Approximate Nearest Neighbors Beyond Space Partitions*

Alexandr Andoni† Aleksandar Nikolov‡ Ilya Razenshteyn§ Erik Waingarten¶

Abstract

We show improved data structures for the high-dimensional approximate nearest neighbor search problem (ANN) for ℓ_p distances for "large" values of p and for generalized Hamming distances. The previous best data structures proceeded by embedding a metric of interest into the ℓ_∞ space or an ℓ_∞-direct sum with simple summands, and then using data structures of Indyk (FOCS 1998, SoCG 2002) for ℓ_∞-ANN. In contrast to this, we bypass the embedding step and proceed by extending the technique underlying the ℓ_∞ data structures to handle ℓ_p and generalized Hamming distances directly. The resulting data structures are randomized, in contrast to Indyk's result for ℓ_∞-ANN, and replicate input points, in contrast with Locality Sensitive Hashing. This leads to ANN data structures with significantly improved approximations over those implied by embeddings, as well as those obtained using all known approaches based on random space partitions. .

1 Introduction

The *c-approximate near neighbor problem* (ANN, from now on) is defined as follows. Given a dataset P of n points lying in a metric space $\mathcal{M} = (X, d_X)$ and a parameter $r > 0$, build a data structure that, given a query point $q \in X$ within distance at most r from the dataset P, returns any data point within distance cr from the query q. The ANN problem has a wide range of applications (see [13, 3] for an overview), and at the same time, gives rise to a vast array of theoretical literature (see surveys [6, 4] as well as theses [1, 27]). In this paper, we focus on the *high-dimensional* case of the ANN problem, where we allow the parameters of a data structure to depend on the "dimension"[1] of the metric space $\mathcal{M}$ only *polynomially.*

In 1998, two seminal works on high-dimensional ANN were published:

1. Indyk and Motwani [21, 16] formulated a general framework based on random space partitions of the ambient metric space (termed Locality-Sensitive Hashing or LSH), and used it to obtain ANN data structures for ℓ_1 and ℓ_2 distances;

2. Indyk [17, 18] showed a deterministic data structure for the ℓ_∞ distance based on a certain geometric decomposition procedure which is notably *not* a space partition. Later in 2002, Indyk further extended this result [19] to handle ℓ_∞-direct sums, where the summands admit efficient ANN data structures.

The paper of Indyk and Motwani [21] pioneered the use of randomized space partitions for ANN, and has since become the foundation of an immense body of work. Since then, the quality of (data-obliviouis and data-dependent) random space partitions for ℓ_1 and ℓ_2 has improved, various extensions of [21] have been explored, and a number of impossibility results have been proved (for an overview, we refer to the recent survey [6]).

The key insight of [21] is to design a distribution over space partitions of ℓ_1/ℓ_2, such that the following event occurs with non-negligible probability: the (unknown) query point q and the near neighbor $p \in P$, which are within distance r, lie in the same part of the partition, while only a negligible fraction of points in P which are farther than cr from q fall in the same part as q.[2] During preprocessing, the data structure samples a random partition from the distribution and places dataset points into their corresponding parts; on a query, the data structure scans through the part of the partition where q lies, and outputs an approximate near neighbor if one is found (which is guaranteed to

*Some of the proofs are deferred to the full version.

†Columbia University, andoni@cs.columbia.edu. Supported in part by NSF grants CCF-1617955, CCF1740833, and Simons Foundation (#491119).

‡University of Toronto, anikolov@cs.toronto.edu

§Microsoft Research, ilyaraz@microsoft.com

¶Stanford University and the Simons Institute, eaw@cs.columbia.edu. This material is based upon work supported by the National Science Foundation under Award No. 2002201.

[1] Depending on the context, dimension can be defined differently, but generally captures the description size of the query. If X is finite, the dimension is commonly defined as $\log |X|$. If $\mathcal{M}$ is a normed space defined on $\mathbb{R}^d$, then the dimension is d. If $\mathcal{M}$ is a Riemannian manifold, then the dimension of $\mathcal{M}$ is a part of the definition etc.

[2] In particular, [21] set parameters so that at most $O(1)$ points from P which are farther than cr from q fall in the same part as q

occur if the event holds). We note that the polynomial space and time overhead comes from repeating the data structure in order to increase the success probability.

For metric spaces beyond ℓ_1 and ℓ_2, this approach of ANN via randomized space partitions has been vastly generalized to yield state-of-the-art ANN data structures for general d-dimensional normed spaces. The generalization follows from building a distribution over space partitions from an upper bound on the metric spectral gap [25, 26]; however, even for ℓ_∞, the approach fails to give an approximation better than $O(\log d)$. In constrast, the data structure of [17] achieves approximation $O(\log \log d)$ for ℓ_∞, thus, the landscape of ANN data structures for general metric spaces (even for general normed spaces) is far from being understood completely.

On a technical level, the data structure for ℓ_∞ of [18] is fundamentally different. Most notably, the polynomial space overhead is inherent throughout the entire execution since points are constantly being *replicated*[3] and the data structure is fully *deterministic* (see Section 1.2.1 for a more detailed discussion). Because of the universality of ℓ_∞ and product spaces, the data structures of Indyk [17, 18, 19] has enabled a number of new ANN results: for Hausdorff [14], Frechet [19], edit [20], Ulam distances [5], ℓ_p norms [1], and general symmetric norms [9]. However, the main contributions of these papers are in their use of metric embeddings as reductions to the ℓ_∞ case; the insights of [18, 19] are applied in a *black-box* fashion. This state of affairs is somewhat unsatisfactory both from the conceptual level, as well as from the quantitative point of view. In particular, one may hope to bypass the embedding step altogether and obtain better data structures by using the (extension of) techniques from [18, 19] for geometries other than ℓ_∞.

In this paper, we make the first step towards that goal. We develop techniques for designing ANN data structures extending [18, 19], going beyond the approach of space partitions. We obtain improved ANN data structures for two metric spaces of interest: for ℓ_p distances for "large" values of p and for generalized Hamming distances, defined as the ℓ_1-direct sum of several copies of a "small" metric space. The new data structures proceed by far-reaching generalizations of the Indyk's approach.

1.1 Results

We obtain improved ANN data structures for two classes of metrics.

[3] In this precise sense, the data structure for ℓ_∞ is not based on space partitions.

ℓ_p distances. Below is the main result that we show for ANN over ℓ_p norms.

THEOREM 1.1. *Fix $d \in \mathbb{N}$, $\alpha \in (0, 1/2]$. For any $n \in \mathbb{N}$ there exists a ANN data structure over ℓ_p^d for n-point datasets with space $O(dn^{1+\alpha})$ and query time $O(dn^\alpha)$, achieving approximation:*

$$c = O_\alpha\left(\log p \cdot \log^{2/p} d\right).$$

To put our result in context, we compare to the best known algorithms for ℓ_p. There are two incomparable results from the prior work. The first one uses the approach of randomized space partitions from [8, 7] and yields $O(p)$ approximation; we note that this bound is best possible via that approach[4]. The second one uses a randomized embedding of the ℓ_p^d norm into the ℓ_∞^d space using exponential random variables [1, 10] and then uses the ANN data structure for ℓ_∞ [18], and eventually yields a $O(\log \log d)$-approximation. Hence the former approach is better if $p \ll \log \log d$, while the reduction to ℓ_∞ is superior whenever $p \gg \log \log d$. Our Theorem 1.1 improves the state-of-the art whenever $\log p \cdot \log^{2/p} d \lesssim p$ (in particular, when $p \gg \log \log d / (\log \log \log d)^{1-\alpha}$, for arbitrary α). For instance, if $p \approx \log \log d$, then both of the previous results give approximation around $\log \log d$, while the approximation obtained from Theorem 1.1 is $O(\log \log \log d)$. We note that our data structure has exponential preprocessing time (also the case for the $O(p)$ approximation from [8, 7])

Generalized Hamming distances. The Generalized Hamming (GH) distance is defined as the ℓ_1-direct sum of d copies of a finite metric space Z (which we think of being small). More precisely, the space consists of tuples $(u_1, u_2, \ldots, u_d)$, where $u_i \in Z$. The distance between two tuples is defined as follows:

$$d((u_1, \ldots, u_d), (v_1, \ldots, v_d)) = \sum_{i=1}^{d} d_Z(u_i, v_i).$$

The GH distance is a generalization of the usual Hamming distance, where Z is the uniform metric over a finite set, and has many natural applications where Hamming distance over some finite alphabet does not capture well that some characters are more "similar" than others (e.g., nucleotides or amino-acids [15]).

We prove the following theorem for ANN under the Generalized Hamming distance.

THEOREM 1.2. *Fix $d \in \mathbb{N}$, a finite metric space (Z, d_Z) whose distances are between 1 and $R \in \mathbb{R}_{\geq 0}$, as well*

[4] It follows from the result of Matousek [24].

as the parameter $\alpha \in [0; 1/2)$. For any $n \in \mathbb{N}$, there exists a data structure for ANN over $\ell_1^d(Z)$ for n-point datasets using space $\mathrm{poly}(d, R, |Z|) \cdot n^{1+\alpha}$ *and query time* $\mathrm{poly}(d, R) \cdot |Z|^{\mathrm{poly}(R, \log |Z|, 1/\alpha)} \cdot n^{\alpha}$, *achieving approximation:*

$$c = O_\alpha (\log \log |Z| + \log R). \tag{1.1}$$

While there are no specialized algorithms for the ANN under GH distance, the existing techniques yield two different approaches, via random partitions and via ℓ_∞ embeddings. First, one can obtain $O(\log |Z|)$ approximation by embedding $\ell_1^d(Z)$ into ℓ_1 using the Bourgain's embedding [12] and then using any ANN under ℓ_1 [6] (based on random space partitions). The second approach obtains $O(\log \log n)$ approximation by embedding $\ell_1(Z)$ into $\ell_\infty(Z)$ using exponential random variables [1, 10] and then using ANN under ℓ_∞-direct sums of simple metrics [19]. Our Theorem above obtains approximation $O(\log \log |Z|)$, for the case when aspect ratio R is bounded by $\mathrm{poly}(\log |Z|)$[5], which yields exponential improvement over what can be obtained using space partitions. Crucially, our approximation is independent of n and d, unlike the reduction to $\ell_\infty(Z)$, and is thus better for the case of a fairly small metric Z.

Let us note as a side remark, that it is likely that the approximation $O(\log \log d)$ is tight for ℓ_∞-ANN [2, 22]. Thus, it is unlikely one can obtain our results merely by improving the ℓ_∞ data structures.

1.2 Techniques

1.2.1 List decompositions The main technical ingredient in the work of Indyk [18] is a decomposition lemma that shows that any n-point dataset $P \subset \mathbb{R}^d$ without a dense cluster of points can be decomposed in a way that allows building a decision tree for the ANN problem with low space and time complexity. Specifically, if no subset of P of diameter $\Delta = O(\log \log d)$ contains more than $\frac{n}{2}$ points, then P can be covered with disjoint sets $L, M_1, M_2, R \subset \mathbb{R}^d$ such that

- the sets are basis-aligned slabs, i.e., for some coordinate $k \in [d]$, and some value τ, $L = \{x : x_k \le \tau\}$, $M_1 = \{x : \tau < x_k \le \tau + 1\}$, $M_2 = \{x : \tau + 1 < x_k \le \tau + 2\}$, $R = \{x : x_k > \tau + 2\}$;
- L and R each contain at least $1/(4d)$ fraction of the points in P;
- the middle slabs are not too large; in particular
$$|P \cap (L \cup M_1 \cup M_2)|^2 + |P \cap (M_1 \cup M_2 \cup R)|^2 \le n^2.$$

Then, a decision tree data structure recursively builds two decision tree data structures for the (overlapping) sets $P_1 = P \cap (L \cup M_1 \cup M_2)$ and $P_2 = P \cap (M_1 \cup M_2 \cup R)$. Each set contains at most $(1 - 1/(4d))\, n$ points, which implies a bound of $O(d \log n)$ on the depth of the tree. The upper bound $|P_1|^2 + |P_2|^2 \le n^2$ implies that the total space is $O(n^2)$. For a query point q, we query the P_1 data structure if $q \in L \cup M_1$, and we query P_2 otherwise. The definition of the sets guarantees that, during this query procedure, q is never separated from any $x \in P$ such that $\|x - q\|_\infty \le 1$ (we consider $r = 1$). If, on the other hand, P contains a dense cluster, then we can store a single point from it, and recurse on points outside of it. Let us note that the sets P_1 and P_2 *do not* form a partition of P, but, rather, they overlap, and this is absolutely crucial for the overall result.

Our results build on this basic tool, and generalize it in several directions. We consider decompositions defined by a list $\mathcal{L}$ of tuples $(L_j, M_{j,1}, M_{j,2}, R_j)_{j=1}^{\ell}$, which induce overlapping sets $P_j = P \cap (L_j \cup M_{j,1} \cup M_{j,2})$ (for $j \le \ell$), and $P_{\ell+1} = P \cap (M_{\ell,1} \cup M_{\ell,2} \cup R_\ell)$. The tuples satisfy similar properties as above: $|P_j|/|P|$ is bounded away from 1, $\sum_{j=1}^{\ell+1} |P_j|^2 \le n^2$, and a description of each P_j can be efficiently stored. This structure defines a natural decision tree, analogous to the construction above. Working with lists allows a finer control of the sizes of the P_j. Additionally, we allow these list decompositions to be randomized, and we allow the query point q to be separated from a near neighbor x with low probability. The randomization is crucial for our results for generalized Hamming distances and ℓ_p norms.

1.2.2 Generalized Hamming distances The data structure for generalized Hamming distances combines list decompositions with a coordinate sampling strategy, thus drawing on techniques previously used for the ℓ_∞^d and ℓ_1^d norms. Recall that a generalized Hamming distance is defined as $\ell_1^d(Z)$, i.e., the ℓ_1-product of d copies of a finite metric Z. We assume that distances between distinct points in Z lie in the interval $[1, R]$. As an initial building block, consider the special case of ANN where near neighbors lie at distance $r = \Theta(dR/c)$ and $c = O(\log \log |Z|)$ is the approximation factor. To build a randomized list decomposition of an n-point set $P \subseteq Z^d$, we first sample a multiset S of m coordinates in $[d]$ and project the points in P to the coordinates in S, treating this projected space as $\ell_1^m(Z)$. A standard application of Hoeffding's inequality shows that, as long as m is a large enough constant relative to c,

[5]Let us note that the hard example for the Bourgain's theorem is the shortest-path metric of a constant-degree expander, which has logarithmic aspect ratio R.

any pair of points in $\ell_1^d(Z)$ have distances which are approximately preserved with constant probability by the projection. In particular, the distance between any query point q and a near neighbor $x \in P$ of q is preserved with constant probability. Moreover, if P does not contain a dense cluster of radius cr, then, with constant probability, the projection of P does not contain dense clusters of slightly smaller radius. Assuming that P does not contain a dense cluster, we may adapt Indyk's decomposition for ℓ_∞^d to obtain a list decomposition of the projected $P \subset \ell_1^m(Z)$ into sets of size $15n/16$. The sets are defined as pre-images of shells in $\ell_1^m(Z)$, and, since this is a metric space of size polynomial in $|Z|$, membership can be decided efficiently. The list decomposition separates a query point q and its near neighbor x only with small constant probability, where all the randomness comes from the sampling of coordinates. This list decomposition allows us to build a randomized decision tree: if there is a dense cluster in P, then we remember a single point in it, and recursively build a decision tree for the points outside the cluster; if there is no dense cluster, we recursively build decision trees for all the sets defined by the list decomposition.

