Near Neighbor Search via Efficient Average Distortion Embeddings

Deepanshu Kush
University of Toronto, Canada

Aleksandar Nikolov
University of Toronto, Canada

Haohua Tang
University of Toronto, Canada

Abstract

A recent series of papers by Andoni, Naor, Nikolov, Razenshteyn, and Waingarten (STOC 2018, FOCS 2018) has given approximate near neighbour search (NNS) data structures for a wide class of distance metrics, including all norms. In particular, these data structures achieve approximation on the order of p for ℓ_p^d norms with space complexity nearly linear in the dataset size n and polynomial in the dimension d, and query time sub-linear in n and polynomial in d. The main shortcoming is the *exponential in d pre-processing time* required for their construction.

In this paper, we describe a more direct framework for constructing NNS data structures for general norms. More specifically, we show via an algorithmic reduction that an efficient NNS data structure for a metric $\mathcal{M}$ is implied by an efficient average distortion embedding of $\mathcal{M}$ into ℓ_1 or the Euclidean space. In particular, the resulting data structures require only *polynomial pre-processing time*, as long as the embedding can be computed in polynomial time.

As a concrete instantiation of this framework, we give an NNS data structure for ℓ_p with *efficient pre-processing* that matches the approximation factor, space and query complexity of the aforementioned data structure of Andoni et al. On the way, we resolve a question of Naor (Analysis and Geometry in Metric Spaces, 2014) and provide an explicit, efficiently computable embedding of ℓ_p, for $p \geq 1$, into ℓ_1 with average distortion on the order of p. Furthermore, we also give data structures for Schatten-p spaces with improved space and query complexity, albeit still requiring exponential pre-processing when $p \geq 2$. We expect our approach to pave the way for constructing efficient NNS data structures for all norms.

2012 ACM Subject Classification Theory of computation → Nearest neighbor algorithms

Keywords and phrases Nearest neighbor search, metric space embeddings, average distortion embeddings, locality-sensitive hashing

Digital Object Identifier 10.4230/LIPIcs.SoCG.2021.50

Related Version *Full Version*: `https://arxiv.org/abs/2105.04712`

Funding *Aleksandar Nikolov*: Supported by an NSERC Discovery Grant (RGPIN-2016-06333), and a Tier II Canada Research Chair.
Haohua Tang: Supported by a Tier II Canada Research Chair.

Acknowledgements We want to thank Sushant Sachdeva for a fruitful discussion.

1 Introduction

In the nearest neighbor problem, a fundamental problem in computational geometry, we are given an n-point subset P of a metric space $\mathcal{M}$ with a distance function $\mathrm{d}_{\mathcal{M}}$, and our goal is to preprocess P into a data structure that, given a query point $q \in \mathcal{M}$, finds a point $x \in P$ minimizing $\mathrm{d}_{\mathcal{M}}(x, q)$. The main parameters of a nearest neighbor search data structure are

- the *pre-processing* time required to construct the data structure given P;
- the *space* taken up by the data structure, in words of memory;
- the *query time* required to answer a nearest neighbor query.

A trivial solution is to store P as a list of points and to answer queries by linear search. Ignoring the time required to compute distances, this solution takes $\Theta(n)$ space, but also requires $\Theta(n)$ query time, which is prohibitively large when we have a large data set and expect to answer many queries. However, in some cases it is possible to use the geometry of the metric to design data structures with much more efficient query procedures and nearly the same space requirements. For instance, Lipton and Tarjan [19] gave a data structure for the nearest neighbor problem in the 2-dimensional Euclidean plane with $O(n \log n)$ pre-processing, $O(n)$ space, and $O(\log n)$ query time. This result has been extended to d-dimensional Euclidean space (see e.g. [21]), and other d-dimensional normed spaces. There is, however, no known nearest neighbor data structure for the d-dimensional Euclidean space that achieves space that is polynomial in n and d, and query time that is polynomial in d, and sublinear in n (i.e., $O(\text{poly}(d) \cdot n^{1-\alpha})$ for some $\alpha > 0$).[1] There is, furthermore, some evidence that no such data structure exists [30].

