on the query time of near neighbor problems with a stronger space restriction (near linear space), although their bound holds for deterministic or exact algorithms. Traditionally, cell probe lower bounds for data structures have been shown using communication complexity arguments [15]. Patrascu and Thorup [18] use a direct sum theorem along with the richness technique to obtain lower bounds for deterministic algorithms. Andoni, Indyk and Patrascu [4] showed randomized lower bounds using communication complexity lower bounds for Lopsided Set Disjointness. In [16, 17], a more direct geometric argument was used to show lower bounds for randomized algorithms based on different variants of expansion of the underlying metric space.

The metric ℓ_∞ is considered in an intriguing paper by Andoni *et al.*[1] who prove a lower bound for deterministic algorithms. The paper uses the richness lemma though the crux of the proof is an interesting isoperimetric bound on ℓ_∞ for a carefully chosen measure. The lower bound they provide is tight for constant query time and matches the upper bound from [9]. In this work we obtain lower bounds for randomized algorithms for two new metric spaces: ℓ_∞ and Earth Movers Distance (EMD). For the ℓ_∞ metric we extend the tight lower bounds of [1] from deterministic to randomized algorithms by computing the notion of 'robust expansion' introduced in [17]. Our's is the first work that looks at the hardness of NNS in EMD metric. Inspired from the Fourier based techniques in the non-embeddability results from [11] we show hardness of NNS in the EMD metric over point sets in the d-dimensional hamming cube. We prove the following hardness guarantees for the case when cell size is $n^{o(1)}$, where n is the number of points in the database. For a given distribution of points and query, a randomized algorithm for (approximate) NNS is one that produces an (approximate) near neighbor with probability at least $2/3$.

Theorem 1. *1. For a $O(\log_{1/\epsilon} \log d)$ approximate NNS in ℓ_∞, any (randomized) t-probe data structure needs space at least $n^{\Omega(1/(\epsilon t))}$*

2. There is distribution of sets from the Hamming cube $\{0,1\}^d$ so that any (randomized) t-probe data structure for an α approximate NNS the EMD metric on this set needs space at least $e^{\Omega(d/(\alpha t))}$ (each set in the distribution can be specified explicitly using $O(d)$ bits).

It is interesting to note that approximate NNS for EMD under this distribution takes exponential space for approximation $O(d^{1-\epsilon})$ for all constant $\epsilon > 0$. Note that lower bounds on NNS on a metric space are stronger than non-embeddability results as once a metric space can be embedded into a well-studied metric space, the algorithms for NNS from the latter will carry over with the appropriate distortion. Thus our results automatically imply robust non-embeddibility results for these metric spaces. While it was known that these metric spaces do not embed into l_1 or l_2 with constant distortion, we now know that they are also not gap embeddable. In particular for the EMD metric on point sets from a d-dimensional hypercube Khot et al[11] showed that it doesn't embed into the ℓ_1 metric with distortion less than d . Our bound generalizes this to gap-inembeddibility:

Theorem 2. *There is no embedding M from the EMD metric space induced by the hamming metric on point sets over $\{0,1\}^d$ to the l_1 metric that satisfies the following gap distortion guarantees.*

$$EMD(u,v) \leq \omega(1) \implies |M(u) - M(v)|_1 \leq 1$$

$$EMD(u,v) = \Omega(d) \implies |M(u) - M(v)|_1 \geq 2$$

We will now review the different notions of metric-expansion from [17] that produce lower bounds for different classes of algorithms, deterministic and randomized. The bounds hold even for even in the average case when the points are chosen uniformly from a certain distribution.

1.1 Expansion and its relation to complexity of NNS

The results in [17] show a relation between the expansion of the metric space and the complexity of NNS. It works with the version of the *Near Neighbor* version of the problem that is parameterized by a search radius r. As in the Nearest Neighbor Search Problem given a query point y the goal is to determine whether the data set contains a point of distance at most r from y. Expansion can be used to consider the case when points are chosen randomly from a distrubution and the query point is a random point from a ball of radius r around one of the database points. Intuitively expansion is the amount by which a set of points expands when we include points in their r neighborhood. If distribution of points is such that the distance between any pair of database points is at least cr then this lower bound also implies hardness for c-approximate NNS.

To compute the expansion we construct an undirected bipartite graph $G = (U, V, E)$ where U and V are all the points in the metric space and and edge is placed between a pair of nodes from U and V if they are at most distance r apart. The data set comes by choosing n points randomly from U and query is a random neighbor from V of a random database point from U (these distrubutions may be non-uniform which we specify in detail later).

Definition 1 (Vertex expansion). *The $\delta-$vertex expansion of the graph is defined as*

$$\Phi_v(\delta) := \min_{A \subset V, |A| \leq \delta |U|} \frac{|N(A)|}{|A|}.$$

Here $N(A)$ denotes the neighborhood of the set A in G. For $A \subset V$, $B \subset U$, let $E(A,B)$ denote the set of edges between A and B in the bipartite graph G. Assume that $|A| = \delta|U|$. Observe that if $E(A,B) = E(A,U)$ then $|B| \geq \Phi_v(\delta)|A|$. In other words $\Phi_v(\delta)$ bounds the measure of the sets that cover all the edges incident on a set of measure δ. The notion of *robust expansion* relaxes this by requiring B to cover at least a γ-fraction of the edges incident on A. This idea is captured in the definition below. For simplicity we assume that $V = U$ and that G is regular. A more subtle definition which takes into account other non-regular graphs is presented later.