The construction above handles the case when far points at distance cr are within a constant factor of the maximal possible distance dR in $\ell_1^d(Z)$. We reduce the general case to this maximal distance case by constructing a randomized embedding of $\ell_1^d(Z)$ into $\ell_1^{d_1}(Y)$, where Y is a metric space with distances in $[1, R]$, and d_1 and $|Y|$ are slightly larger than d and $|Z|$, respectively. The embedding has constant distortion, and maps pairs of points at distance cr in $\ell_1^d(Z)$ to points at distance $\Theta(d_1/R)$. This allows us to use the data structure for the maximal distance case, with the number of coordinates m sampled depending also on R.

1.2.3 ℓ_p distances While the data structure for ℓ_p^d also relies on randomized list decomposition, the randomization is not due to coordinate sampling. Similarly to [8], we consider a distributional version of the problem, where a pair $(\boldsymbol{x}, \boldsymbol{q})$ of a query point $\boldsymbol{q} \in \mathbb{R}^d$ and a dataset point $\boldsymbol{x}$ that is a near neighbor of $\boldsymbol{q}$ (i.e., at distance at most 1 from it) is sampled from a distribution μ. At a high level, solving this distributional problem for any distribution μ will imply the randomized data structure for the usual worst-case setting, via von Neumann's minimax theorem (see the proof of Theorem 3.1 for an illustration of this). We note, however, that his black-box use of the minimax theorem gives data structures that require exponential time pre-processing of the dataset, similarly to the data structures in [8].

To solve the distributional problem, it is sufficient to construct a (deterministic) list decomposition $\mathcal{L}$ for any distribution μ with left marginal ρ (i.e., the distribution on dataset points) such that any set P_j induced by $\mathcal{L}$ has small measure with respect to ρ, the probability that the pair $(\boldsymbol{x}, \boldsymbol{q}) \sim \mu$ is separated is an arbitrarily small constant, and the sets P_j do not overlap too much with respect to ρ, i.e., $\sum_{j=1}^{\ell+1} \rho(P_j)^2 \le 1$. In the case of ℓ_∞^d, Indyk's decomposition guarantees such a decomposition as long as there is no ball of radius $\Delta = O(\log\log d)$ whose measure is bigger than $1/2$ under ρ. Our goal is to, instead, get a decomposition assuming that there is no ball of radius $\Delta = O((\log p)(\log d)^{1/p})$ and measure $1/2$ under ρ. This, however, appears to be insufficient with our methods, and, instead, we prove the existence of a good list decomposition assuming that the marginal of ρ on any σd coordinates, for $\sigma = \Omega(1)$, does not contain a $\ell_p^{\sigma d}$ ball of radius Δ and measure $1/2$. Despite this qualitatively stronger assumption, we can still use the list decomposition to build an efficient data structure. In particular, when a list decomposition exists, we recursively build a data structure on the sets defined by it. When one does not exist, then the assumption is violated, and some σd-coordinate subspace contains a dense cluster; we can then recurse on both the points outside the cluster, and on the points inside the cluster, projected to the remaining $(1-\sigma)d$ coordinates. In both recursive calls, we are making progress, either by decreasing the dataset size, or decreasing the dimension of the problem. This allows us to still bound the approximation factor, and the space and time complexity of the data structure.

2 List Decompositions

This section sets up some of the main definitions we will need in order to define and analyze the ANN data structures. For now, we use X to denote an arbitrary finite subset of points. It is useful to think of X as a discretization of the universe from which the input pointset will be drawn. In the following definitions it will also be useful to think of the measure ρ as the empirical measure of an n-point set P, i.e., the measure given by $\rho(x) = \frac{1}{n}$ for every $x \in P$.

DEFINITION 2.1. (LIST DECOMPOSITIONS) *Let ρ be a probability distribution supported on X. For $\ell \in \mathbb{N}$, a* list decomposition $\mathcal{L}$ of length ℓ with respect to ρ *is given by a list of quadruples of disjoint subsets of X, $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i\in[\ell]}$. We have $L_1 \cup M_{1,1} \cup M_{1,2} \cup R_1 = X$, and, moreover, for every $i \in [\ell-1]$, we have*

$$\rho(L_i) \le \rho(R_i) \qquad \textit{and}$$

$$L_{i+1} \cup M_{i+1,1} \cup M_{i+1,2} \cup R_{i+1} = M_{i,1} \cup M_{i,2} \cup R_i,$$

We denote the query map *of $\mathcal{L}$ as $\mathsf{Q}_{\mathcal{L}} \colon X \to [\ell+1]$*

defined by:

$$\mathsf{Q}_{\mathcal{L}}(q) = \begin{cases} \min\{i \in [\ell] : q \in L_i \cup M_{i,1}\} & \exists i \in [\ell] : q \in L_i \cup M_{i,1} \\ \ell + 1 & q \in M_{\ell,2} \cup R_\ell \end{cases}$$

A list decomposition generalizes the ℓ_∞^d algorithm of Indyk [18], in the case an underlying point-set does not contain a dense ball. Notice that a list decomposition $\mathcal{L}$ corresponds to a nested sequence of overlapping subsets of X. Namely, $(L_1, M_{1,1}, M_{1,2}, R_1)$ first partitions X into four sets, where L_1 is smaller than R_1 with respect to ρ. After that, the subsequent quadruple $(L_2, M_{2,1}, M_{2,2}, R_2)$ is a partition of $M_{1,1} \cup M_{1,2} \cup R_1$, and the process continues. Therefore, if $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i \in [\ell]}$ is a list-decomposition of X with respect to some distribution ρ, the collection $\{L_1, L_2, \dots, L_\ell\}$ are disjoint subsets of X.

For a list-decomposition $\mathcal{L}$, we will build a data structure that proceeds by decomposing the dataset P into $\ell + 1$ (possibly overlapping) subsets $P_1(\mathcal{L}), \dots, P_{\ell+1}(\mathcal{L}) \subset P$ according to

$$(2.2) \quad P_i(\mathcal{L}) = \begin{cases} P \cap (L_i \cup M_{i,1} \cup M_{i,2}) & i \in [\ell] \\ P \cap (R_\ell \cup M_{\ell,1} \cup M_{\ell,2}) & i = \ell + 1 \end{cases}.$$

Then, on a particular query $q \in X$, the data structure will search for the dataset point within the set $P_{\mathsf{Q}_{\mathcal{L}}(q)}(\mathcal{L})$.

The next definition allows us to control the blow-up in space complexity resulting from the overlap between the sets $P_i(\mathcal{L})$.

Definition 2.2. (β-bounded overlap) *Let ρ be a probability distribution over X, and for $\ell \in \mathbb{N}$, let $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i \in [\ell]}$ be a list decomposition of length ℓ with respect to ρ. For $\beta \in \mathbb{R}_{\geq 0}$, we say that $\mathcal{L}$* has a β-bounded overlap with respect to ρ *when*

$$\sum_{i=1}^{\ell+1} \rho(P_i(\mathcal{L}))^{1+\beta} \leq 1.$$

The following definition requires that the sets $P_i(\mathcal{L})$ have measure bounded away from 1, so that we can guarantee that any application of the data structure defined by the list decomposition significantly decreases the number of points that need to be processed.

Definition 2.3. (ξ-progress) *Let ρ be a probability distribution over X, and for $\ell \in \mathbb{N}$, let $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i \in [\ell]}$ be a list decomposition of length ℓ with respect to ρ. For $\xi \in (0, 1)$, we say that $\mathcal{L}$ makes ξ-progress for ρ if $\rho(P_i(\mathcal{L})) \leq 1 - \xi$ holds for all $i \in [\ell + 1]$.*

The final definition we need is that of splitting of points. We emphasize that there will be an asymmetry between dataset and query points in the next definition.

Definition 2.4. (split points) *Let ρ be a probability distribution over X, and for $\ell \in \mathbb{N}$, let $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i \in [\ell]}$ be a list decomposition of length ℓ with respect to ρ. Let $(x, y) \in X \times X$ be points, we say that $\mathcal{L}$ splits (x, y) if there exists $i \in [\ell]$ such that 1) $x \in L_i$ and $y \in M_{i,2} \cup R_i$, or 2) $x \in R_i$ and $y \in L_i \cup M_{i,1}$.*

Definition 2.4 will measure the failure events of the data structure. Specifically, notice that if $x \in P$ is any dataset point and $q \in X$ is a query close to x, then (x, q) is split if and only if

$$x \notin P_{\mathsf{Q}_{\mathcal{L}}(q)}(\mathcal{L}).$$

To see the first direction, notice that if there exists $i \in [\ell]$ where $x \in L_i$ and $q \in M_{i,2} \cup R_i$, then $\mathsf{Q}_{\mathcal{L}}(q) > i$. Since, for every $i' > i$, $L_{i'} \cup M_{i',1} \cup M_{i',2} \cup R_{i'} \subset M_{i,1} \cup M_{i,2} \cup R_i$ and is disjoint from L_i, we have $x \notin P_{\mathsf{Q}_{\mathcal{L}}(q)}(\mathcal{L})$. (The argument for when (x, q) are split because $x \in R_i$ and $q \in L_i \cup M_{i,1}$ follows similarly).

For the other direction, suppose that (x, q) are *not* split, which means every $i \in [\ell]$ satisfies $x \in M_{i,1} \cup M_{i,2} \cup R_i$ whenever $q \in M_{i,2} \cup R_i$ and $x \in L_i \cup M_{i,1} \cup M_{i,2}$ whenever $q \in L_i \cup M_{i,1}$. Let $i^* = \mathsf{Q}_{\mathcal{L}}(q)$, and notice that for every $i < i^*$, $q \in M_{i,2} \cup R_i$, which means $x \in M_{i,1} \cup M_{i,2} \cup R_i = L_{i+1} \cup M_{i+1,1} \cup M_{i+1,2} \cup R_{i+1}$. As a result, $x \in L_{i^*} \cup M_{i^*,1} \cup M_{i^*,2} \cup R_{i^*}$. By definition of i^*, if $i^* < \ell + 1$, then $q \in L_{i^*} \cup M_{i^*,1}$ so that $x \in L_{i^*} \cup M_{i^*,1} \cup M_{i^*,2}$; on the other hand, if $i^* = \ell + 1$, then $q \in M_{\ell,2} \cup R_\ell$, so $x \in M_{\ell,1} \cup M_{\ell,2} \cup R_\ell$. In conclusion, we have that in the event that (p, q) are not split, we maintain the promise that p is inside the (proper) subset of dataset points which the data structure will query given q.

For the rest of this document, for any probability distribution ρ supported on X, as well as $\ell \in \mathbb{N}$ and $\beta \in \mathbb{R}_{\geq 0}$, let

$$\mathsf{L}(\rho, \ell, \beta, \xi) = \left\{ \mathcal{L} : \begin{array}{c} \mathcal{L} \text{ is a list decomposition of length } \ell \\ \text{with respect to } \rho \text{ with } \beta\text{-bounded overlap} \\ \text{and } \xi\text{-progressing} \end{array} \right\}.$$

We may now re-state the theorem of [18], which results in a data structure for ANN over ℓ_∞^d with approximation $O(\log_{1+\beta} \log d)$.

Theorem 2.1. (Main Theorem implicit in [18]) *Fix $d \in \mathbb{N}$ and $\beta \geq 0$. Let $n \in \mathbb{N}$ and $P \subset \mathbb{R}^d$ be* *any subset of n points, and denote ρ_P as the empirical probability distribution supported on P. Suppose that for any $A \subset \mathbb{R}^d$ where*

$$\mathrm{diam}_{\ell_\infty}(A) = \sup_{x,y \in A} \|x - y\|_\infty \le 4\lceil \log_{1+\beta} \log(4d) \rceil,$$

we have $|A \cap P| \le n/2$. Then, there exists $\mathcal{L} \in \mathsf{L}(\rho_P, 1, \beta, 1/(4d))$ such that for all $x \in P$ and all $q \in \mathbb{R}^d$ with $\|x - q\|_\infty \le 1$, $\mathcal{L}$ does not split (x, q).

The data structure of [18] proceeds by iteratively applying Theorem 2.1 to decompose the points into two sets in the case there are no clusters of $n/2$ points within diameter $4\lceil \log_{1+\beta} \log(4d) \rceil$,[6] and simply storing one point in the case such a cluster exists. Hence, once Theorem 2.1 is established, the data structure builds a $O(d \log(n))$-depth tree which *never* splits points (p, q) when $\|p - q\|_\infty \le 1$.

3 Random List Decompositions for ℓ_p^d

We now present the main decomposition theorem for ℓ_p norms, where we construct random list decompositions for point sets with a *projection-resilient separation* property. By a separation property, here we mean the property that no large fraction of the data set points are too clustered. Informally, projection-resilient separation will mean that the points are separated, and remain so when considering subsets of coordinates. For example, Theorem 2.1 implies a decomposition which does not split close points assuming that the point-set is no more than half the dataset points lie in any set of diameter $4\lceil \log_{1+\beta} \log(4d) \rceil$. The *projection-resilient separation* property (to be defined shortly) is a strengthening of the condition of Theorem 2.1 which enables a random list decomposition such that points that are close in the ℓ_p norm are split with low probability after ruling out clusters of smaller diameter.

In order to avoid issues of measurability when proving results about distributions over $\mathbb{R}^d$, we will again work with a fixed and arbitrary finite set $X \subset \mathbb{R}^d$, and will always consider discrete probability distributions ρ whose support is X.

Definition 3.1. *(σ-projection-resilient (Δ, η)-separation in $\mathbb{R}^d$.) Fix $d \in \mathbb{N}$, let $S \subset [d]$ be any set, and let $\Pi_S : \mathbb{R}^d \to \mathbb{R}^{|S|}$ be the projection of points to coordinates in S. Let ρ be any distribution supported on X, and let $\Delta \in \mathbb{R}_{\ge 0}$ and $\eta, \sigma \in (0, 1)$. We say that ρ satisfies* σ-projection-resilient (Δ, η)-separation *in $\mathbb{R}^d$ if for every subset $S \subset [d]$ with $|S| \ge \sigma d$, and every subset $A \subset \mathbb{R}^{|S|}$ of $\mathrm{diam}_{\ell_p^{|S|}}(A) \le \Delta$,*

$$\rho\left(\Pi_S^{-1}(A)\right) = \Pr_{\boldsymbol{x} \sim \rho}\left[\Pi_S(\boldsymbol{x}) \in A\right] \le \eta.$$

The assumption in Theorem 2.1 is equivalent to requiring that ρ_P satisfies 1-projection resilient $(\Delta, \frac{1}{2})$-separation, where $\Delta = 4\lceil \log_{1+\beta} \log(4d) \rceil$. When $\sigma < 1$, however, projection resilient separation makes the stronger requirement that there are no dense clusters even after projection on coordinate subspaces of large dimension.

The rest of this section is devoted to proving the following theorem. Due to the large number of parameters in the theorem, we give a high level overview of their meaning. It is useful to think of $d \in \mathbb{N}$ as the (fixed) ambient dimension and $p \in [1, \infty)$ indicating the ℓ_p norm. At a high level, ϵ will upper bound the probability that close points are split, and $1 - \zeta$ the fraction of dataset points we want the list-decompositions to be useful for. Intuitively, we want the theorem to be useful for as many points as possible and the probability of splitting close points small (i.e., having ζ and ϵ close to 0). Recall that parameters of the list-decomposition ℓ, β and ξ control the size of the list, the space overhead these will incur, and the progress we will make. The size of the list will always be $4d$, the space overhead β is specified by an algorithm designer (typically a small constant), and the progress ξ to a small constant (so it suffices to consider $O(\log n)$ list-decompositions). However, as alluded to earlier, these list-decompositions will only be possible under projection-resilient separation of the dataset P, which are governed with parameters σ, η and Δ; most importantly, is the setting of Δ, as the approximation obtained will depend linearly on it.