The nearest neighbor search problem finds a multitude of applications beyond computational geometry, in areas as diverse as databases, computer vision, and machine learning. For example, it is used to find joinable tables in publicly available data [24]; for object recognition [22] and shape matching [10] in computer vision; to solve analogical reasoning tasks [23]; in machine learning, the k-Nearest Neighbors classifier is a common baseline. In these applications, often both the data set size n and the dimension d are large, making query times that are linear in n or exponential in d unacceptable. It makes sense then to relax this problem in the hope of allowing for efficient data structures in the high-dimensional regime. A common relaxation is to allow returning an approximate nearest neighbor to the query point q, i.e., a point $x \in P$ for which $\mathrm{d}_{\mathcal{M}}(x, q) \le c \min_{y \in P} \mathrm{d}_{\mathcal{M}}(y, q)$ for some approximation factor $c > 1$. A long and fruitful line of work, recently surveyed in [4], has shown that it is possible to construct data structures for this approximate nearest neighbor problem over certain spaces such as the d-dimensional Euclidean or Manhattan distance that use space $O(n^{1+\varepsilon} \cdot \text{poly}(d))$ and support queries in time $O(n^{\varepsilon} \cdot \text{poly}(d))$, for a constant $\varepsilon < 1$ that goes to 0 as the approximation factor c goes to infinity.

Rather than solving the nearest neighbor search problem directly, it is more convenient to fix a scale for the distance, and work with the (c, r)-near neighbor search ((c, r)-NNS) problem, defined below.

▶ **Definition 1.** *In the (c, r)-near neighbor search ((c, r)-NNS) problem, we are given a set of n points P in a metric space $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$, and are required to build a data structure so that given a query point $q \in \mathcal{M}$ with the guarantee that $\mathrm{d}_{\mathcal{M}}(x^*, q) \le r$ for some $x^* \in P$, we can use the data structure to output a point $x \in P$ satisfying $\mathrm{d}_{\mathcal{M}}(x, q) \le cr$ with probability at least $\frac{2}{3}$.*

It was shown by Indyk and Motwani [15] that the approximate nearest neighbor problem can be reduced to solving $\text{poly}(\log n)$ instances of the (c, r)-NNS problem. Therefore, we focus on the latter problem from this point onward.

Most (but not all) efficient data structures for the NNS problem in the high-dimensional regime are based on the idea of locality sensitive hashing (LSH), introduced by Indyk and Motwani [15]. A locality sensitive family of hash functions is a probability distribution $\mathcal{H}$ over random functions $h : \mathcal{M} \to \Omega$ such that pairs of close points are much more likely to be mapped by h to the same value than far points. In particular, pairs of points at distance at

[1] Here and in the rest of the paper, we use the notation $\text{poly}(A)$ to denote the class of polynomials in the expression A.

most r get mapped to the same bucket with probability at least p_1, while pairs of points at distance at least cr get mapped to the same bucket with probability at most p_2, with $p_1 > p_2$. Indyk and Motwani showed that an LSH family implies a data structure for the (c, r)-NNS problem with space $O(n^{1+\rho} \log_{1/p_2}(n))$, and query time $O(n^\rho \log_{1/p_2}(n))$, where $\rho = \frac{\log(1/p_1)}{\log(1/p_2)}$. Moreover, they constructed LSH families for the Hamming and Manhattan (i.e., ℓ_1) distance with $\rho \approx \frac{1}{c}$. Subsequent work also showed the existence of LSH families for the Euclidean distance (i.e., ℓ_2), as well as the ℓ_p metric for $1 \le p \le 2$, and improved the parameters [2, 12]. The LSH definition above has the property that the distribution $\mathcal{H}$ is independent of the dataset P. Sometimes, however, data structures with better trade-offs can be constructed by allowing $\mathcal{H}$ to depend on properties of P [3, 5, 8].