Definition 2 (Robust expansion). *G has robust-expansion $\Phi_r(\delta, \gamma)$ if $\forall A, B \subseteq V$ satisfying $|A| \leq \delta|V|, |B| \leq \Phi(\delta, \gamma)|A|$, it is the case that $\frac{|E(A,B|}{|E(A,V)|} \leq \gamma$. Note that $\Phi_r(\delta, 1) = \Phi_v(\delta)$.*

Lower bounds for NNS based on the above notions of expansion were proven in [17]; the deterministic lower bounds use expansion and the randomized lower bounds make use of robust-expansion We now state the bounds for randomized algorithms. For technical reasons, it also assumes that the metric space satisfies a property called *weak-independence* which simply means that two balls of radius r centered at randomly chosen points are sufficiently disjoint with high probability $1 - o(1/n^2)$. Here m denotes the number of cells used by the algorithm where each cell can hold a word of size w bits.

Theorem 3. *[17] There exists an absolute constant γ such that the following holds. Any randomized algorithm for a* weakly-independent *instance of Near Neighbor problem which is correct with probability at least half (where the probability is taken over the sampling of the input and the algorithm), satisfies the following inequalities:*

$$\frac{m^t w}{n} \geq \Phi_r\left(\frac{1}{m^t}, \frac{\gamma}{t}\right) \tag{1}$$

These theorems, combined with known isoperimetric inequalities yield most known cell probe lower bounds for near neighbor problems. There is also some evidence that the connection between expansion and hardness of NNS is tight for constant t – this has been shown to hold for cases when the graph G is symmetric [17].

The bipartite graph $G = (U, V, E)$ may be weighted by a a probability distribution e over the edges E. Let $\mu(u) = e(u, V) = \sum_{v \in V} e(u, v)$ be the induced distribution on U, and let $\nu(v) = e(U, v)$ be the induced distribution on V. For $x \in U$, we denote by ν_x the conditional distribution of the endpoints in V of edges incident on u, i.e. $\nu_x(y) = e(x, y)/e(x, V)$. Thus ν_y is a distribution over (or concentrated over) the r-neighborhood of y. In this case we select n points $x_1, \ldots, x_n$ independently from the distribution μ uniformly at random. This defines the database distribution. To generate the query, we pick an $i \in [n]$ uniformly at random, and sample y independently from ν_{x_i}. The tuple (G, e) satisfies γ-*weak independence (WI)* if $\Pr_{x,z \sim \mu, y \sim \nu_x}[(y, z) \in E] \leq \frac{\gamma}{n}$. Thus, weak independence ensures that with probability $(1 - \gamma)$, for the instance generated as above, x is indeed the unique neighbor in G of y in $\{x_1, \ldots, x_n\}$. The following definition generalizes the notion of robust-expansion to weighted bipartite graphs.

Definition 3. *[17] [Robust Expansion] The γ-robust expansion of a set $A \subseteq V$ is*

$$\phi_r(A, \gamma) \stackrel{def}{=} \min_{B \subseteq U : e(B,A) \geq \gamma e(U,A)} \mu(B)/\nu(A).$$

2 Robust expansion of l_∞

In this section we prove a bound on the robust expansion of l_∞ under a variant of the distribution introduced in [1]. Let $G_d = (U, V, E)$ be the l_∞ graph on $U = V = [1, \ldots, m]^d$, i.e. $u \in U$ is connected to $v \in V$ iff $||u - v||_\infty \leq 1$. We now define a distribution τ on the edges of G_d.

We start by defining the distribution for G_1, the one-dimensional l_1 graph (see Fig. 1). The distribution on G_d for general d will be the product of distributions on G_1. We let

$$\tau^1_{i,j} = 2^{-(1/\epsilon)^i} \text{ if } j = i+1 \text{ and } i \text{ is odd}$$
$$\tau^1_{i,j} = 2^{-(1/\epsilon)^j} \text{ if } j = i-1 \text{ and } i \text{ is odd}$$
$$\tau^1_{1,0} = 1 - \sum_{i\geq 1} 2^{-(1/\epsilon)^i}, \text{ and } \tau^1_{i,j} = 0 \text{ o.w.}$$

We denote the induced one-dimensional distributions by

$$\mu^1_u = \sum_{v\in N^1(u)} \tau^1_{(u,v)}, \nu^1_v = \sum_{u\in N^1(v)} \tau^1_{(u,v)}.$$