We encourage the reader to consider setting the parameters σ, η, ξ, ζ and β to constants, and then, ϵ to a sufficiently small constant. Finally, this gives a setting of the separation Δ for which the list-decompositions will hold.

Theorem 3.1. *Fix $d \in \mathbb{N}$ and $p \in [1, \infty)$, as well as $\epsilon, \sigma, \zeta \in (0, 1)$, $\eta \in (0, 1/100]$, $\xi \in (0, 1/20]$, $\beta \in (0, 1]$ and let*

$$\Delta = 6 \cdot \lceil \log_{1+\beta}(2p/\beta) \rceil \cdot \left(\frac{8 \ln(8d)}{\epsilon \cdot (1 - \sigma)} \right)^{1/p}. \tag{3.3}$$

Let $n \in \mathbb{N}$ and $P \subset X$ be any n point subset. Denote ρ_P as the empirical distribution of P and assume ρ_P satisfies σ-projection-resilient (Δ, η)-separation. Then, there exists a subset $P_0 \subset P$ of size at least $(1 - \zeta)n$ and a distribution $\mathcal{H}$ supported on $\mathsf{L}(\rho_P, 4d, \beta, \xi)$ such that

[6] Notice that the decomposition is into two sets since the length of the list-decomposition in Theorem 2.1 is 1.

for every $x \in P_0$ and every $y \in X$ where $\|x - y\|_p \le 1$,

$$\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x,y)] \le \frac{10\epsilon}{\zeta}.$$

Moreover, every set $L_j, M_{j,1}, M_{j,2}, R_j$ has the form $\{x : a \le x_k \le b, x \notin \cup_{i<j} L_i\}$ (where we set $L_0 = \emptyset$) for some coordinate $k \in [d]$ and reals $a, b \in \mathbb{R}$.

Theorem 3.1 follows from the following lemma, whose proof is the technical bulk of this section.

Lemma 3.2. *Fix $d \in \mathbb{N}$ and $p \in [1,\infty)$, as well as $\epsilon, \sigma \in (0,1)$, $\eta \in (0, 1/100]$, $\xi \in (0, 1/20]$, $\beta \in (0,1]$, and let*

$$\Delta = 6 \cdot \lceil \log_{1+\beta}(2p/\beta)\rceil \cdot \left(\frac{8\ln(8d)}{\epsilon\cdot(1-\sigma)}\right)^{1/p}. \tag{3.4}$$

Let ρ be any distribution supported on $X \subset \mathbb{R}^d$ which satisfies σ-projection-resilient (Δ,η)-separation. Suppose μ is a probability distribution supported on pairs $X \times X$ whose distance in ℓ_p is at most 1 and whose left marginal is ρ. Then, there exists $\ell \le 4d$ and a list decomposition $\mathcal{L} \in \mathsf{L}(\rho, \ell, \beta, \xi)$ which satisfies

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\mathcal{L} \text{ splits } (\boldsymbol{x},\boldsymbol{y})] \le 10\epsilon. \tag{3.5}$$

Moreover, every set $L_j, M_{j,1}, M_{j,2}, R_j$ has the form $\{x : a \le x_k \le b, x_k \notin L_{j-1}\}$ (where we set $L_0 = \emptyset$) for some coordinate $k \in [d]$ and reals $a, b \in \mathbb{R}$.

Proof. [Proof of Theorem 3.1 assuming Lemma 3.2] We will first apply von Neumann's minimax theorem, and then use Markov's inequality to show that such a subset $P_0 \subset P$ must exist. In particular, let $\mathcal{U} \subset \mathbb{R}^{X\times X}$ be the convex set of all probability distributions μ supported on pairs $X \times X$ whose left marginal is exactly ρ_P. Let $m = |\mathsf{L}(\rho_P, 4d, \beta, \xi)|$ and let $\mathcal{V} \subset \mathbb{R}^m$ be the convex set of probability distributions supported on $\mathsf{L}(\rho_p, 4d, \beta, \xi)$. Notice that Lemma 3.2 implies

$$\max_{\mu\in\mathcal{U}} \min_{\mathcal{H}\in\mathcal{V}} \mathop{\mathbf{E}}_{\mathcal{L}\sim\mathcal{H}}\left[\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\mathcal{L} \text{ splits } (\boldsymbol{x},\boldsymbol{y})]\right] \le 10\epsilon,$$

where $\mathcal{H}$ may simply be the distribution which is a point mass at a deterministically chosen list-decomposition $\mathcal{L} \in \mathsf{L}(\rho_P, 4d, \beta, \xi)$, specified by Lemma 3.2. Notice that the function

$$\mathop{\mathbf{E}}_{\mathcal{L}\sim\mathcal{H}}\left[\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\mathcal{L} \text{ splits } (\boldsymbol{x},\boldsymbol{y})]\right] = \sum_{\mathcal{L}\in\mathsf{L}(\rho_p,4d,\beta,\xi)} \sum_{(x,y)\in X\times X} \mathcal{H}(\mathcal{L}) \cdot \mu(x,y) \cdot \mathbf{1}\{\mathcal{L} \text{ splits } (x,y)\}$$

is linear in $\mathcal{H}$ for any fixed μ and linear in μ for any fixed $\mathcal{H}$. Hence, we apply von Neumann's minimax theorem in order to say

$$\min_{\mathcal{H}\in\mathcal{V}} \max_{\mu\in\mathcal{U}} \mathop{\mathbf{E}}_{\mathcal{L}\sim\mathcal{H}}\left[\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\mathcal{L} \text{ splits } (\boldsymbol{x},\boldsymbol{y})]\right] \le 10\epsilon. \tag{3.6}$$

Fix $\mathcal{H} \in \mathcal{V}$ to the minimizer of (3.6). Suppose that there exists a subset $P' \subset P$ of at least ζn points such that for each $x \in P'$, there exists $y_x \in X$ with $\|x - y_x\|_p \le 1$ where

$$\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x, y_x)] > \frac{10\epsilon}{\zeta}. \tag{3.7}$$

Then, consider the distribution $\mu \in \mathcal{U}$ given by letting a sample $(\boldsymbol{x},\boldsymbol{y}) \sim \mu$ be generated by first picking $\boldsymbol{x} \sim \rho_P$, and if $\boldsymbol{x} \in P'$, then letting $\boldsymbol{y} = y_x$, and otherwise choosing $\boldsymbol{y} \sim B_{\ell_p^d}(\boldsymbol{x}, 1) \cap X$ arbitrarily. Notice that by (3.6),

$$\begin{aligned} 10\epsilon &\ge \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (\boldsymbol{x},\boldsymbol{y})]\right] \\ &\ge \zeta \sum_{x\in P'} \Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x, y_x)] > 10\epsilon, \end{aligned}$$

which is a contradiction. Hence, the set of points $x \in P$ for which there exists $y_x \in X$ with $\|x - y_x\|_p \le 1$ satisfying (3.7) is smaller than ζn, so that letting $P_0 = P \setminus P'$ completes the proof. □

3.1 Proof of Lemma 3.2

Definition 3.3. (Bad and Good Coordinates) *Let μ be a probability distribution supported on pairs in $X \times X$ with ℓ_p distance at most 1. For $\sigma \in (0,1)$, we say that a coordinate $k \in [d]$ is* σ-good *with respect to μ if*

$$\mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[|\boldsymbol{x}_k - \boldsymbol{y}_k|^p] \le \frac{1}{(1-\sigma)\cdot d}. \tag{3.8}$$

Otherwise, we say the coordinate $k \in [d]$ is σ-bad *with respect to μ.*

Lemma 3.4. *For any probability distribution μ supported on pairs $X \times X$ at ℓ_p distance at most 1, and any $\sigma \in (0,1)$, there exists at most $(1-\sigma)d$ coordinates which are σ-bad.*

Proof. Let $B \subset [d]$ be the set of σ-bad coordinates. Recall that since μ is supported on pairs at ℓ_p distance at most 1, we have

$$\begin{aligned} 1 &\ge \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\|\boldsymbol{x}-\boldsymbol{y}\|_p^p] \ge \sum_{k\in B} \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[|\boldsymbol{x}_k - \boldsymbol{y}_k|^p] \\ &\ge \frac{1}{(1-\sigma)d}\cdot|B|. \end{aligned}$$

Rearranging gives that $|B| \le (1-\sigma)\cdot d$. □

The following lemma is the cornerstone of the argument. At a high level, we consider a distribution on pairs μ, with left marginal ρ, as well as a σ-good coordinate k and consider a particular kind of partition of $\mathbb{R}^d$ into sets L, M, R.[7] If these partitions exist, these lead to the construction of the list decomposition proving Lemma 3.2. The next lemma assumes such partitions do not exist, and shows that under this case, the marginal distribution of ρ projected onto the k-th coordinate is very concentrated.[8] Applying this argument to all σ-good coordinates will imply that absence of these partitions result in a low-diameter cluster with respect to ρ, and hence violates σ-projection-resilient (Δ, η)-separation.

DEFINITION 3.5. *Let ρ be any probability distribution supported on X. For any $k \in [d]$ and $\gamma \in (0, 1)$, we let $t_{k,\gamma,\rho} \in \mathbb{R}$ be*

$$t_{k,\gamma,\rho} = \inf \left\{ t \in \mathbb{R} : \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \rho} [\boldsymbol{x}_k \le t] \ge \gamma \right\}.$$

The point $m_\rho \in \mathbb{R}^S$ is the median *point of ρ, i.e., for $k \in S$, $(m_\rho)_k = t_{k,1/2,\rho}$.*

Note that $t_{k,\gamma,\rho}$, as a function of γ, is just the inverse of the CDF of the marginal of ρ on coordinate k.

LEMMA 3.6. *Fix any $\beta > 0$ and let $s \in \mathbb{R}_{\ge 0}$ be any parameter. Let ρ be any probability distribution over X. Let $k \in [d]$, and suppose that for $p \ge 1$ and for every $\tau_1 \in \mathbb{R}$ with $\tau_1 \le (m_\rho)_k$, the set*

$$L = \left\{ x \in \mathbb{R}^d : x_k \le \tau_1 \right\}, \qquad \textit{and for}$$

$$\tau_2 = \left(\frac{s}{\rho(L)} \right)^{1/p},$$

the sets

$$M = \left\{ x \in \mathbb{R}^d : x_k \in (\tau_1, \tau_1 + 2\tau_2) \right\} \qquad \textit{and}$$

$$R = \left\{ x \in \mathbb{R}^d : x_k \ge \tau_1 + 2\tau_2 \right\}$$

satisfy

$$\rho(L \cup M)^{1+\beta} + \rho(M \cup R)^{1+\beta} \ge 1 \tag{3.9}$$

whenever $\rho(L) \ge 1/(4d)$, then,

$$\begin{aligned} &\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \rho} \left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1} \left\{ t_{k,1/(4d),\rho} \le \boldsymbol{x}_k \le (m_\rho)_k \right\} \right] \\ &\le s \cdot \left(6 + 2 \cdot \lceil \log_{1+\beta}(p/\beta) \rceil \right)^p \cdot \ln(8d). \end{aligned} \tag{3.10}$$

Proof. [Proof of Lemma 3.6] The proof will proceed in two main steps. The first is to apply an argument of [18], where (3.9) aids in upper bounding $|t_{k,\gamma,\rho} - m_k|$ for any $\gamma \in [1/(4d), 1/2]$. Then, (3.10) will follow from integrating the upper bound on $|t_{k,\gamma,\rho} - (m_\rho)_k|^p$ over all $\gamma \in [1/(4d), 1/2]$. For the first step, let $\gamma \in [1/(4d), 1/2]$, and consider the following (iteratively defined) sequence of numbers $(\kappa_i)_{i \in \mathbb{Z}_{\ge 0}}$ and tuples of sets $\{(L_i, M_i, R_i)\}_{i \in \mathbb{Z}_{\ge 0}}$:

$$\begin{aligned} \kappa_0 &= t_{k,\gamma,\rho} \qquad \text{and} \qquad L_i = \left\{ x \in \mathbb{R}^d : x_k \le \kappa_i \right\} \\ \kappa_{i+1} &= \kappa_i + 2 \left(\frac{s}{\rho(L_i)} \right)^{1/p} \\ M_i &= \left\{ x \in \mathbb{R}^d : x_k \in (\kappa_i, \kappa_{i+1}) \right\} \\ R_i &= \left\{ x \in \mathbb{R}^d : x_k \ge \kappa_{i+1} \right\} \qquad \text{for all } i \in \mathbb{N}. \end{aligned}$$

For all $i \ge 0$, the sequence $(\rho(L_i))_{i \in \mathbb{Z}_{\ge 0}}$ is non-decreasing and always bounded below by $\gamma \ge 1/(4d)$, which means that as long as $\kappa_i \le (m_\rho)_k$, (3.9) holds with (L_i, M_i, R_i). Denote

$$i_0 = \max \left\{ i \in \mathbb{Z}_{\ge 0} : \kappa_i \le (m_\rho)_k \right\}, \tag{3.11}$$

Furthermore, notice that $L_i \cup M_i = L_{i+1}$ and $\rho(M_i \cup R_i) = (1 - \rho(L_i))$. Hence, we apply (3.9) to all $i \in \{0, \dots, i_0\}$ to conclude

$$\begin{aligned} 1 &\le \rho(L_i \cup M_i)^{1+\beta} + \rho(M_i \cup R_i)^{1+\beta} \\ &= \rho(L_{i+1})^{1+\beta} + (1 - \rho(L_i))^{1+\beta} \\ &\le \rho(L_{i+1})^{1+\beta} + 1 - \rho(L_i), \end{aligned}$$

where we used the fact that $\beta > 0$ and $(1 - \rho(L_i)) \in [0, 1]$ to say $(1 - \rho(L_i))^{1+\beta} \le 1 - \rho(L_i)$. Therefore, for every $i \in \{0, \dots, i_0\}$, $\rho(L_{i+1})^{1+\beta} \ge \rho(L_i)$, which implies that for every $i \in \{0, \dots, i_0\}$,

$$\frac{1}{\rho(L_i)} \le \left(\frac{1}{\rho(L_0)} \right)^{\frac{1}{(1+\beta)^i}} = \left(\frac{1}{\gamma} \right)^{\frac{1}{(1+\beta)^i}}. \tag{3.12}$$

Since $\rho(L_{i_0}) \le 1/2$ because $\kappa_{i_0} \le (m_\rho)_k$, we may conclude that

$$2 \le \left(\frac{1}{\gamma} \right)^{\frac{1}{(1+\beta)^{i_0}}}. \tag{3.13}$$

In addition, we have

$$\kappa_{i+1} - \kappa_i \le 2 s^{1/p} \cdot \frac{1}{\rho(L_i)^{1/p}} \le 2 s^{1/p} \left(\frac{1}{\gamma} \right)^{\frac{1}{p} \cdot \frac{1}{(1+\beta)^i}}, \tag{3.14}$$

[7] Specifically, we will split M into two sets M_1 and M_2, and use (L, M_1, M_2, R) as one of the tuples in the list-decomposition.

[8] This notion of concentration is formalized in terms upper-bounding the p-th moment of a random variable relating the distance from $\boldsymbol{x}_k$ to the median when $\boldsymbol{x} \sim \rho$.

which means, since $\kappa_{i_0+1} > (m_\rho)_k$, we may upper bound $|t_{k,\gamma,\rho} - (m_\rho)_k|$ by

$$\begin{aligned} |t_{k,\gamma,\rho} - (m_\rho)_k| &\le \sum_{i=0}^{i_0} (\kappa_{i+1} - \kappa_i) \\ &\le 2s^{1/p} \sum_{i=0}^{i_0} (1/\gamma)^{\frac{1}{p} \cdot \frac{1}{(1+\beta)^i}}, \\ &= 2s^{1/p} \sum_{i=0}^{i_0} c^{(1+\beta)^i/p}, \end{aligned} \tag{3.15}$$

where we let

$$c = \left(\frac{1}{\gamma}\right)^{\frac{1}{(1+\beta)^{i_0}}} \ge 2,$$

and the inequality follows from (3.13). In order to upper bound the sum in the right-hand side of (3.15), let $j_0 = \lceil \log_{1+\beta}(p/\beta) \rceil$, and notice that for $i > j_0$,

$$c^{(1+\beta)^i/p} \ge c \cdot c^{(1+\beta)^{i-1}/p},$$

which, implies that after $i \ge j_0$ summands begin to increase (more than) geometrically with constant 2. Hence,

$$\begin{aligned} \sum_{i=0}^{i_0} c^{(1+\beta)^i/p} &\le (3 + j_0) \cdot c^{(1+\beta)^{i_0}/p} \\ &\le (3 + \lceil \log_{1+\beta}(p/\beta) \rceil) \cdot \left(\frac{1}{\gamma}\right)^{1/p} \end{aligned}$$

Finally, substituting into (3.15) implies

$$|t_{k,\gamma,\rho} - (m_\rho)_k|^p \le s \cdot \left(6 + 2 \cdot \lceil \log_{1+\beta}(p/\beta) \rceil\right)^p \cdot \frac{1}{\gamma}.$$

In order to show (3.10), notice that

$$\begin{aligned} &\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1}\left\{t_{k,1/(4d),\rho} \le \boldsymbol{x}_k \le (m_\rho)_k\right\}\right] \\ &\qquad = \int_{1/(4d)}^{1/2} |t_{k,\gamma,\rho} - (m_\rho)_k|^p d\gamma \\ &\le s \cdot \left(6 + 2 \cdot \lceil \log_{1+\beta}(p/\beta) \rceil\right)^p \cdot \ln(8d). \end{aligned}$$

∎

LEMMA 3.7. *Fix $\epsilon, \sigma \in (0,1)$, $\eta \in (0,1/2)$, and $\beta \in (0,1]$, and*

$$\Delta = 6 \cdot \lceil \log_{1+\beta}(2p/\beta) \rceil \cdot \left(\frac{4\ln(8d)}{\epsilon \cdot (1-2\eta) \cdot (1-\sigma)}\right)^{1/p}.$$

Let ρ be a probability distribution supported on X, and μ be a probability distribution supported on pairs in $X \times X$ with ℓ_p distance at most 1, and whose left-marginal is ρ. Suppose ρ satisfies σ-projection-resilient (Δ, η)-separation. Then there exist disjoint sets $L, M_1, M_2, R \subset \mathbb{R}^d$ such that $X \subseteq L \cup M_1 \cup M_2 \cup R$, and the following conditions are satisfied:

1. *We have $1/(4d) \le \rho(L) \le 1/2$ and $\rho(R) \ge 1/10$.*
2. *The sets give a list-decompositions with β-bounded overlap, i.e.,*
$$\rho(L \cup M_1 \cup M_2)^{1+\beta} + \rho(M_1 \cup M_2 \cup R)^{1+\beta} \le 1.$$
3. *Points sampled from μ are split with low probability, i.e.,*
$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\boldsymbol{y} \in R \cup M_2 \wedge \boldsymbol{x} \in L] \le \epsilon\rho(L) \qquad \textit{and}$$
$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\boldsymbol{y} \in L \cup M_1 \wedge \boldsymbol{x} \in R] \le \epsilon\rho(L).$$
4. *Each of the sets L, M_1, M_2, R has the form $\{x : a \le x_k \le b\}$ for some coordinate $k \in [d]$ and reals $a, b \in \mathbb{R}$.*

Proof. We proceed by contradiction. Let $S \subset [d]$ be the set of σ-good coordinates, for which we know, by Lemma 3.4, that $|S| \ge \sigma d$. For each $k \in S$, let

$$\kappa_{k,-} = \inf\left\{t \in \mathbb{R} : \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\rho}[\boldsymbol{x}_k \le t] \ge \frac{1}{4d}\right\} \qquad \text{and}$$

$$\kappa_{k,+} = \sup\left\{t \in \mathbb{R} : \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\rho}[\boldsymbol{x}_k \ge t] \ge \frac{1}{4d}\right\},$$

and consider the subset $\mathcal{U} \subset \mathbb{R}^d$ where

$$\mathcal{U} = \left\{x \in \mathbb{R}^d : \forall k \in S, x_k \in [\kappa_{k,-}, \kappa_{k,+}]\right\}$$

$$\text{which satisfies} \qquad \rho(\mathcal{U}) \ge 1/2$$

from a union bound. We now apply Lemma 3.6 to every $k \in S$ with the parameter

$$s = \frac{1}{\epsilon \cdot (1-\sigma) \cdot d}.$$

Specifically, assume for the sake of contradiction, that for every $k \in S$ and $\tau \in \mathbb{R}$ with $\tau \le m_k$, the set

$$L = \left\{x \in \mathbb{R}^d : x_k \le \tau\right\} \qquad \text{and} \qquad \tau_2 = \left(\frac{s}{\rho(L)}\right)^{1/p},$$

as well as the sets

$$\begin{aligned} M_1 &= \left\{x \in \mathbb{R}^d : \tau \le x_k \le \tau + \tau_2\right\} \\ M_2 &= \left\{x \in \mathbb{R}^d : \tau + \tau_2 \le x_k \le \tau + 2\tau_2\right\} \\ R &= \left\{x \in \mathbb{R}^d : x_k \ge \tau + 2\tau\right\} \end{aligned}$$

satisfying $1/(4d) \le \rho(L)$, have

$$\rho(L \cup M_1 \cup M_2)^{1+\beta} + \rho(M_1 \cup M_2 \cup R)^{1+\beta} \ge 1.$$

We conclude that every $k \in S$ satisfies

$$\begin{aligned}&\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1}\left\{\kappa_{k,-} \le \boldsymbol{x}_k \le (m_\rho)_k\right\}\right] \\ &\le \frac{\left(6 + 3\cdot\lceil\log_{1+\beta}(p/\beta)\rceil\right)^p \cdot \ln(8d)}{\epsilon\cdot(1-\sigma)\cdot d} \\ &\le \frac{\left(6\cdot\lceil\log_{1+\beta}(2p/\beta)\rceil\right)^p \cdot \ln(8d)}{\epsilon\cdot(1-\sigma)\cdot d}.\end{aligned}$$

Analogously, we apply the same argument to the *right side* of $(m_\rho)_k$. Assume, for the sake of contradiction, that for every $k \in S$ and $\tau \in \mathbb{R}$ with $t \ge (m_\rho)_k$, the set $L = \{x \in \mathbb{R}^d : x_k \ge \tau\}$ and $\tau_2 = (s/\rho(L))^{1/p}$, as well as $M_1 = \{x \in \mathbb{R}^d : \tau - \tau_2 \le x_k \le \tau\}$, $M_2 = \{x \in \mathbb{R}^d : \tau - 2\tau_2 \le x_k \le \tau - \tau_2\}$, and $R = \{x \in \mathbb{R}^d : x_k \le \tau - 2\tau_2\}$ satisfying $\rho(L) \ge 1/(4d)$ also has $\rho(L \cup M_1 \cup M_2)^{1+\beta} + \rho(M_1 \cup M_2 \cup R)^{1+\beta} \ge 1$. Then every $k \in S$ satisfies

$$\begin{aligned}&\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1}\{\kappa_{k,+} \ge \boldsymbol{x}_k \ge (m_\rho)_k\}\right] \\ &\le \frac{\left(6\cdot\lceil\log_{1+\beta}(2p/\beta)\rceil\right)^p \cdot \ln(8d)}{\epsilon\cdot(1-\sigma)\cdot d}.\end{aligned}$$

We conclude

$$\begin{aligned}&\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[\|\Pi_S(\boldsymbol{x}) - m_\rho\|_p^p \cdot \mathbf{1}\left\{\boldsymbol{x} \in \mathcal{U}\right\}\right] \\ &\le \sum_{i\in S}\Bigg(\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1}\{\kappa_{k,-} \le \boldsymbol{x}_k \le (m_\rho)_k\}\right] \\ &+ \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\rho}\left[|\boldsymbol{x}_k - (m_\rho)_k|^p \cdot \mathbf{1}\{\kappa_{k,+} \ge \boldsymbol{x}_k \ge (m_\rho)_k\}\right]\Bigg) \\ &\le \frac{2}{\epsilon\cdot(1-\sigma)} \cdot \left(6\cdot\lceil\log_{1+\beta}(p/\beta)\rceil\right)^p \cdot \ln(8d).\end{aligned}$$

Therefore, we can define a subset $A \subset \mathbb{R}^{|S|}$ as an ℓ_p ball around $m_\rho \in \mathbb{R}^{|S|}$ of diameter

$$\begin{aligned}&\mathrm{diam}_{\ell_p^{|S|}}(A) \\ &\le 6\cdot\lceil\log_{1+\beta}(2p/\beta)\rceil\left(\frac{2\ln(8d)}{\epsilon\cdot(1-\sigma)\cdot(1/2-\eta)}\right)^{1/p},\end{aligned}$$

which satisfies, by Markov's inequality applied to p-th moments of $\|\Pi_S(\boldsymbol{x}) - m_\rho\|_p \cdot \mathbf{1}\{\boldsymbol{x} \in \mathcal{U}\}$,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\rho}\left[\Pi_S(\boldsymbol{x}) \notin A \wedge \boldsymbol{x} \in \mathcal{U}\right] \le 1/2 - \eta,$$

Since $\rho(\mathcal{U}) \ge 1/2$, we have that $\mathbf{Pr}_{\boldsymbol{x}\sim\rho}[\Pi_S(\boldsymbol{x}) \in A \wedge \boldsymbol{x} \in \mathcal{U}] \ge \eta$. This contradicts the assumption that ρ satisfies σ-projection-resilient (Δ, η)-separation. Hence, there exists a coordinate $k \in S$, and k-axis-aligned disjoint subsets $L, M_1, M_2, R \subset \mathbb{R}^d$ satisfying $1/(4d) \le \rho(L) \le 1/2$ where

$$\rho(L \cup M_1 \cup M_2)^{1+\beta} + \rho(M_1 \cup M_2 \cup R)^{1+\beta} \le 1.$$

The above inequality proves (2). It also implies $\rho(R) \ge 1/10$ since, otherwise, we would have

$$\begin{aligned}&1 < (9/10)^2 + (1/2)^2 \le \\ &\rho(L \cup M_1 \cup M_2)^{1+\beta} + \rho(M_1 \cup M_2 \cup R)^{1+\beta} \le 1,\end{aligned}$$

where we used that $\beta \le 1$ and $\rho(L) \le 1/2$. This proves (1). Furthermore, recall that k is a σ-good coordinate, so for any $C \subseteq \mathbb{R}^d$, we have

$$\begin{aligned}\epsilon s = \frac{1}{(1-\sigma)\cdot d} &\ge \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[|\boldsymbol{x}_k - \boldsymbol{y}_k|^p\right] \\ &\ge \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[|\boldsymbol{x}_k - \boldsymbol{y}_k|^p \cdot \mathbf{1}\{\boldsymbol{x} \in C\}\right].\end{aligned}$$

Notice that, by applying Markov's inequality to the expectation above with $C = L$ and $C = R$, we have

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[\boldsymbol{y} \in M_2 \cup R \wedge \boldsymbol{x} \in L\right] \\ &\le \mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[\left(|\boldsymbol{x}_k - \boldsymbol{y}_k|^p \ge \frac{s}{\rho(L)}\right) \wedge (\boldsymbol{x} \in L)\right] \le \epsilon\cdot\rho(L), \\ &\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[\boldsymbol{y} \in L \cup M_1 \wedge \boldsymbol{x} \in R\right] \\ &\le \mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[\left(|\boldsymbol{x}_k - \boldsymbol{y}_k|^p \ge \frac{s}{\rho(L)}\right) \wedge (\boldsymbol{x} \in R)\right] \\ &\le \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}\left[|\boldsymbol{x}_k - \boldsymbol{y}_k|^p \cdot \mathbf{1}\{\boldsymbol{x} \in R\}\right] \cdot \frac{\rho(L)}{s} \\ &\le \epsilon\cdot\rho(L).\end{aligned}$$

The above inequalities prove (3). □

4 ANN for ℓ_p from Theorem 3.1

For any $d \in \mathbb{N}$, we consider $X_d \subset \mathbb{R}^d$ to be a discretization of points in $\mathbb{R}^d$ that the algorithm will receive. For example, if the algorithm receives points in $\mathbb{R}^d$ represented as a sequence of d numbers, each with b-bit-sized words, then X_d is of size 2^{bd}. We focus attention on a specific style of data structure which are DAG-like.

Definition 4.1. *For $d \in \mathbb{N}$, A DAG-like data structure for ANN over X_d is a rooted directed acyclic graph $G = (V, E)$ where each node $v \in V$ has an associated function f_v. For each node $v \in V$ with at least one child, let $N(v) \subset V$ be the out-neighborhood of v, and let $\deg(G) = \max_{v\in V}|N(v)|$ denote the maximum out-degree of any node in V. Let the query-depth,* $\mathrm{depth}(G)$, *be the length of the longest path, and* $\text{width}(G)$ *be the maximum width when* G *is expressed as a layered graph. Every node* $v \in V$ *contains the description of a function* $f_v : X_d \to N(v) \cup X_d \cup \{\perp\}$*. Let* $r \in V$ *be the (unique) root of* G*; a query* $q \in X$ *is executed by beginning at* r *and walking down the path specified by* $f_v(q)$ *if* $f_v(q) \in N(v)$*, outputting* $f_v(q)$ *if* $f_v(q) \in X_d$*, or outputting* $\perp$ *(to indicate failure) if* $f_v(q) = \perp$*.*

We observe that a DAG-like data structure with efficient encodings (with respect to space and time) of f_v may be stored and traversed efficiently. For our case, f_v will be encoded in $O(d)$ words, and there exists an efficient algorithm which takes as input the encoding of a function f_v, as well as a point $q \in X_d$, and outputs $f_v(q)$ in $O(d)$ time. This results in a data structure whose space complexity is

$$\text{Space: } \text{width}(G) \cdot \text{depth}(G) \cdot O(d),$$

and whose time complexity is

$$\text{Time: } \text{depth}(G) \cdot O(d).$$

THEOREM 4.1. (ANN FOR ℓ_p^d) *Fix* $d \in \mathbb{N}$, $\beta, \alpha \in (0,1]$*, and*

$$c = O\left(\lceil \log_{1+\beta}(p/\beta)\rceil \cdot \left(\frac{\log^2 d}{\alpha^{1/\alpha+1}\log(1/\alpha)}\right)^{1/p}\right).$$

For any $n \in \mathbb{N}$ *and a dataset* $P \subset X_d$ *of* n *points, there exists a DAG-like data structure for* $(1,c)$*-ANN over* ℓ_p^d *over the dataset* P*, such that the data structure is of width* $O(dn^{1+\beta})$ *and query depth* $O(dn^{\alpha})$*, and succeeds with probability at least* $n^{-\alpha}$*.*

We prove Theorem 4.1 after some necessary lemmas, which break deal with the cases when the dataset satisfies σ-projection-resilient (Δ, η)-separation, and when it does not.