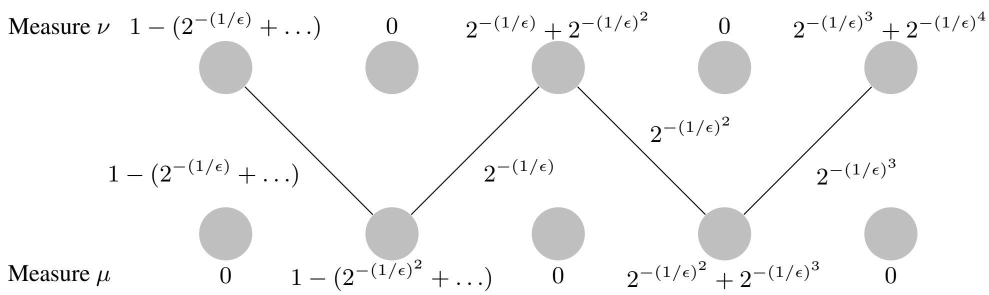

Fig. 1. Distribution on G_1

The d-dimensional distribution τ^d over edges is defined by $\tau^d_{(u,v)} = \prod_{i=1}^d \tau^1_{u_i,v_i}$. We fist note that this induces a product distribution on the vertices $u \in U$, where $\mu^d(u) = \prod_{i=1}^d \mu_1(u_i)$. In what follows we will use the notation $e_d(A,B) = \sum_{e\in E\cap(A\times B)} \tau^d_e$. We also omit superscripts in μ^d, ν^d, e_d and τ^d_e whenever this does not cause confusion.

The main component of our lower bound is a strengthened isoperimetric inequality for l_∞ under the distribution that we just defined. The main technical lemma will be

Lemma 1. *Let $G_d = (U, V, E)$ denote the l_∞ graph. For any $A \subseteq U, B \subseteq V$ one has $e(A,B) \leq (\mu(A)\nu(B))^{1/(1+\delta)}$ for some $\delta = \Theta(\epsilon)$ and all sufficiently small ϵ.*

A bound on robust expansion follows from Lemma 1 (details are deferred to the full version):

Lemma 2. *Let $G_d = (U, V, E)$ denote the l_∞ graph. For any $A \subseteq U, B \subseteq V$ such that $e(B, A) \geq \gamma e(A, V)$ one has $\nu(B) \geq \gamma^{1+\delta}(\mu(A))^\delta$ for some $\delta = \Theta(\epsilon)$ and sufficiently small ϵ.*

The proof Lemma 1 is by induction on the dimension, and we start by outlining the proof strategy for the base case, i.e. $d = 1$. For $d = 1$, Lemma 1 turns into

Lemma 3. *Let G_1 denote the l_∞ graph in dimension 1 with the measure τ defined as above. There exist constants $\gamma, \epsilon^* > 0$ such that for every $x, y \in \mathbb{R}_+^V$ for $\epsilon < \epsilon^*$ and $\delta = \gamma\epsilon$ one has*

$$\sum_{(i,j)\in E(G_1)} x_i \tau_{i,j} y_j \leq \left(\sum_i \mu_i x_i^{1+\delta}\right)^{1/(1+\delta)} \left(\sum_i \nu_i y_i^{1+\delta}\right)^{1/(1+\delta)} \tag{2}$$

It will be convenient to make a substitution to ensure that the rhs is the product of unweighted $(1+\delta)$-norms. Set $u_i = \mu_i^{1/(1+\delta)} x_i, v_i = \nu_i^{1/(1+\delta)} y_i$, so that (2) becomes

$$\sum_{(i,j)\in E} u_i \mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-1/(1+\delta)} v_j \leq ||u||_{1+\delta} ||v||_{1+\delta}. \tag{3}$$

We prove the bound (3) in two steps. In particular, we break the graph G_1 into two pieces that overlap by one vertex, prove stronger versions of (3) for both subproblems, and then piece them together to obtain (3).

In the first step we concentrate on the subgraph induced by vertices on both sides with indices in $[0 : 2]$. This amounts to only considering distributions that are zero outside of $[0 : 2]$. We prove in Lemma 4 that a strengthened version of (3) holds under these restrictions. In particular, we show in the full version that

Lemma 4. *There exist constants $\epsilon^*, \gamma > 0$ such that for all $v_0, v_2 \geq 0$ one has for all $\epsilon < \epsilon^*, \delta = \gamma\epsilon$*

$$\tau_{10} v_0 + \tau_{12} v_2 \leq \left(\nu_0 v_0^{1+\delta} + (1 - \Omega(\delta^5))\nu_1 v_1^{1+\delta}\right)^{1/(1+\delta)} \tag{4}$$

It should be noted that while (3) depends on both u and v, the inequality in (4) only depends on u. This is because only the single vertex v_1 has a nonzero weight among vertices in $[0 : 2]$, and hence can be cancelled from both sides. The $1 + O(\delta^5)$ term multiplying u_2 on the lhs represents the main strengthening, and will be crucially important for combining the inequalities for different parts of the graph later.