LEMMA 4.2. *(Sufficiency of List-Decompositions with Definitions 2.2, 2.3, and 2.4) Fix* $d, \ell \in \mathbb{N}$*, monotone functions* $f_1, f_2 : \mathbb{N} \to \mathbb{N}$*, with* $f_1(d) \geq \ell + 2$*, and* $\beta, \xi, \epsilon, \alpha, \zeta \in (0,1)$*,* $\zeta \leq 1 - \xi$ *and* $c > 1$*. Let* $n \in \mathbb{N}$*, such that*

$$\frac{1}{f_2(d) \cdot n^{\alpha}} + (1-\xi)^{\alpha} + \zeta^{\alpha} \leq 1. \tag{4.16}$$

Consider any dataset $P \subset X_d$ *of size* n*, and denote by* ρ_P *the empirical distribution of* P*. Suppose the following two conditions hold:*

1. *(Data-dependent list-decompositions) There exists a distribution* $\mathcal{H}$ *supported on* $\mathsf{L}(\rho_P, \ell, \beta, \xi)$ *and a subset* $P_0 \subset P$ *of size* $|P_0| \geq (1-\zeta)n$ *such that for any* $x \in P_0$ *and* $y \in X_d$ *where* $\|x - y\|_p \leq 1$*,*
$$\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x,y)] \leq \epsilon.$$
2. *(Inductive assumption) Suppose that for any* $m \leq (1-\xi)n$ *and any dataset in* X_d *of size* m *there exists a DAG-like data structure for* $(1,c)$*-ANN over* ℓ_p^d *of width at most* $f_1(d)\cdot m^{1+\beta}$ *and query depth at most* $f_2(d) \cdot m^{\alpha}$ *succeeding with probability at least* $m^{-\varrho}$*, where*
$$\varrho = \frac{\log\left(\frac{1}{1-\epsilon}\right)}{\log\left(\frac{1}{1-\xi}\right)}. \tag{4.17}$$

Then, there exists a DAG-like data structure for $(1,c)$*-ANN over* ℓ_p^d *with the dataset* P *of width at most* $f_1(d) \cdot n^{1+\beta}$ *and query-depth at most* $f_2(d) \cdot n^{\alpha}$ *which succeeds with probability at least* $n^{-\varrho}$*.*

Proof. We build a DAG-like data structure recursively. Let r be the root, corresponding to the dataset $P \subset X_d$ of size n. We sample $\mathcal{L} \sim \mathcal{H}$ and decompose P into $\ell + 1$ sets $P_1(\mathcal{L}), \ldots, P_{\ell+1}(\mathcal{L}) \subset P$ according to (2.2). For each $i \in [\ell+1]$, we use the inductive assumption to sample a DAG-like data structure D_i for points in $P_i(\mathcal{L})$; letting v_i be the root node of D_i, we make v_i a child of r. The query map $\mathsf{Q}_{\mathcal{L}} : X_d \to [\ell+1]$ specifies f_r, by mapping a query $q \in X_d$ to the node v_i where $\mathsf{Q}_{\mathcal{L}}(q) = i$. In addition, we use the inductive assumption again in order to sample a DAG-like data structure D_0 for the dataset $P \setminus P_0$ containing at most ζn points (where we use the fact $\zeta \leq 1 - \xi$), and let r_0 be the root of D_0. For every $i \in [\ell+1]$ and node v in D_i, we modify the function f_v so that every input $q \in X_d$ where $f_v(q) = \perp$, now has $f_v(q) = r_0$ (see Figure 1).

For any $m \in \mathbb{N}$, let $T(m)$ denote the maximum query depth of a data structure built for m-point datasets, which by assumption is at most $f_2(d)\cdot m^{\alpha}$. Let $T_0(n)$ be the maximum query depth of a data structure for n point datasets satisfying condition 1. Then, for any $\mathcal{L} \in \mathsf{L}(\rho_P, \ell, \beta, \xi)$,

$$\begin{aligned}
&T_0(n) \\
&\leq 1 + T((1-\xi)n) + T(\zeta n) \\
&\leq f_2(d)n^{\alpha}\left(1/(f_2(d)\cdot n^{\alpha}) + (1-\xi)^{\alpha} + \zeta^{\alpha}\right) \\
&\leq f_2(d) \cdot n^{\alpha},
\end{aligned}$$

where in the last inequality, we use (4.16). In order to bound the width of the data structure, let $W(m)$ be the maximum width of DAG-like data structure for datasets of size m, and let $W_0(n)$ be the maximum width

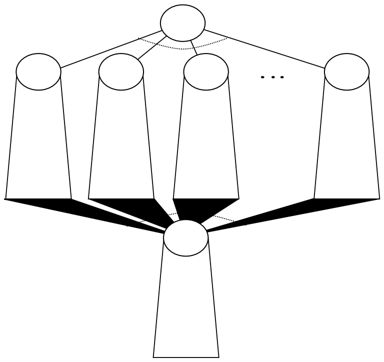

Figure 1: Composition of data structures after one application of the list-decomposition $\mathcal{L} \sim \mathcal{H}$ to search through $r_1, \dots, r_{\ell+1}$, and then search through r_0.

of a DAG-like data structure for datasets P of n points satisfying condition 1. Then, for any $\mathcal{L} \in \mathsf{L}(\rho_P, \ell, \beta, \xi)$,

$$\begin{aligned} &W_0(n) \leq \\ &\max \Big\{W(n \cdot \rho(R_{\ell+1} \cup M_{\ell+1,1} \cup M_{\ell+1,2})) \\ &+ \sum_{i=1}^{\ell+1} W(n \cdot \rho(L_i \cup M_{i,1} \cup M_{i,2})), W(\zeta n)\Big\} \\ &\leq f_1(d) \cdot n^{1+\beta} \max\{1, \zeta^{\alpha}\} \leq f_1(d) \cdot n^{1+\beta}, \end{aligned}$$

where we used the fact that every $\mathcal{L} \in \mathsf{L}(\rho_P, \ell, \beta, \xi)$ has β-bounded overlap and the fact that $f_1(d) \geq 5d$. Finally, suppose $q \in X_d$ and $x \in P$ where $\|q - x\|_p \leq 1$. If $x \in P \setminus P_0$, then the probability of success is at least $(\zeta n)^{-\varrho} \geq n^{-\varrho}$, since the query will (if not already succeeded), query q on r_0, and there, we consider the event that D_0 succeeds. On the other hand, if $x \in P_0$, then the data structure succeeds when x and q are not split by $\mathcal{L} \sim \mathcal{H}$, and for $\mathsf{Q}_{\mathcal{L}}(q) = i$, the data structure D_i succeeds. This occurs with probability at least

$$(1-\epsilon) \cdot (1-\xi)^{-\varrho} \cdot n^{-\varrho} \geq n^{-\varrho},$$

by the setting of ϱ in (4.17). This concludes the proof. ∎

Lemma 4.3. *(Sufficiency of Failure of σ-Projection-Resilient (Δ, η)-Separation) Fix $d, C_t \in \mathbb{N}$, a monotone function $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, and $\sigma, \eta, \alpha, \delta \in (0,1)$, $\Delta \in \mathbb{R}_{\geq 0}$. Suppose that*

$$(1-\sigma)^{C_t} + (1-\eta)^{\alpha} \leq 1, \tag{4.18}$$

and let

$$c(d) = (\Delta + 1) \cdot \left(\frac{\log d}{\log\left(\frac{1}{1-\sigma}\right)} \right)^{1/p}. \tag{4.19}$$

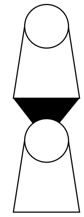

Figure 2: Composing D_1 and D_2 for the proof of Lemma 4.3.

Let $n \in \mathbb{N}$, and consider a dataset $P \subset X_d$ of size n, and denote by ρ_P the empirical distribution of P. Suppose the following three conditions hold:

1. *(ρ_P fails σ-projection-resilient (Δ, η)-separation) There exists a subset $S \subset [d]$ with $|S| \geq \sigma d$, as well as a subset $A \subset \mathbb{R}^{|S|}$ of $\mathrm{diam}_{\ell_p^{|S|}}(A) \leq \Delta$ such that*

$$\tilde{P} = \{p \in P : \Pi_S(p) \in A\} \quad \text{satisfies} \quad |\tilde{P}| \geq \eta n.$$

2. *(Inductive assumption for smaller dimension) Suppose that for any $d_0 \leq (1-\sigma)d$ and any dataset of $m \leq n$ points, there exists a DAG-like data structure for $(1, c(d_0))$-ANN over $\ell_p^{d_0}$ that has width at most $f(m, d_0)$ and has query-depth at most $m^{\alpha} \cdot (d_0)^{C_t}$, and fails with probability at most δ.*

3. *(Inductive assumption for smaller datasets) Suppose that for any $m \leq (1-\eta)n$ and any dataset of m points, there exists a DAG-like data structure for $(1, c(d))$-ANN over ℓ_p^d that has width $f(m, d)$, has query-depth $m^{\alpha} \cdot d^{C_t}$, and fails with probability at most δ.*

Then, there exists a DAG-like data structure for $(1, c(d))$-ANN over ℓ_p^d with the dataset P of width at most $f(n, d)$ and query-depth at most $n^{\alpha} \cdot d^{C_t}$ which fails with probability at most δ.

Proof. [Proof of Theorem 4.1] The proof proceeds by induction on $n, d \in \mathbb{N}$. We notice that for any $d \in \mathbb{N}$, the case $n = 1$ is trivial, by storing a single point. Suppose we assume the theorem for all $n < n_0$ by inductive assumption, and since for $d = 1$, the theorem holds trivially by a binary search tree, we consider $d_0 \in \mathbb{N}$ such that for all $d < d_0$, the theorem statement is true with n_0 and d. Let $P \subset X_{d_0}$ be a dataset of n_0

points. We consider setting:

$$\eta = \frac{1}{100}, \qquad \xi = \frac{1}{20}, \qquad \zeta = (\xi\alpha/2)^{1/\alpha}$$
$$\text{and} \qquad \sigma = 1 - \alpha \cdot \eta$$

so that for every $n \geq 1/(2\xi\alpha)^{1/\alpha}$, we satisfy (4.16), and for $C_t \geq 1$, we satisfy (4.18). Furthermore, we let

$$\epsilon = \frac{\alpha \cdot \zeta}{1000},$$

so that

$$\varrho = \frac{\log\left(\frac{1}{1-10\epsilon/\zeta}\right)}{\log\left(\frac{1}{1-\xi}\right)} \leq \alpha.$$

Consider $\Delta \in \mathbb{R}_{\geq 0}$ as set in (3.3), and suppose that the empirical distribution ρ_P satisfies σ-projection-resilient (Δ, η)-separation. Then, we apply Theorem 3.1 to obtain a distribution $\mathcal{H}$ supported on $\mathsf{L}(\rho_P, 4d, \beta, \xi)$ and a subset $P_0 \subset P$ be a subset of size at least $(1-\zeta)n$ such that for any $x \in P_0$ and $y \in X_{d_0}$ with $\|x - y\|_p \leq 1$,

$$\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x, y)] \leq \epsilon.$$

We may apply Lemma 4.2, where we notice that the inductive assumption is satisfied since $(1-\xi)n \leq n_0$. Thus, we obtain a data structure with the specified guarantees for P.

On the other hand, suppose P fails to satisfy σ-projection-resilient (Δ, η)-separation. Then, we notice again, that the inductive hypothesis with $d < d_0$ and $n < n_0$ holds, so that we may apply Lemma 4.3. $\square$

5 Random List Decompositions for $\ell_1^d(Z)$

We now present the random list decompositions for $\ell_1^d(Z)$. The main theorem in this section is stated next, but we defer the proof until some necessary lemmas are set in place. We notice that unlike the case of ℓ_p norms, the (r, cr)-ANN problem may not be equivalent, in terms of complexity, for all $r \in \mathbb{R}_{\geq 0}$, since the metric space (Z, d_Z) does not necessarily scale. For a metric space (Z, d_Z) of diameter $R \in \mathbb{R}_{\geq 0}$, the random list decompositions for $\ell_1^d(Z)$ presented next are especially tailored to the setting of $cr = \Theta(dR)$, corresponding to the maximum scale in $\ell_1^d(Z)$. When $cr \ll dR$, we will introduce one more ingredient to reduce to the case where the list decompositions will apply.

Theorem 5.1. *Let (Z, d_Z) be a finite metric space of diameter $R \in \mathbb{R}_{\geq 0}$, and $d \in \mathbb{N}$, as well as $\beta \geq 0$, $\epsilon \in (0, 1/16)$, $c \in (0, 1)$ and set*

$$\tau = \left\lceil \log_{1+\beta}\left(\frac{16 \ln |Z|}{c^2}\right)\right\rceil, m = \frac{128(\tau+1)^2 \ln(1/\epsilon)}{c^2},$$
$$\text{and}\, \eta \leq \frac{1}{4|Z|^m}. \tag{5.20}$$

Let $n \in \mathbb{N}$ and $P \subset Z^d$ be any subset of n points, and denote ρ_P as the empirical probability distribution of P. Suppose that for any $A \subset Z^d$ where

$$\mathrm{diam}_{\ell_1(Z)}(A) = \sup_{x,y\in A} d_{\ell_1^d(Z)}(x, y) \leq c \cdot d \cdot R,$$

we have $|A \cap P| \leq \eta n$. Then, there exists a distribution $\mathcal{H}$ supported on $\mathsf{L}(\rho_P, |Z|^m, \beta, 1/16)$ such that for any $x, y \in Z^d$ with

$$d_{\ell_1^d(Z)}(x, y) = \sum_{i=1}^{d} d_Z(x_i, y_i) \leq \frac{1}{16} \cdot \frac{c \cdot d \cdot R}{\tau + 1} \tag{5.21}$$

we have

$$\Pr_{\mathcal{L}\sim\mathcal{H}}[\mathcal{L} \text{ splits } (x, y)] \leq 2\epsilon.$$

For the remainder of the section, we fix $n, d \in \mathbb{N}$, and a finite metric space (Z, d_Z) of diameter at most $R \in \mathbb{R}_{\geq 0}$. We will show $\mathcal{L}$ exists by analyzing an algorithm which attempts to construct $\mathcal{L}$. For a parameter $m \in \mathbb{N}$, consider the procedure `BuildList` which takes as input P, a (multi-)set $S \subset [d]$ of size m, as well as additional parameters, and either outputs a list-decomposition $\mathcal{L} \in \mathsf{L}(\rho_P, |Z|^m, \beta, \xi)$ or outputs `fail`. (see Figure 5).

Subroutine `BuildList`$(P, S, \beta, \xi, \tau, \Delta)$

Input: A dataset $P \subset Z^d$ of size n, as well as a multi-set $S \subset [d]$ of size m, and parameters $\beta \geq 0$, $\xi \in (0, 1)$, $\tau \in \mathbb{N}$ and $\Delta \in \mathbb{R}_{\geq 0}$.
Output: either a list-decomposition $\mathcal{L} \in \mathsf{L}(\rho_P, |Z|^m, \beta, \xi)$, or "fail."

- Initialize an index $j = 1$ which increases as we build $\mathcal{L} = \{(L_j, M_{j,1}, M_{j,2}, R_j)\}_j$. Initialize $P_1 \leftarrow P$ and $R_0 \cup M_{j-1,0} \cup M_{j-1,2} = Z^d$, and proceed as follows until we break out of the following loop, or we output "fail":
  1. If $|P_j| \leq (1-\xi)n$, break out of the loop.
  2. Pick the first point $x \in P_j$ satisfying
  $$|\{y \in P_j : \Pi_S(y) = \Pi_S(x)\}| \geq \frac{|P_j|}{|Z|^m}.$$
  3. For $k = 1, \ldots, \tau$, consider the subsets of Z^m given by
  $$L^{(k)} = \Pi_S^{-1}\left(B_{\ell_1^m(Z)}(\Pi_S(x), k\Delta)\right),$$
  $$M^{(k)} = \Pi_S^{-1}\left(B_{\ell_1^m(Z)}(\Pi_S(x), (k+1)\Delta) \setminus L^{(k)}\right)$$
  $$R^{(k)} = \Pi_S^{-1}\left(Z^m \setminus (M^{(k)} \cup L^{(k)})\right).$$

4. If there exists $k_0 \in [\tau]$ such that

$$\rho_{P_j}(L^{(k_0)} \cup M^{(k_0)})^{1+\beta} + \rho_{P_j}(M^{(k_0)} \cup R^{(k_0)})^{1+\beta} \leq 1, \quad \text{and}$$

$$\rho_{P_j}(L^{(k_0)}) \leq \rho_{P_j}(R^{(k_0)}) \tag{5.22}$$

then, we set $L_j = L^{(k_0)} \cap (R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2})$, $R_j = R^{(k_0)} \cap (R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2})$ and

$$M_{j,1} = \Pi_S^{-1}\left(M^{(k)} \cap B_{\ell_1^m(Z)}(\Pi_S(x), (k+1/2)\Delta)\right) \cap (R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2}),$$
$$M_{j,2} = \Pi_S^{-1}\left(M^{(k)} \setminus M_{j,1}\right) \cap (R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2}).$$

Increment j, and let $P_{j+1} = P_j \cap R_j \cup M_{j,1} \cup M_{j,2}$.