In the second step we consider the subgraph of G_1 induced by vertices with indices in $[2 : +\infty]$. This amounts to considering distributions that are zero on the the first two vertices on each side of the graph. For this case we prove

Lemma 5. *Let G_1 denote the l_∞ graph in dimension 1 with the measure τ defined as above. There exist constants $\gamma, \epsilon^* > 0$ such that for every $x, y \in \mathbb{R}_+^V$ for $\epsilon < \epsilon^*$ and $\delta = \gamma\epsilon$ one has*

$$\sum_{(i,j)\in E(G_1), i>1} x_i \tau_{i,j} y_j \leq 2^{-1/\epsilon} \left(\sum_i \mu_i x_i^{1+\delta}\right)^{1/(1+\delta)} \left(\sum_i \nu_i y_i^{1+\delta}\right)^{1/(1+\delta)} \tag{5}$$

The $2^{-1/\epsilon}$ term represents the strengthening with respect to (3) and will be crucial for combining (4) and (5). Combining (4) and (5), we then get the result (essentially) by an application of Cauchy-Schwarz and norm inequalities. One complication will be the fact that (4) and (5) overlap by v_2, but we will be able to handle this since the strengthened inequalities ensure that v_2 appears in (4) and (5) with weights that sum up to at most 1. We now give

Proof of Lemma 5: We need to bound $\sum_{(i,j)\in E, i\geq 2} u_i \mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-1/(1+\delta)} v_j$. In order to do that, we decompose the edges of G_1 restricted to $[2:+\infty]$ into two edge disjoint matchings M_1 and M_2: $M_1 = \{(i,j) \in E(G_1) : j = i-1, i,j \geq 2\}$, $M_2 = \{(i,j) \in E(G_1) : j = i+1, i,j \geq 2\}$.

First, suppose that $(i,j) \in M_1$, i.e. $j = i-1$ and$i = 2k+1$, where $k \geq 1$ since we are considering distributions restricted to $[2:+\infty]$. We have

$$\mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-1/(1+\delta)} \leq 2^{(1/\epsilon)^{(k+1)}/(1+\delta)} \cdot 2^{-(1/\epsilon)^{k+1}} \cdot 2^{(1/\epsilon)^k/(1+\delta)} = 2^{(1/\epsilon)^k (1-\delta/\epsilon)/(1+\delta)}.$$

For $\delta \geq 4\epsilon$ and sufficiently small constant ϵ $\mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-(1-2\epsilon)} \leq 2^{-2(1/\epsilon)^k} \leq 2^{-2/\epsilon}$, where we used the fact that $k \geq 1$. A similar argument shows that the same holds for all $(i,j) \in M_2$. Thus, for $r = 1, 2$

$$\sum_{(i,j)\in M_r} u_i \mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-1/(1+\delta)} v_j \leq 2^{-2/\epsilon} \sum_{(i,j)\in E, i\geq 2} u_i v_j \leq 2^{-2/\epsilon} \sqrt{\sum_{i\geq 2} u_i^2} \sqrt{\sum_{j\geq 2} v_j^2}$$

by Cauchy-Schwarz. Since for all x one has $||x||_p \geq ||x||_q$ when $p \leq q$, we conclude that for $r = 1, 2$

$$\sum_{(i,j)\in M_r} u_i \mu_i^{-1/(1+\delta)} \tau_{ij} \nu_j^{-1/(1+\delta)} v_j \leq 2^{-2/\epsilon} \left(\sum_{i\geq 2} u_i^{1+\delta}\right)^{1/(1+\delta)} \left(\sum_{j\geq 2} v_j^{1+\delta}\right)^{1/(1+\delta)},$$

as required. Putting the estimates for M_1 and M_2 together, we get

$$\sum_{(i,j)\in E(G_1), i\geq 2} x_i \tau_{i,j} y_j \leq 2^{-1/\epsilon} \left(\sum_i \mu_i x_i^{1+\delta}\right)^{1/(1+\delta)} \left(\sum_i \nu_i y_i^{1+\delta}\right)^{1/(1+\delta)}.$$

□

We now prove Lemma 3, and then use it as the base case for induction on dimension.

Proof of Lemma 3: By Lemma 4 we have

$$\sum_{(i,j)\in E(G_1), i,j\leq 2} x_i \tau_{i,j} y_j \leq \left(\mu_1 x_1^{1+\delta}\right)^{1/(1+\delta)} \left(\nu_0 y_0^{1+\delta} + (1-\Omega(\delta^5))\nu_2 y_2^{1+\delta}\right)^{1/(1+\delta)}, \tag{6}$$

For convenience, let $A := \left(\mu_1 x_1^{1+\delta}\right)^{1/(1+\delta)}$, $B := \left(\nu_1 y_1^{1+\delta} + (1-\Omega(\delta^5))\nu_2 y_2^{1+\delta}\right)^{1/(1+\delta)}$. Furthermore, by Lemma 5

$$\sum_{(i,j)\in E(G_1), i\geq 2, j\geq 2} x_i \tau_{i,j} y_j \leq 2^{-1/\epsilon} \left(\sum_i \mu_i x_i^{1+\delta}\right)^{1/(1+\delta)} \left(\sum_i \nu_i y_i^{1+\delta}\right)^{1/(1+\delta)}, \tag{7}$$

and we define for convenience $C := \left(\sum_i \mu_i x_i^{1+\delta}\right)^{1/(1+\delta)}$ and $D := 2^{-1/\epsilon}\left(\sum_i \nu_i y_i^{1+\delta}\right)^{1/(1+\delta)}$.

First, we get by combining (6) and (7) that

$$\sum_{(i,j)\in E(G_1)} x_i \tau_{i,j} y_j \le A \cdot B + C \cdot D \tag{8}$$

Applying Cauchy-Schwarz and norm inequalities to the rhs of (8), we get

$$A\cdot B + C\cdot D \le \sqrt{A^2+C^2}\sqrt{B^2+D^2} \le \left(A^{1+\delta}+C^{1+\delta}\right)^{1/(1+\delta)}\left(B^{1+\delta}+D^{1+\delta}\right)^{1/(1+\delta)}. \tag{9}$$

Combining (8) and (9), we obtain

$$\sum_{(i,j)\in E(G_1)} x_i \tau_{i,j} y_j \le \left(\nu_0 y_0^{1+\delta} + \nu_2(1-\Omega(\delta^5)+2^{-(1+\delta)/\epsilon})y_2^{1+\delta} + \sum_{j>2}\nu_j y_j^{1+\delta}\right)^{1/(1+\delta)}$$
$$\cdot \left(\mu_1 x_1^{1+\delta} + \sum_{i>1}\mu_i x_i^{1+\delta}\right)^{1/(1+\delta)} \le \left(\sum_{i\ge 0}\mu_i x_i^{1+\delta}\right)^{\frac{1}{1+\delta}} \left(\sum_{j\ge 0}\nu_j y_j^{1+\delta}\right)^{\frac{1}{1+\delta}}$$

□

Proof of Lemma 1: We use induction on d. The base case $d = 1$ is given by Lemma 3. We now describe the inductive step $d-1 \to d$.

Let $A \subseteq U, B \subseteq V$. For each i let $A_i = \{u \in A : u_i = i\}, B_i = \{u \in A : u_i = i\}$. Then by our definition of edge weights $e_d(A,B) = \sum_{(i,j)\in E(G_1)} \tau_{ij} e_{d-1}(A_i, B_j)$. By the inductive hypothesis we have $e_{d-1}(A_i,B_j) \le (\mu_{d-1}(A_i)\mu_{d-1}(B_j))^{1/(1+\delta)}$, and hence

$$e_d(A,B) \le \sum_{(i,j)\in E(G_1)} \tau_{ij}(\mu_{d-1}(A_i)\mu_{d-1}(B_j))^{1/(1+\delta)}.$$

Now by Lemma 3 we have

$$\sum_{(i,j)\in E(G_1)} \tau_{ij}(\mu_{d-1}(A_i)\mu_{d-1}(B_j))^{1/(1+\delta)} \le \left(\sum_i \mu_i^1 \mu_{d-1}(A_i)\sum_j \mu_j^1\mu_{d-1}(B_j)\right)^{1/(1+\delta)}$$
$$= (\mu_d(A)\mu_d(B))^{1/(1+\delta)}.$$

□

Theorem 4. *$O(\log_{1/\epsilon}\log d)$-approximate NNS for l_∞ requires space $n^{\Omega(1/(\epsilon t))}$ even with randomization.*

Proof. The proof follows by first showing that the distance between a pair of points drawn from our distribution is $\Omega(\log_{1/\epsilon}\log d)$ and applying Theorem 3 together with Lemma 2. The details are deferred to the full version.

3 Earth mover distance

In this section we derive lower bounds on the cell probe complexity of nearest neighbor search for Earth mover distance (also known, as transportation cost metric) over $\mathbb{F}_2^d$. Our approach is based on lower bounding the robust expansion of EMD over quotients of $\mathbb{F}_2^d$ with respect to the dual of a random linear code. Quotients of $\mathbb{F}_2^d$ with respect to random linear codes have been used in [11] to derive non-embeddability results for EMD over $\mathbb{F}_2^d$ into l_1. Here we extend these non-embeddability results to hardness of nearest neighbor search. As a by-product of our approach, we also prove that EMD over $\mathbb{F}_2^d$ is not gap-embeddable into l_1 with distortion less than $\Omega(d)$.

Let (X, d) be a metric space. The earth mover distance between two sets $A, B \subseteq X$, such that $|A| = |B|$ is defined by

$$EMD(A, B) = \min_{\pi: A \to B} \sum_{x \in A} d(x, \pi(x)), \tag{10}$$

where the minimum is taken over all bijective mappings π from A to B. For the purposes of our lower bounds, the metric space (X, d) will be the binary hypercube $(\mathbb{F}_2^d, ||\cdot||_1)$ with Hamming distance as the metric, and A, B will be subsets of $\mathbb{F}_2^d$ of a special form. In particular, A and B will be cosets of $\mathbb{F}_2^d$ with respect to the action of a carefully chosen group (in fact, a linear code with large minimum distance).