5. Otherwise, if there is no $k_0 \in [\tau]$ satisfying (5.22), output "fail".

- If we have not yet output "fail", then output $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i\in[j]}$.

Lemma 5.1. *Consider $m \in \mathbb{N}$, a multi-set S of $[d]$, and parameters $\xi \in (0,1)$, $\beta \geq 0$, $\tau \in \mathbb{N}$, and $\Delta \in \mathbb{R}_{\geq 0}$. Let $P \subset Z^d$ be any subset of Z^d of size n, and let ρ_P be the empirical probability distribution of P. Suppose that* `BuildList`$(P, S, \beta, \xi, \tau, \Delta)$ *does not output "fail", then it outputs a list-decomposition $\mathcal{L} \in \mathsf{L}(\rho_P, |Z|^m, \beta, \xi)$.*

Proof. First, we notice that `BuildList`$(P, S, \beta, \xi, \tau, \Delta)$ outputs $\mathcal{L}$ when the inner loop terminates due to $|P_j| \leq (1-\xi)n$. In particular, until the last iteration of the loop, the algorithm generates a nested sequence $P = P_1 \supset P_2 \supset \cdots \supset P_j$ and $L_i \cup M_{i,1} \cup M_{i,2} \cup R_i \subset R_{i-1} \cup M_{i-1,1} \cup M_{i-1,2}$. Notice that for all $i \in [j]$,

$$\rho_P(L_i) = \frac{|P_i|}{n} \cdot \rho_{P_i}(L_i) \geq \frac{|P_i|}{n} \cdot \frac{n(1-\xi)}{|Z|^m \cdot |P_i|} \geq \frac{(1-\xi)}{|Z|^m},$$

which upper bounds $j \leq |Z|^m$. Note by (5.22), $\rho_{P_i}(L_i) \leq \rho_{P_i}(R_i)$. The fact that $\mathcal{L}$ has β-bounded overlap follows from (5.22), as well as an induction. In particular, we have

$$\begin{aligned}
&\rho_P(R_j \cup M_{j,1} \cup M_{j,2})^{1+\beta} + \sum_{i=1}^{j} \rho_P(L_i \cup M_{i,1} \cup M_{i,2})^{1+\beta} \\
&= \rho_P(R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2})^{1+\beta} \\
&\quad \big(\rho_{P_j}(R_j \cup M_{j,1} \cup M_{j,2})^{1+\beta} \\
&\quad + \rho_{P_j}(L_j \cup M_{j,1} \cup M_{j,2})^{1+\beta}\big) \\
&\quad + \sum_{i=1}^{j-1} \rho_P(L_i \cup M_{i,1} \cup M_{i,2})^{1+\beta} \\
&\leq \rho_P(R_{j-1} \cup M_{j-1,1} \cup M_{j-1,2})^{1+\beta} \\
&\quad + \sum_{i=1}^{j-1} \rho_P(L_i \cup M_{i,1} \cup M_{i,2})^{1+\beta}.
\end{aligned}$$

□

Lemma 5.2. *Let $m \in \mathbb{N}$, a multi-set S of $[d]$, parameters $\xi \in (0,1)$, $\beta \geq 0$, $\tau \in \mathbb{N}$, $\Delta \in \mathbb{R}_{\geq 0}$, a subset $P \subset Z^d$ of size n. If* `BuildList`$(P, S, \beta, \xi, \tau, \Delta)$ *outputs "fail", then there exists a subset $P_0 \subset P$ of size at least*

$$n(1-\xi) \cdot \min\left\{\exp\left(-\frac{m \ln |Z|}{(1+\beta)^\tau}\right), \frac{1}{2}\right\} \tag{5.23}$$

such that every $x, y \in P_0$ satisfies

$$\sum_{i\in S} d_Z(x_i, y_i) \leq 2(\tau+1) \cdot \Delta. \tag{5.24}$$

Proof. Suppose first that `BuildList`(P, S, β, ξ, τ) outputs "fail" because on the final iteration of the loop, we violate the second condition of (5.22), i.e., $\rho_{P_j}(L^{(k_0)}) \geq \rho_{P_j}(R^{(k_0)})$. Notice that this means $\rho_{P_j}(L^{(k_0)} \cup M^{(k_0)}) \geq 1/2$. Thus, let $P_0 = P_j \cap \left(L^{(k_0)} \cap M^{(k_0)}\right)$. Notice that $|P_0| \geq (1-\xi)n/2$, and therefore we satisfy (5.23). Finally, by virtue of being a ball in Z^m of radius at most $(k_0+1)\Delta \leq (\tau+1)\Delta$, satisfies (5.24).

On the other hand, suppose that `BuildList`(P, S, β, ξ, τ) outputs "fail" because the on final iteration of the loop, we violate the first condition of (5.22) always, i.e., for all $k_0 \in [\tau]$, $\rho_{P_j}(L^{(k_0)} \cup M^{(k_0)})^{1+\beta} + \rho_{P_j}(M^{(k_0)} \cup R^{(k_0)})^{1+\beta} \geq 1$. Then, since $L^{(k_0+1)} = L^{(k_0)} \cup M^{(k_0)}$, and $\rho_{P_j}(R^{(k_0)} \cup M^{(k_0)}) = 1 - \rho_{P_j}(L^{(k_0)})$, we have,

$$\begin{aligned}
1 &\leq \rho_{P_j}(L^{(k_0+1)})^{1+\beta} + \left(1 - \rho_{P_j}(L^{(k_0)})\right)^{1+\beta} \\
&\Rightarrow \rho_{P_j}(L^{(k_0+1)}) \\
&\geq \rho_{P_j}(L^{(k_0)})^{\frac{1}{(1+\beta)}}.
\end{aligned}$$

Therefore, $\rho_{P_j}(L^{(\tau)} \cup M^{(\tau)}) \geq \rho_{P_j}(L^{(1)})^{\frac{1}{(1+\beta)^\tau}}$, which implies (5.23). The bound (5.24) is immediate from the definition of $L^{(\tau)} \cup M^{(\tau)}$. □

LEMMA 5.3. *Let $m \in \mathbb{N}$, parameters $\xi, \eta, c \in (0,1)$, $\beta \geq 0$, and let $\tau \in \mathbb{N}$ and $\Delta \in \mathbb{R}_{\geq 0}$ be*

$$\tau = \left\lceil \log_{1+\beta}\left(\frac{16 \ln |Z|}{c^2}\right)\right\rceil,$$

$$\Delta = \frac{cmR}{4(\tau+1)} \qquad \text{and} \qquad \eta \leq \frac{1-\xi}{2|Z|^m}$$

Let $P \subset Z^d$ be a subset of n points. Suppose that any $A \subset Z^d$ where

$$\mathrm{diam}_{\ell_1^d(Z)}(A) \leq c \cdot d \cdot R,$$

must satisfy $|P \cap A| \leq \eta n$. Letting $\mathcal{S}_m$ be the uniform distribution over multi-sets $\mathbf{S}$ from $[d]$ of size m. Then,

$$\Pr_{\mathbf{S}\sim\mathcal{S}_m}[\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau, \Delta) \text{ outputs "fail"}] \leq \frac{4}{1-\xi} \cdot \exp\left(-\frac{c^2 m}{16}\right).$$

Proof. Consider the following random process:

1. Sample $\boldsymbol{x} \sim P$ uniformly, and then sample $\boldsymbol{y} \sim P$ conditioned on $d_{\ell_1^d(Z)}(\boldsymbol{x}, \boldsymbol{y}) \geq c \cdot d \cdot R/2$.
2. Sample $\mathbf{S} \sim \mathcal{S}_m$, which may be equivalently sampled as $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \sim [d]$ and writing $\mathbf{S} = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m\}$, output

$$d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(\boldsymbol{x}), \Pi_{\mathbf{S}}(\boldsymbol{y})).$$

On the one hand, we have

$$\Pr_{\substack{\boldsymbol{x},\boldsymbol{y}\sim P\\ \mathbf{S}\sim\mathcal{S}_m}}\left[d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(\boldsymbol{x}), \Pi_{\mathbf{S}}(\boldsymbol{y})) \leq 2(\tau+1)\Delta\right] \tag{5.25}$$

$$\geq \Pr_{\mathbf{S}\sim\mathcal{S}}[\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau) \text{ outputs "fail"}] \cdot \left((1-\xi)\cdot \min\left\{\exp\left(-\frac{m \ln |Z|}{(1+\beta)^\tau}\right), \frac{1}{2}\right\} - \eta\right)^2. \tag{5.26}$$

To justify (5.25), consider the event, when $\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau)$ outputs "fail," and by Lemma 5.2, let $P_0 \subset P$ be the low-diameter subset of sufficiently many points. Consider the event $\boldsymbol{x} \sim P_0$ and that $\boldsymbol{y} \sim P_0$ with distance at least $c \cdot d \cdot R/2$ from $\boldsymbol{x}$. Since $|P \cap B_{\ell_1^d(Z)}(\boldsymbol{x}, cdR/2)| \leq \eta n$, the number of points in P_0 with distance at least $cdR/2$ from $\boldsymbol{x}$ is at least $|P_0| - \eta n$. This shows (5.25).

On the other hand, since $2(\tau+1)\Delta \leq cmR/4$, and $d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(x), \Pi_{\mathbf{S}}(y))$ is a sum of i.i.d random variables bounded between 0 and R, we may apply McDiarmid's inequality to any two points $x, y \in P$ at distance at least $cdR/2$. Specifically, using the fact that $2(\tau+1)\Delta \leq cmR/4$, we have

$$\Pr_{\substack{\boldsymbol{x},\boldsymbol{y}\sim P\\ \mathbf{S}\sim\mathcal{S}\sim[d]}}\left[d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(\boldsymbol{x}), \Pi_{\mathbf{S}}(\boldsymbol{y})) \leq 2(\tau+1)\Delta\right]$$

$$\leq \exp\left(-\frac{2(cmR/4)^2}{mR^2}\right) = \exp\left(-\frac{c^2 \cdot m}{8}\right). \tag{5.27}$$

Combining (5.27) and (5.25), and minding the bounds on τ and η finishes the lemma. □

Proof. [Proof of Theorem 5.1] We let $\Delta = \frac{cmR}{4(\tau+1)}$ and $\xi = 1/16$, so that we may apply Lemma 5.3, and let $\mathcal{H}$ be the distribution over $\mathsf{L}(\rho_P, |Z|^m, \beta, \xi)$ given by letting $\mathcal{L} \sim \mathcal{H}$ be given by $\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau, \Delta)$ where $\mathbf{S} \sim \mathcal{S}_m$ conditioned on the procedure not outputting "fail." Notice that for these settings of parameters ξ, c and m,

$$\Pr_{\mathbf{S}\sim\mathcal{S}_m}[\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau, \Delta) \text{ outputs "fail"}] \leq \frac{4}{1-\xi} \cdot \exp\left(-\frac{c^2 m}{16}\right) \leq \frac{1}{2}.$$

For any $x, y \in Z^d$ satisfying (5.21), the list-decomposition $\mathcal{L} = \{(L_i, M_{i,1}, M_{i,2}, R_i)\}_{i\in[j]}$ generated from $\texttt{BuildList}(P, S, \beta, \xi, \tau, \Delta)$ splits (x, y) only if $d_{\ell_1^m(Z)}(\Pi_S(x), \Pi_S(y)) \geq \Delta/2$. In particular, consider $i \in [j]$, and let $M_{i,1}$ and $M_{i,2}$ be two shells around a point $z \in \ell_1^m(Z)$ of width $\Delta/2$. Then, if $x \in L_i$ and $y \in M_{i,2} \cup R_i$,

$$\begin{aligned} &d_{\ell_1^m(Z)}(\Pi_S(x), \Pi_S(y)) \\ &\geq \left|d_{\ell_1^m(Z)}(\Pi_S(y), z) - d_{\ell_1^m(Z)}(\Pi_S(x), z)\right| \\ &\geq \frac{\Delta}{2}, \end{aligned}$$

and a similar calculation shows $x \in R_i$ and $y \in L_i \cup M_{i,1}$ only when $d_{\ell_1^m(Z)}(\Pi_S(x), \Pi_S(y)) \geq \Delta/2$. Since

$$\frac{\Delta}{2} - \mathop{\mathbf{E}}_{\mathbf{S}\sim\mathcal{S}_m}\left[d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(x), \Pi_{\mathbf{S}}(y))\right] = \frac{\Delta}{2} - \frac{m}{d} \cdot d_{\ell_1^d(Z)}(x, y) \geq \frac{\Delta}{4},$$

and $d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(x), \Pi_{\mathbf{S}}(y))$ is a sum of m i.i.d random variables bounded between 0 and R, we apply McDiarmid's inequality to conclude

$$\begin{aligned} &\Pr_{\mathbf{S}\sim\mathcal{S}_m}\left[d_{\ell_1^m(Z)}(\Pi_{\mathbf{S}}(x), \Pi_{\mathbf{S}}(y)) \geq \frac{\Delta}{2}\right] \leq \\ &\exp\left(-\frac{\Delta^2}{8mR^2}\right) = \exp\left(-\frac{c^2 m}{128 \cdot (1+\tau)^2}\right) \\ &\leq \epsilon, \end{aligned}$$

for the setting of m. Notice that conditioning of $\texttt{BuildList}(P, \mathbf{S}, \beta, \xi, \tau, \Delta)$ not outputting "fail" increases the above probability by at most a factor of 2, which concludes the theorem. □

6 ANN for $\ell_1^d(Z)$ from Theorem 5.1

6.1 ANN for $\ell_1^d(Z)$ in the maximal scale

We give the main theorem for ANN over $\ell_1^d(Z)$ at the maximal scale. In particular, if the finite metric space (Z, d_Z) has diameter $R \in \mathbb{R}_{\geq 0}$, we give a data structure for (r, cr)-ANN problem over $\ell_1^d(Z)$ which achieves approximation $O(\log \log |Z|)$ in the case $cr = \Theta(dR)$ with query time $|Z|^{\mathrm{poly}(\log\log|Z|)} \cdot n^{o(1)}$ and space $O(d \log |Z| n^{1+\beta})$.

THEOREM 6.1. *Fix $n, d \in \mathbb{N}$, a finite metric space (Z, d_Z) of diameter $R \in \mathbb{R}_{\geq 0}$, as well as parameters $\beta, \varrho \geq 0$ and $c_0 \in (0, 1)$. Let*

$$c = 16\left(\left\lceil \log_{1+\beta}\left(\frac{16 \ln |Z|}{c_0^2}\right)\right\rceil + 1\right) + 1 \qquad \text{and}$$

$$r = \frac{c_0 \cdot d \cdot R}{c - 1}. \tag{6.28}$$

There exists a data structure for (r, cr)-ANN over $\ell_1^d(Z)$ using space $O(d \log |Z| n^{1+\beta})$ and query time $|Z|^{O(c^2 \ln(1/\varrho)/c_0^2)} \cdot \log n$ which succeeds with probability $n^{-\varrho}$.