Let C denote a linear code, i.e. a linear subspace of $\mathbb{F}_2^d$ of dimension $\Omega(d)$ and minimum distance $\Omega(d)$. Such codes are known to exist [14]. In particular, it can be seen that a random linear code of dimension $\Omega(d)$ satisfies this conditions with high probability. We will use the notation for the dual code

$$C^\perp = \{y \in \mathbb{F}_2^d : (y, x) \equiv 0 \mod 2, \forall x \in C\},$$

where $(x, y) = \sum_{i=1}^d x_i y_i$. For a vector $u \in \mathbb{F}_2^d$ we denote the coset of u with respect to the dual code $C^\perp$ by $\mathbf{u} = \{w \in \mathbb{F}_2^d : w - u \in C^\perp\}$. Thus, $\mathbf{u}$ is the set of vectors in $\mathbb{F}_2^d$ that can be obtained from u by translating it by an element of $C^\perp$. In what follows we consider EMD on such subsets $\mathbf{u}$ of the hypercube. The following simple property of EMD restricted to cosets of $\mathbb{F}_2^d$ with respect to $C^\perp$ will be very useful. Recall that by (10) $EMD(\mathbf{u}, \mathbf{v})$ is the cost of the bijective mapping π from A to B that minimizes total movement $\sum_{x \in A} ||x - \pi(x)||_1$. We now show that when EMD is restricted to cosets of $C^\perp$, i.e. $A = \mathbf{u}, B = \mathbf{v}$ for some $u, v \in \mathbb{F}_2^d$, the minimum over mappings π is achieved for a mapping that simply translates each element of a coset $\mathbf{u}$ by a fixed vector w to get $\mathbf{v}$ (the proof is deferred to the full version.):

Fact 5 *For $\mathbf{u}, \mathbf{v} \in \mathbb{F}_2^d / C^\perp$ one has $EMD(\mathbf{u}, \mathbf{v}) = |C^\perp| \cdot \min_{a \in \mathbf{u}, b \in \mathbf{v}} ||a - b||_1$.*

Our estimates of robust expansion of EMD on $\mathbb{F}_2^d / C^\perp$ will use Fourier analysis on the hypercube, so we give the necessary definitions now. The Fourier basis is given by Walsh functions $W_A : \mathbb{F}_2^d \to \mathbb{R}, A \subseteq \{1, \ldots, d\}$ is denoted by

$$W_A(x) = (-1)^{\sum_{j \in A} x_j}, x = (x_1, \ldots, x_d) \in \mathbb{F}_2^d.$$

Thus, $\{W_A : A \subseteq \{1, \ldots, d\}\}$ is an orthonormal basis of $L_2(\mathbb{F}_2^d, \sigma)$, where $\sigma(x) = 2^{-d}, x \in \mathbb{F}_2^d$ is the uniform measure on $\mathbb{F}_2^d$. For each $f : \mathbb{F}_2^d \to \mathbb{R}$ one has $f =$

$\sum_{A\subseteq\{1,\ldots,d\}} \hat{f}(A)W_A$, where $\hat{f}(A) = \int_{\mathbb{F}_2^d} f(x)W_A(x)d\sigma(x)$. Parseval's indentity states that

$$\int_{\mathbb{F}_2^d} f(x)g(x)d\sigma(x) = \sum_{A\subseteq\{1,\ldots,d\}} \hat{f}(A)\hat{g}(A)$$

for all $f, g \in L_2(\mathbb{F}_2^d, \sigma)$. We will often use the notation $(f, g) = \int_{\mathbb{F}_2^d} f(x)g(x)d\sigma(x)$. We will also use the non-uniform measure $\sigma_\epsilon(x) = \epsilon^{\sum_{i=1}^d x_i}(1-\epsilon)^{d-\sum_{i=1}^d x_i}$.

We now define the distribution on inputs that we will use for our lower bounds. For $r \in (0, d)$ let $G = (U, V, E)$, where $U = V = \mathbb{F}_2^d/C^\perp$ denote the complete bipartite graph. We now define distributions on U, V and the edges of G. Let μ and ν denote the uniform distribution on U and V respectively. The distribution on pairs is given first sampling $\mathbf{u} \in U$ uniformly, and then letting

$$\mathbf{v} = \mathbf{u} + Z, \tag{11}$$

where $\mathbf{Pr}[Z = z] = \sigma_{r/d}(z)$, i.e. Z is a point in $\mathbb{F}_2^d$ obtained by setting each coordinate independently to 1 with probability r/d and 0 with probability $1 - r/d$. Here for a coset $\mathbf{u}$ and a point $z \in \mathbb{F}_2^d$ we write $\mathbf{u} + z$ to denote the coset obtained from $\mathbf{u}$ by adding z to each $u \in \mathbf{u}$. We note that this is equivalent to sampling a uniformly random $\mathbf{u}$, then sampling a uniformly random point $u \in \mathbf{u}$, letting $v = u + Z$ and declaring $\mathbf{v}$ to be the resulting coset. In particular, this yield the following distribution on edges;

$$\tau_{\mathbf{u},\mathbf{v}} = \frac{1}{2^d} \sum_{u\in\mathbf{u}, v\in\tau} \sigma_{r/d}(u - v). \tag{12}$$

The distance between $\mathbf{u}$ and $\mathbf{v}$ sampled according to this distribution is $O(r)$ with high probability: $\mathbf{Pr}_{(\mathbf{u},\mathbf{v})\in E}[EMD(\mathbf{u}, \mathbf{v}) > \gamma r] \leq e^{-\Omega((\gamma-1)r)}$, i.e. pairs sampled from our distribution are nearby with high probability. On the other hand, two uniformly random cosets are at distance $\Omega(d)$ with high probability:

Lemma 6. *Let $\mathbf{u}, \mathbf{v}$ denote uniformly random points in $\mathbb{F}_2^d/C^\perp$. Then $\mathbf{Pr}[EMD(\mathbf{u}, \mathbf{v}) > c'd] \geq 1 - 2^{-\Omega(d)}$ for a constant $c' > 0$.*

We now turn to lower bounding the robust expansion. It will be convenient to use the following notation. For $A \in \mathbb{F}_2^d/C^\perp$ we will write $\mathbf{1}_A$ to denote the indicator function of A lifted to $\mathbb{F}_2^d$, i.e. $\mathbf{1}_A(x)$ equals 1 if $x \mod C^\perp = A$ and 0 otherwise. Our main lemma relies on the following crucial property of functions that are constant on cosets of $C^\perp$, proved in [11]. In particular, any such function necessarily has zero Fourier coefficients corresponding to non-empty sets of small size:

Lemma 7. *[11] Assume that $f : \mathbb{F}_2^d \to \mathbb{R}$ satisfies for every $x \in \mathbb{F}_2^d$ and for all $y \in C^\perp$, $f(x + y) = f(x)$. Suppose that the minimum distance of C is d_0. Then $\hat{f}(S) = 0$ for all $|S| < d_0, S \neq \emptyset$.*

The function $\mathbf{1}_A(x)$ satisfies the preconditions of Lemma 7 for $A \in \mathbb{F}_2^d/C^\perp$, and hence we have $\hat{\mathbf{1}}_A(S) = 0$ for $|S| \leq c'd$, $S \neq \emptyset$.

We now bound the robust expansion of EMD under our distribution. Similarly to section 2, we first bound the weight of edges going between a pair of sets A, B. As before, we use the notation $e(A,B)=\sum_{\mathbf{u}\in A,\mathbf{v}\in B}\tau_{\mathbf{u},\mathbf{v}}$. It will be convenient to express $e(A,B)$ in terms of the Bonami-Beckner operator $T_\rho : L_2(\mathbb{F}_2^d,\sigma)\to L_2(\mathbb{F}_2^d,\sigma)$. For a function $f\in L_2(\mathbb{F}_2^d,\sigma)$ one has $T_\rho f(x)=\mathbf{E}_{z\sim\sigma_{1-2\rho}}[f(x+z)]$, where we will use $\rho=1-2r/d$. The proof of the following claim is given in the full version:

Claim 6 *For any $A,B\in\mathbb{F}_2^d/C^\perp$ one has $e(A,B)=(T_\rho\mathbf{1}_A,\mathbf{1}_B)$, where $(f,g)=\int_{\mathbb{F}_2^d} f(x)g(x)\sigma(x)$.*

Our main lemma, which bounds the weight of edges going between a pair $A,B\in V$ is

Lemma 8. *Let C be a linear code of dimension $\Omega(d)$ and minimum distance $\Omega(d)$. Let $\mathbb{F}_2^d/C^\perp$ denote the quotient of $\mathbb{F}_2^d$ with respect to the dual code $C^\perp$, and consider the distribution over edges given by the noise operator with parameter $\rho=1-2r/d$ as in (11). Then for any $r<d/4$ one has $e(A,B)\le\mu(A)\mu(B)+e^{-\Omega(r)}\sqrt{\mu(A)\mu(B)}$.*

Proof. Consider any two sets $A,B\subseteq\mathbb{F}_2^d/C^\perp$. By Claim 6, we have $e(A,B)=(T_\rho\mathbf{1}_A,\mathbf{1}_B)$. We now use the fact that $\mathbf{1}_A$ is constant on quotients of $C^\perp$, and hence by Lemma 7 one has $\hat{\mathbf{1}}_A(S)=0$ for all $S\subseteq\{0,1\}^n$, $S\neq\emptyset$, with $|S|\le cd$. Since

$$T_\rho\mathbf{1}_A=\sum_{S\subseteq\{0,1\}^d}(1-2\rho)^{|S|}\widehat{\mathbf{1}}_A(S)W_S,\tag{13}$$

we have $||T_\rho f||\le e^{-cr}||f||$ for all $f\in L_2(\mathbb{F}_2^d,\sigma)$, such that $(f,\mathbf{1}_1)=0$. Here we denote the constant function equal to 1 by $\mathbf{1}$. We also use the fact that if $(f,\mathbf{1})=0$, then $(T_\rho f,\mathbf{1})=0$, as can be seen directly from (13). For $A\subset\mathbb{F}_2^d/C^\perp$ we will write $|\mathbf{1}_A|$ to denote l_1-norm of $\mathbf{1}_A$ (in particular, $|\mathbf{1}_A|=|C^\perp|\cdot|A|$), where $|A|$ is the number of elements in A. We now have

$$\begin{aligned}(T_\rho\mathbf{1}_A,\mathbf{1}_B)&=\left(\frac{|\mathbf{1}_A|}{2^d}\mathbf{1}+T_\rho(\mathbf{1}_A-\frac{|\mathbf{1}_A|}{2^d}\mathbf{1}),\frac{|\mathbf{1}_B|}{2^d}\mathbf{1}+(\mathbf{1}_B-\frac{|\mathbf{1}_B|}{2^d}\mathbf{1})\right)\\&=\left(\frac{|\mathbf{1}_A|}{2^d}\mathbf{1},\frac{|\mathbf{1}_B|}{2^d}\mathbf{1}\right)+\left(T_\rho(\mathbf{1}_A-\frac{|\mathbf{1}_A|}{2^d}\mathbf{1}),\mathbf{1}_B-\frac{|\mathbf{1}_B|}{2^d}\mathbf{1}\right)\end{aligned}$$

since the cross terms cancel due to orthogonality. Thus,

$$(T_\rho\mathbf{1}_A,\mathbf{1}_B)\le 2^{-2d}|\mathbf{1}_A||\mathbf{1}_B|+e^{-2\rho cd}||\mathbf{1}_A-\frac{|\mathbf{1}_A|}{2^d}\mathbf{1}||||\mathbf{1}_B-\frac{|\mathbf{1}_B|}{2^d}\mathbf{1}||,$$

and since $\rho d=r$, we get

$$e(A,B)\le\frac{|\mathbf{1}_A|}{2^d}\cdot\frac{|\mathbf{1}_B|}{2^d}+e^{-2\rho cd}\sqrt{\frac{|\mathbf{1}_A|}{2^d}\cdot\frac{|\mathbf{1}_B|}{2^d}}\le\mu(A)\mu(B)+e^{-\Omega(r)}\sqrt{\mu(A)\mu(B)}.$$

Using Lemma 8 we can now bound the robust expansion of EMD over $\mathbb{F}_2^d/C^\perp$:

Lemma 9. *Let C be a linear code of dimension $d/4$ such that the distance of $C^\perp$ is at least $c'd$ for some constant $c>0$. Then the γ-robust expansion of EMD over $\mathbb{F}_2^d/C^\perp$ at distance r is at least $(\gamma/2)^2e^{\Omega(r)}$.*

Theorem 7. *α-approximate NNS with t probes for d-dimensional EMD requires $e^{\Omega(d/(\alpha t))}$ space, even with randomization.*

Proof. Set $r = \Theta(d/\alpha)$. By Lemma 6 the distance between points is $\Omega(d)$ whenever $d \geq c \log n$ for a sufficiently large $c > 0$, which gives the weak independence property. The distance to the near point is $\Theta(r)$ with probability $1 - n^{-\Omega(1)}$. The robust expansion is at least $(\gamma/2)^2 e^{\Omega(r)}$ by Lemma 9, so the result follows by Theorem 3.

Proof of Theorem 2: Suppose that such an embedding exists. Then one can build a NNS data structure of size $n^{O(1)}$ to solve $3/2$-approximate NNS in l_1, implying a $o(d)$-approximate NNS for EMD. However, this would contradict Theorem 7 when $d = \Omega(\log n)$. □

References

1. Alexandr Andoni, Dorian Croitoru, and Mihai Patrascu. Hardness of nearest neighbor under l-infinity. *FOCS*, 2008.
2. Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Commun. ACM*, 51(1):117–122, 2008.
3. Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Earth mover distance over high-dimensional spaces. *SODA*, pages 343–352, 2008.
4. Alexandr Andoni, Piotr Indyk, and Mihai Patrascu. On the optimality of the dimensionality reduction method. *FOCS*, pages 449–458, 2006.
5. Omer Barkol and Yuval Rabani. Tighter bounds for nearest neighbor search and related problems in the cell probe model. *STOC*, pages 388–396, 2000.
6. Allan Borodin, Rafail Ostrovsky, and Yuval Rabani. Lower bounds for high dimensional nearest neighbor search and related problems. *STOC*, pages 312–321, 1999.
7. Amit Chakrabarti, Bernard Chazelle, Benjamin Gum, and Alexey Lvov. A lower bound on the complexity of approximate nearest-neighbor searching on the hamming cube. *STOC*, pages 305–311, 1999.
8. Amit Chakrabarti and Oded Regev. An optimal randomised cell probe lower bound for approximate nearest neighbour searching. *FOCS*, pages 473–482, 2004.
9. Piotr Indyk. On approximate nearest neighbors under l_∞ norm. *J. Comput. Syst. Sci*, 63, 2001.
10. Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. *STOC*, page 604613, 1998.
11. Subhash Khot and Assaf Naor. Nonembeddability theorems via fourier analysis. *FOCS'05*.
12. E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *STOC*, page 614623, 1998.
13. Ding Liu. A strong lower bound for approximate nearest neighbor searching. *Inf. Process. Lett.*, 92(1):23–29, 2004.
14. F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland: New York, NY, 1977.
15. Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.
16. Rina Panigrahy, Kunal Talwar, and Udi Wieder. A geometric approach to lower bounds for approximate near-neighbor search and partial match. *FOCS*, pages 414–423, 2008.
17. Rina Panigrahy, Kunal Talwar, and Udi Wieder. Lower bounds on near neighbor search via metric expansion. *FOCS*, 2010.
18. Mihai Patrascu and Mikkel Thorup. Higher lower bounds for near-neighbor and further rich problems. *FOCS*, pages 646–654, 2006.