Proof. The data structure proceeds similarly to that of [18] and recursively builds a randomized decision tree. We let $\gamma \in (0, 1)$ be a sufficiently small universal constant and $\epsilon = \gamma\varrho$, $\xi = 1/16$, $\tau \in \mathbb{N}$, $m \in \mathbb{N}$ according to (5.20), $\eta = 1/(4|Z|^m)$, and

$$\varrho = \frac{\log\left(\frac{1}{1-2\epsilon}\right)}{\log\left(\frac{1}{1-\xi}\right)}.$$

We check whether there exists a subset $A \subset Z^d$ satisfying

$$\mathrm{diam}_{\ell_1^d(Z)}(A) \leq (c - 1)r. \tag{6.29}$$

containing at least ηn points from P. Suppose that such a set exists. Then, let $x \in P \cap A$ and $P' = P \setminus A$. We store the point x and recursively build a data structure for the points in P'. On a query $q \in Z^m$, we first check whether $d_{\ell_1^d(Z)}(q, x) \leq cr$. If this is the case, we return x. Otherwise, we notice that (6.29) implies that if $p \in P$ satisfies $d_{\ell_1^d(Z)}(p, q) \leq r$, that $p \notin A$, so that $p \in P'$.

Suppose, on the other hand, that no such $A \subset Z^d$ exists. In this case, we apply Theorem 5.1 and obtain a distribution $\mathcal{H}$ supported on $\mathsf{L}(\rho_P, |Z|^m, \beta, 1/16)$ for which any two points within distance at most r are split with probability less than 2ϵ. We sample $\mathcal{L} \sim \mathcal{H}$ and use $\mathcal{L}$ to decompose the dataset into at most $|Z|^m$ parts, each of which contains less than $15n/16$ points. Then, we recursively build a data structure for points falling in each non-empty part.

By virtue of $\mathcal{L}$ having β-bounded overlap, and the fact that we recurse only on non-empty parts, an argument analogous to the width bound of Lemma 4.2 shows that the total space usage of the data structure is $O(d \log |Z| n^{1+\beta})$. We notice that on any execution of a query, the number of nodes where we find a low-diameter set of ηn points is at most $O(|Z|^m \log n)$ and processing each one takes $O(d)$ time. Furthermore, on any execution of a query, the number of nodes which apply a list decomposition is at most $\log_{1/(1-\xi)} n$, since on each such application, the number of points considered decreases by a factor of $1 - \xi$, and the total time needed to process each such application is at most $O(m|Z|^m)$, since these correspond to computing distances to points in Z^m. Thus, the time complexity of one execution of the data structure is $O(m|Z|^m \log n)$. For a query $q \in Z^d$ such that $p \in P$ satisfies $d_{\ell_1^d(Z)}(p, q) \leq r$, the probability of success of the data structure is at least the probability that we never sample $\mathcal{L} \sim \mathcal{H}$ which splits (p, q). This probability is at least

$$(1 - 2\epsilon)^{\log_{1/(1-\xi)} n} = n^{-\varrho}.$$

□

6.2 Densification for $\ell_1^d(Z)$

DEFINITION 6.1. *Let (Z, d_Z) be a finite metric space and $k \in \mathbb{N}$. We say the pair $(x, y) \in Z^k \times Z^k$ is an* edge *of Z^k if there exists a unique $i \in [k]$ where $x_i \neq y_i$; in this case, we say the edge of Z^k is in* direction i. *Notice that*

$$d_{\ell_1^k(Z)}(x, y) = d_Z(x_i, y_i).$$

LEMMA 6.2. *Let (Z, d_Z) be a finite metric space whose distances are between 1 and $R \in \mathbb{R}_{\geq 0}$. For $k \in \mathbb{N}$, there exists a finite metric space (Y, d_Y) with distances between 1 and R of size $|Y| \leq |Z|^{O(R)}$, as well as an embedding $f\colon Z^k \to Y$ such that every edge of Z^k, $(x, y) \in Z^k \times Z^k$, satisfies $d_{\ell_1^k(Z)}(x, y) = d_Y(f(x), f(y))$. Furthermore, given $x \in Z^k$, the embedding $f(x)$ proceeds by computing intersections of $O(R \log |Z|)$ many subsets of $|Z|$ with the set $\{x_1, \ldots, x_k\}$.*

The proof of Lemma 6.2 follows from the probabilistic method, where we describe a probability distribution over finite metric spaces $(\mathbf{Y}, d_{\mathbf{Y}})$ and embeddings $\boldsymbol{f}\colon Z^k \to \mathbf{Y}$ and show in Claims 6.3, 6.4, and 6.5, that the necessary guarantees are satisfied with non-zero probability. Consider a parameter $m \in \mathbb{N}$ which we will specify in Claim 6.5 to $O(R \log |Z|)$, and consider the following distribution $\mathcal{D}$ over metric spaces (Y, d_Y), whose object set is associated with $\{0,1\}^m$, and embeddings $f : Z^k \to Y$:

1. Independently sample a collection of m uniform sets $\mathcal{A} = (\mathbf{A}_1, \dots, \mathbf{A}_m)$, where each $\mathbf{A}_j \subset Z$. and let $G_{\mathcal{A}} = (\{0,1\}^m, E_{\mathcal{A}}, \boldsymbol{w}_{\mathcal{A}})$ be the weighted graph (parameterized by $\mathcal{A}$) which includes an edge $(a, b) \in E_{\mathcal{A}}$ if and only if there exists a unique pair of distinct points $x, y \in Z$ such that for all $j \in [m]$, $a_j \oplus b_j = |\mathbf{A}_j \cap \{x, y\}| \mod 2$, and if so, assigns the weight $w_{\mathcal{A}}(a, b) = d_Z(x, y)$.

2. We let the metric space $(Y_{\mathcal{A}}, d_{Y_{\mathcal{A}}})$ be the metric completion of the weighted graph $G_{\mathcal{A}} = (\{0,1\}^m, E_{\mathcal{A}}, w_{\mathcal{A}})$. In other words, for any two $a, b \in \{0,1\}^m$, $d_{Y_{\mathcal{A}}}(a, b)$ is the sum of the weights in the shortest path between a and b in $G_{\mathcal{A}}$ (if there exists a path between a and b), or D if a and b are disconnected.

3. We consider the $g_{\mathcal{A}} : Z^* \to \{0,1\}^m$, defined by letting, for each $z = (z_1, \dots, z_\ell) \in Z^\ell$ and $j \in [m]$

$$g_{\mathcal{A}}(z)_j = \sum_{i=1}^{\ell} |\mathbf{A}_j \cap \{z_i\}| \mod 2.$$

   The embedding $f_{\mathcal{A}} : Z^k \to \{0,1\}^m$ is given by the restriction of $\boldsymbol{g}$ to Z^k.

4. We output the metric space $(Y_{\mathcal{A}}, d_{Y_{\mathcal{A}}})$ and the embedding $f_{\mathcal{A}} : Z^k \to Y_{\mathcal{A}}$.

Claim 6.3. *Consider a fixed collection $\mathcal{A} = (A_1, \dots, A_m)$ for which every $1 \le \ell \le 2(R-1)$ and every $z \in Z^\ell$ where some $z_i \in Z$ appears an odd number of times satisfies $g_{\mathcal{A}}(z) \neq 0$. Then, for any edge of Z^k, $(x, y) \in Z^k \times Z^k$, $d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y)) \ge d_{\ell_1^k(Z)}(x, y)$.*

Proof. We first show that $d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y)) \ge d_{\ell_1^k(Z)}(x, y)$. Let $(x, y) \in Z^k \times Z^k$ be an edge of Z^k in direction i, such that letting $a = f_{\mathcal{A}}(x)$ and $b = f_{\mathcal{A}}(y)$, there is a lowest-weight path in $G_{\mathcal{A}}$ between a and b which we denote $c = (c_1, c_2, \dots, c_r) \in (\{0,1\}^m)^r$ where $c_1 = a$ and $c_r = b$. For each $t \in [r-1]$, let $(x^{(t)}, y^{(t)}) \in Z \times Z$ denote the unique distinct pair satisfying $(c_t)_j \oplus (c_{t+1})_j = |A_j \cap \{x^{(t)}, y^{(t)}\}| \mod 2$, and notice that the total weight of the path is $\sum_{t \in [r-1]} w_{\mathcal{A}}(x^{(t)}, y^{(t)})$, and that since c is the lowest-weight path, the weight is exactly $d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y))$. Suppose first that $r \ge R$, since the weight of every edge (c_t, c_{t+1}) is at least 1, the path would have weight at least $R \ge d_{\ell_1^k(Z)}(x, y)$. Hence, we may assume $r < R$. Consider the point $p \in Z^*$ iteratively defined as follows:

- Initially, we let $p_1 = (y_i, x_i) \in Z^2$.
- For $t \in [1, r-1]$, consider the edge $(c_t, c_{t+1}) \in \{0,1\}^m \times \{0,1\}^m$, and let p_{t+1} be the concatenation of p_t and $(x^{(t)}, y^{(t)})$.
- We let $p = p_r \in Z^{2r}$.

We claim that for all $j \in [m]$, we $g_{\mathcal{A}}(p)_j = 0$. The reason is that

$$\begin{aligned} g_{\mathcal{A}}(p)_j &= |A_j \cap \{x_i, y_i\}| \\ &\quad + \sum_{t=1}^{r-1} |A_j \cap \{x^{(t)}, y^{(t)}\}| \mod 2, \end{aligned} \tag{6.30}$$

$$= |A_j \cap \{x_i, y_i\}| + \sum_{t=1}^{r-1} (c_t)_j \oplus (c_{t+1})_j \mod 2, \tag{6.31}$$

by definition of $(x^{(t)}, y^{(t)})$, and since the sum in the right-hand side of (6.31) telescopes, $g_{\mathcal{A}}(p)_j = |A_j \cap \{x_i, y_i\}| + (c_1)_j + (c_r)_j \mod 2$. Notice that c_1 and c_r are exactly $g_{\mathcal{A}}(x)$ and $g_{\mathcal{A}}(y)$, respectively, so we may re-write

$$\begin{aligned} g_{\mathcal{A}}(p) &= |A_j \cap \{x_i, y_i\}| \\ &+ (g_{\mathcal{A}}(x))_j + (g_{\mathcal{A}}(y))_j \mod 2 \\ &= |A_j \cap \{x_i, y_i\}| \\ &+ \sum_{i_0=1}^{k} (|A_j \cap \{x_{i_0}\}| + |A_j \cap \{y_{i_0}\}|) \mod 2, \end{aligned}$$

and the above sum is 0 since $x_{i_0} = y_{i_0}$ for all $i \neq i_0$. Since $r < R$, $p \in Z^\ell$ for some $\ell \le 2(R-1)$ and since $g_{\mathcal{A}}(p) = 0$, we must have every $p_i \in Z$ appears an even number of times. As a result, the collection of edges $\{(x^{(t)}, y^{(t)})\}_{t \in [r-1]} \subset \binom{Z}{2}$ can be partitioned into a disjoint collection of cycles and a path from x_i to y_i.[9] Therefore,

$$\begin{aligned} & d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y)) \\ & = \sum_{t \in [r-1]} d_Z(x^{(t)}, y^{(t)}) \ge \\ & \sum_{\substack{t \in [r-1] \\ \text{form path } x_i \mapsto y_i}} d_Z(x^{(t)}, y^{(t)}) \\ & \ge d_Z(x_i, y_i). \end{aligned}$$

□

[9]Here, we view the path and cycles as a subset of the complete graph on Z.

CLAIM 6.4. *Consider a fixed collection $\mathcal{A} = (A_1, \dots, A_m)$ for which every $1 \le \ell \le 4$ and every $z \in Z^\ell$ where some $z_i \in Z$ appears an odd number of times satisfies $g_{\mathcal{A}}(z) \neq 0$. Then, for any edge of Z^k, $(x, y) \in Z^k \times Z^k$, $d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y)) \le d_{\ell_1^k(Z)}(x, y)$.*

Proof. In order to show $d_{Y_{\mathcal{A}}}(f_{\mathcal{A}}(x), f_{\mathcal{A}}(y)) \le d_Z(x_i, y_i)$, we first notice that the distinct pair $x_i, y_i \in Z$ satisfies $(f_{\mathcal{A}}(x))_j \oplus (f_{\mathcal{A}}(y))_j = |A_j \cap \{x_i, y_i\}| \mod 2$ for every $j \in [m]$. Thus, the only reason we would fail to satisfy the claim is if the pair $x_i, y_i \in Z$ was not unique. For the sake of contradiction, suppose that there was another distinct pair $u, v \in Z$ where $(f_{\mathcal{A}}(x))_j \oplus (f_{\mathcal{A}}(y))_j = |A_j \cap \{u, v\}| \mod 2$ for all $j \in [m]$, then, the string $(x_i, y_i, u, v) \in Z^4$, satisfies that every $j \in [m]$,

$$\begin{aligned}
&|A_j \cap \{x_i, y_i\}| \mod 2 \\
&= (f_{\mathcal{A}}(x))_j \oplus (f_{\mathcal{A}}(y))_j = |A_j \cap \{u, v\}| \mod 2 \\
&\Longrightarrow 0 = |A_j \cap \{u\}| + |A_j \cap \{v\}| \\
&+ |A_j \cap \{x_i\}| + |A_j \cap \{y_i\}| \mod 2
\end{aligned}$$

which implies that $g_{\mathcal{A}}(x_i, y_i, u, v) = 0$, but some element appears an odd number of times since the distinct pairs x_i, y_i and u, v are different; a contradiction. ∎

CLAIM 6.5. *There exists a large constant $C > 0$ such that for $m = C \cdot R \log_2(|Z|)$, with high probability over the draw of the collection $\mathcal{A} = (\mathbf{A}_1, \dots, \mathbf{A}_m)$, every $1 \le \ell \le \max\{2(R-1), 4\}$ and every $z \in Z^\ell$ for which some z_i appears an odd number of times satisfies $g_{\mathcal{A}}(z) \neq 0$.*

Proof. Consider a fixed $1 \le \ell \le \max\{2(D-1), 4\}$ and let $z \in Z^\ell$ which appears an odd number of times. Then, we have

$$\Pr_{\mathbf{A} \subset Z}\left[1 = \sum_{i=1}^{\ell} |\mathbf{A} \cap \{z_i\}| \mod 2\right] \ge \frac{1}{4}.$$

For some $u \in Z$, let $I_u \subset [\ell]$ be all indices where $z_i = u$, and suppose that $|I_u|$ is odd. Then, $u \in \mathbf{A}$ occurs with probability $1/2$, and once this happens, $\sum_{i \in I_u} |\mathbf{A} \cap \{z_i\}| \mod 2 = 1$. Therefore, $\sum_{i=1}^{\ell} |\mathbf{A} \cap \{z_i\}| = 0$ only when $\sum_{i \notin I_u} |\mathbf{A} \cap \{z_i\}| \mod 2 = 1$, which occurs with probability at most $1/2$. Since we repeat it m times, the probability that $g_{\mathcal{A}}(z) = 0$ is at most $(3/4)^m$. By the setting of m, we may take a union bound over all $1 \le \ell \le \max\{2(R-1), 4\}$ and $z \in Z^\ell$ for which an element appears an odd number of times. ∎

LEMMA 6.6. *Let (Z, d_Z) be any finite metric space whose distances are between 1 and $R \in \mathbb{R}_{\ge 0}$, $d \in \mathbb{N}$, and $r, \tau \ge 1$, and let*

$$k = \frac{d}{2 \cdot r \cdot \tau \cdot R}.$$

There exists a finite metric space (Y, d_Y) whose distances are between 1 and R, and with $|Y| \le |Z|^{O(R)}$, as well as

$$d_1 = O(\tau^2 R^4 \cdot d \log |Z|).$$

and a distribution over embeddings $\mathbf{F} \colon \ell_1^d(Z) \to \ell_1^{d_1}(Y)$, computable in $\mathrm{poly}(d, \tau, R, |Z|)$ time and space such that the following holds. For every $a, b \in Z^d$ where $d_{\ell_1^d(Z)}(a, b) \in [r, \tau r]$,

$$\frac{d_{\ell_1^{d_1}(Y)}(\mathbf{F}(a), \mathbf{F}(b))}{d_1} \ge \frac{1}{4} \cdot \frac{k}{d} \cdot d_{\ell_1^d(Z)}(a, b). \tag{6.32}$$

and for every $a, b \in Z^d$ with $d_{\ell_1^d(Z)}(a, b) \le r$,

$$\mathop{\mathbf{E}}_{\mathbf{F}}\left[\frac{d_{\ell_1^{d_1}(Y)}(\mathbf{F}(a), \mathbf{F}(b))}{d_1}\right] \le 4 \cdot \frac{k}{d} \cdot d_{\ell_1^d(Z)}(a, b). \tag{6.33}$$

Proof. We take the metric space (Y, d_Y) as well as the embedding $f \colon Z^k \to Y$ obtained from Lemma 6.2. Consider the following randomized procedure of building an embedding:

1. We sample $\mathcal{I} = (\boldsymbol{i}^{(1)}, \dots, \boldsymbol{i}^{(d_1)})$ uniformly from all multi-sets of k indices, where $\boldsymbol{i}^{(t)} = (\boldsymbol{i}_1^{(t)}, \dots, \boldsymbol{i}_k^{(t)})$ and $\boldsymbol{i}_j^{(t)} \in [d]$.
2. We let $F_{\mathcal{I}} \colon Z^d \to Y^{d_1}$ be the map given by concatenating $F_{\boldsymbol{i}^{(1)}}, \dots, F_{\boldsymbol{i}^{(t)}} \colon Z^k \to Y$, where
$$F_{\boldsymbol{i}^{(t)}}(a) = f(a_{\boldsymbol{i}_1^{(t)}}, \dots, a_{\boldsymbol{i}_k^{(t)}}) \in Y.$$

In order to see that the required time and space for the embedding $F_{\mathcal{I}}$ is $\mathrm{poly}(d, \tau, R, |Z|)$, notice that in order to compute the value of $F_{\mathcal{I}}(a)$, we need to evaluate f on d_1 many k-tuples of indices in $[d]$, and each evaluation of f simply checks intersections against d_1 sets in $|Z|$. Therefore, by the parameter settings of d_1 and k, the encoding of $F_{\mathcal{I}}$, as well as the time necessary to compute $F_{\mathcal{I}}(a)$ is at most $\mathrm{poly}(d, \tau, R, |Z|)$.

We consider a fixed pair $a, b \in Z^d$ and notice that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathcal{I}}\left[\frac{d_{\ell_1^{d_1}(Y)}(F_{\mathcal{I}}(a), F_{\mathcal{I}}(b))}{d_1}\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{i}}\left[d_Y(F_{\boldsymbol{i}}(a), F_{\boldsymbol{i}}(b))\right] = \mu(a, b).
\end{aligned} \tag{6.34}$$

where $\boldsymbol{i} = (\boldsymbol{i}_1, \dots, \boldsymbol{i}_k)$ is a k-tuple of random indices $\boldsymbol{i}_j \sim [d]$. Furthermore, the random variable $d_{\ell_1^{d_1}(Y)}(F_{\mathcal{I}}(a), F_{\mathcal{I}}(b))$ in the expectation on the right-hand side of (6.34) is a sum of independent random variables, each taking values between 1 and R. We show that if $d_{\ell_1^d(Z)}(a, b) \in [r, \tau r]$, then we have

$$\mu(a, b) \ge \frac{1}{2} \cdot \frac{k}{d} \cdot d_{\ell_1^d(Z)}(a, b).$$

If we have established this fact, we have, by McDiarmid's inequality

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\mathcal{I}}&\left[\frac{d_{\ell_1^{d_1}(Z)}(F_{\mathcal{I}}(a), F_{\mathcal{I}}(b))}{d_1} \leq \frac{\mu(a,b)}{2}\right] \\ &\leq \exp\left(-\frac{\mu(a,b)^2 d_1}{2R^2}\right) \\ &\leq \exp\left(-\frac{d_1}{32\cdot\tau^2\cdot R^4}\right)\end{aligned}$$

so that setting $d_1 = 128\tau^2 R^4 d \cdot \log|Z|$ and taking a union bound over at most $|Z|^{2d}$ pairs of points $a, b \in Z^d$ with $d_{\ell_1^d(Z)}(a,b) \in [r, \tau r]$, we have that with probability at least $1/2$ over the draw of $\mathcal{I}$, (6.32) holds. Therefore, if we considered sampling $\mathbf{F}$ according to the distribution over $\mathcal{I}$ conditioned on (6.32) holding, we have that for every $a, b \in Z^d$,

$$\mathop{\mathbf{E}}_{\mathbf{F}}\left[\frac{d_{\ell_1^{d_1}(Y)}(\mathbf{F}(a), \mathbf{F}(b))}{d_1}\right] \leq 2\mu(a,b),$$

because distances are always non-negative. It remains to upper and lower bound $\mu(a,b)$, which we do next.

Consider the indicator random variable $\mathbf{X}$ of the event defined over the randomness in a draw of $\boldsymbol{i} = (\boldsymbol{i}_1, \dots, \boldsymbol{i}_k)$ that there exists at most one index $j \in [k]$ where $a_{\boldsymbol{i}_j} \neq b_{\boldsymbol{i}_j}$. In this case, the pair $(a_{\boldsymbol{i}_1}, \dots, a_{\boldsymbol{i}_k})$ and $(b_{\boldsymbol{i}_1}, \dots, b_{\boldsymbol{i}_k})$ form an edge of Z^k in direction $\boldsymbol{i}_j$, and by Lemma 6.2,

$$d_Y(F_{\boldsymbol{i}}(a), F_{\boldsymbol{i}}(b)) = d_Z(a_{\boldsymbol{i}_j}, b_{\boldsymbol{i}_j}). \tag{6.35}$$

Notice that conditioned on there being a unique index $j \in [k]$ where $a_{\boldsymbol{i}_j} \neq b_{\boldsymbol{i}_j}$, the index $\boldsymbol{i}_j$ is uniformly distributed over indices $i \in [d]$ where $a_i \neq b_i$. Whenever $\mathbf{X}$ occurs, either $a_{\boldsymbol{i}_j} = b_{\boldsymbol{i}_j}$ and (6.35) is 0, or there exists a unique $j \in [k]$ where $a_{\boldsymbol{i}_j} \neq b_{\boldsymbol{i}_j}$, and (6.35) is non-zero. Letting γ be the probability, over the draw of $\boldsymbol{i}$, that $\mathbf{X} = 1$,

$$\begin{aligned}\gamma \cdot \frac{k}{d} \cdot d_{\ell_1^d(Z)}(a,b) &\leq \mu(a,b) \\ &\leq \frac{k}{d}\cdot d_{\ell_1^d(Z)}(a,b) + (1-\gamma)R,\end{aligned}$$

Letting $N = |\{i \in [d] : a_i \neq b_i\}|$, we have

$$1 - \gamma \leq \binom{k}{2}\left(\frac{N}{d}\right)^2 \leq \binom{k}{2}\left(\frac{d_{\ell_1^d(Z)}(a,b)}{d}\right)^2,$$

since each coordinate where $a_i \neq b_i$ contributes at least 1 to $d_{\ell_1^d(Z)}(a,b)$. Consider first the case $d_{\ell_1^d(Z)}(a,b) \in [r, \tau r]$, and notice that

$$\begin{aligned}\mu &\geq \left(1 - \binom{k}{2}\cdot\frac{r^2\tau^2}{d^2}\right)\cdot\frac{k}{d}\cdot d_{\ell_1^d(Z)}(a,b) \\ &\geq \frac{1}{2}\cdot\frac{k}{d}\cdot d_{\ell_1^d(Z)}(a,b).\end{aligned}$$

On the other hand, for every $a, b \in Z^d$ with $d_{\ell_1^d(Z)}(a,b) \leq r$ we have

$$\begin{aligned}\mu(a,b) &\leq \frac{k}{d}\cdot d_{\ell_1^d(Z)}(a,b)\left(1 + \frac{k}{2}\cdot\frac{r}{d}\cdot R\right) \\ &\leq 2\cdot\frac{k}{d}\cdot d_{\ell_1^d(Z)}(a,b).\end{aligned}$$

□

6.3 ANN for $\ell_1^d(Z)$ in all scales

LEMMA 6.7. *Fix $n, d \in \mathbb{N}$ and a finite metric space (Z, d_Z) whose distances are between 1 and R, as well as parameters $r \in \mathbb{R}_{\geq 0}$ and $\epsilon \in (0, 1/2)$. There exists a distribution $\mathcal{H}$ supported on functions $f \colon Z^d \to [n^2]$ computable in time and space* $\mathrm{poly}(d, |Z|)$ *such that for any dataset $P \subset Z^d$ of size n, the following holds. Let $q \in Z^d$ be any point, and suppose there exists $p \in P$ where $d_{\ell_1^d(Z)}(p, q) \leq r$, then with probability at least $n^{-\epsilon}/2$ over the choice of $\boldsymbol{f} \sim \mathcal{H}$,*

- *$\boldsymbol{f}(p) = \boldsymbol{f}(q)$, and*
- *Whenever $p' \in P$ satisfy $\boldsymbol{f}(p') = \boldsymbol{f}(q)$, then $d_{\ell_1^d(Z)}(q, p') \leq O(r \log|Z|/\epsilon)$.*

THEOREM 6.2. *Fix $n, d \in \mathbb{N}$, a finite metric space (Z, d_Z) of whose distances are between 1 and $R \in \mathbb{R}_{\geq 0}$, as well as parameters $\beta, \varrho \geq 0$, $\epsilon \in (0, 1/2)$, and $r \geq 1$. Let*

$$c = O\left(\left\lceil \log_{1+\beta}\left(\frac{R\log|Z|}{\epsilon}\right)\right\rceil + 1\right). \tag{6.36}$$

There exists a data structure for (r, cr)-ANN over $\ell_1^d(Z)$ using space $\mathrm{poly}(d, R, |Z|) \cdot n^{1+\beta}$ *and query time* $\mathrm{poly}(d, R) \cdot |Z|^{\mathrm{poly}(R, \log|Z|/\epsilon)\ln(1/\varrho)} \cdot \log n$ *which succeeds with probability $n^{-\varrho-\epsilon}/4$.*

References

[1] A. ANDONI, *Nearest Neighbor Search: the Old, the New, and the Impossible*, PhD thesis, MIT, 2009. Available at `http://www.mit.edu/~andoni/thesis/main.pdf`.

[2] A. ANDONI, D. CROITORU, AND M. PATRASCU, *Hardness of nearest neighbor under l-infinity*, in 2008 49th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2008, pp. 424–433.

[3] A. ANDONI AND P. INDYK, *Near-optimal hashing algorithms for near neighbor problem in high dimension*, Communications of the ACM, 51 (2008), pp. 117–122.

[4] ———, *Nearest neighbors in high-dimensional spaces*, in Handbook of Discrete and Computational Geometry, Chapman and Hall/CRC, 2017, pp. 1135–1155.

[5] A. Andoni, P. Indyk, and R. Krauthgamer, *Overcoming the ℓ_1 non-embeddability barrier: Algorithms for product metrics*, in Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA), 2009, pp. 865–874.

[6] A. Andoni, P. Indyk, and I. Razenshteyn, *Approximate nearest neighbor search in high dimensions*, (2018). This survey accompanied the ICM 2018 talk of Piotr Indyk. arXiv preprint arXiv:1806.09823.

[7] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten, *Hölder homeomorphisms and approximate nearest neighbors*, in 2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2018, pp. 159–169.

[8] A. Andoni, A. Naor, A. Nikolov, I. P. Razenshteyn, and E. Waingarten, *Data-dependent hashing via nonlinear spectral gaps*, in Proceedings of the Symposium on Theory of Computing (STOC), 2018.

[9] A. Andoni, H. L. Nguyen, A. Nikolov, I. Razenshteyn, and E. Waingarten, *Approximate near neighbors for general symmetric norms*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, 2017, pp. 902–913.

[10] A. Andoni, A. Nikolov, I. P. Razenshteyn, and E. Waingarten, *Approximate near neighbors for general symmetric norms*, in Proceedings of the Symposium on Theory of Computing (STOC), 2017. Previously in `http://arxiv.org/abs/1611.06222`.

[11] J. Bourgain, *On lipschitz embedding of finite metric spaces in hilbert space*, Israel Journal of Mathematics, 52 (1985), pp. 46–52.

[12] ———, *On Lipschitz embedding of finite metric spaces into Hilbert space*, Israel Journal of Mathematics, 52 (1985), pp. 46–52.

[13] T. Darrell, P. Indyk, and G. Shakhnarovich, *Nearest-neighbor Methods in Learning and Vision: Theory and Practice*, MIT Press, 2005.

[14] M. Farach-Colton and P. Indyk, *Approximate nearest neighbor algorithms for hausdorff metrics via embeddings*, Proceedings of the Symposium on Foundations of Computer Science (FOCS), (1999).

[15] N. C. Goonesekere and B. Lee, *Context-specific amino acid substitution matrices and their use in the detection of protein homologs*, Proteins: Structure, Function, and Bioinformatics, 71 (2008), pp. 910–919.

[16] S. Har-Peled, P. Indyk, and R. Motwani, *Approximate nearest neighbor: Towards removing the curse of dimensionality*, Theory of Computing, 8 (2012), pp. 321–350. Previously in FOCS'98.

[17] P. Indyk, *On approximate nearest neighbors in non-euclidean spaces*, in Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS '1998), 1998.

[18] ———, *On approximate nearest neighbors under ℓ_∞ norm*, Journal of Computer and System Sciences, 63 (2001), pp. 627–638.

[19] ———, *Approximate nearest neighbor algorithms for Frechet metric via product metrics*, Proceedings of the ACM Symposium on Computational Geometry (SoCG), (2002), pp. 102–106.

[20] ———, *Approximate nearest neighbor under edit distance via product metrics*, in Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA), 2004, pp. 646–650.

[21] P. Indyk and R. Motwani, *Approximate nearest neighbors: Towards removing the curse of dimensionality*, in Proceedings of the 30th ACM Symposium on the Theory of Computing (STOC '1998), 1998, pp. 604–613.

[22] M. Kapralov and R. Panigrahy, *NNS lower bounds via metric expansion for ℓ_∞ and EMD*, in Proceedings of the 39th International Colloquium on Automata, Languages and Programming (ICALP '2012), 2012, pp. 545–556.

[23] N. Linial, E. London, and Y. Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica, 15 (1995), pp. 215–245.

[24] J. Matoušek, *On embedding expanders into ℓ_p spaces*, Israel Journal of Mathematics, 102 (1997), pp. 189–197.

[25] M. Mendel and A. Naor, *Nonlinear spectral calculus and super-expanders*, Publications mathématiques de l'IHÉS, 119 (2014), pp. 1–95.

[26] A. Naor, *Comparison of metric spectral gaps*, Analysis and Geometry in Metric Spaces, 2 (2014).

[27] I. Razenshteyn, *High-dimensional similarity search and sketching: algorithms and hardness*, PhD thesis, Massachusetts Institute of Technology, 2017.