Until recently, relatively little was known about the NNS problem beyond the ℓ_p spaces for $1 \le p \le 2$, and the ℓ_∞^d space,[2] for which Indyk gave an efficient deterministic decision tree data structure with approximation $O(\log \log d)$ [14]. Data structures for other spaces can be constructed by reducing to these special cases via bi-Lipschitz embeddings. I.e., if for some metric $\mathcal{M}$ we can find an efficiently computable injection $f : \mathcal{M} \to \ell_2^d$ such that $\|f(x) - f(y)\|_2 \approx \mathrm{d}_{\mathcal{M}}(x, y)$ for all $x, y \in \mathcal{M}$, then we can use NNS data structures for ℓ_2^d to solve the NNS problem in $\mathcal{M}$. The best approximation factor achievable by this approach depends on the *distortion* of f, which measures how well $\|f(x) - f(y)\|_2$ approximates $\mathrm{d}_{\mathcal{M}}(x, y)$ in the worst case, and is defined as $\|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$, where

$$\|f\|_{\mathrm{Lip}} \coloneqq \sup_{x \ne y, x, y \in \mathcal{M}} \frac{\|f(x) - f(y)\|}{\mathrm{d}_{\mathcal{M}}(x, y)}, \qquad \|f^{-1}\|_{\mathrm{Lip}} \coloneqq \sup_{x \ne y, x, y \in \mathcal{M}} \frac{\mathrm{d}_{\mathcal{M}}(x, y)}{\|f(x) - f(y)\|_2}$$

are the Lipschitz constants of f and its inverse, respectively. Although this approach does yield some non-trivial results (see [4] for a survey), it only produces data structures with approximation $c \ge d^{\frac{1}{2} - \frac{1}{p}}$ even in the special case of ℓ_p^d (with $p > 2$), as this is the best possible distortion achievable by a bi-Lipszhitz embedding into ℓ_2 (see, e.g., [11]). It is natural to ask if the best approximation achievable by an efficient NNS data structure for a metric $\mathcal{M}$ is characterized by the optimal distortion of a bi-Lipschitz embedding into ℓ_2. More generally, a fundamental problem in high-dimensional computational geometry is to *determine the geometric properties of a metric space that allow efficient and accurate NNS data structures.*

A recent line of work showed that the answer to the first question above is negative, and there exist efficient NNS data structures with approximation much better than what is implied by bi-Lipszhitz embeddinds into ℓ_2^d or ℓ_1^d [6, 7]. These papers give data-dependent LSH families for any d-dimensional normed space with approximation factor that is sub-polynomial in the dimension d. A sample theorem is the following.

▶ **Theorem 2** ([6]). *For any $r > 0$, $p \ge 2$ and any $\varepsilon \in (0, 1)$, there is some $c \lesssim \frac{p}{\varepsilon}$ such that the following holds. For any set P of n points in $\mathbb{R}^d$ such that for all $x \in \mathbb{R}^d$ we have $|B_{\ell_p^d}(x, cr) \cap P| \le \frac{n}{2}$, there exists a probability distribution on axis-aligned boxes S satisfying*

$$\mathbb{P}\left[\frac{n}{4} \le |S \cap P| \le \frac{3n}{4}\right] = 1$$
$$\|x - y\|_p \le r \implies \mathbb{P}[|S \cap \{x, y\}| = 1] \le \varepsilon.$$

[2] Recall the ℓ_p^d norm on $\mathbb{R}^d$: $\|x\|_p \coloneqq \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$ for $1 \le p < \infty$, and $\|x\|_\infty \coloneqq \max_{i=1}^d |x_i|$.

Here, $B_{\ell_p^d}(x, cr) = \{y \in \mathbb{R}^d : \|y-x\|_p \le cr\}$ is the ℓ_p^d-ball of radius cr centered at x. Moreover, here and in the rest of the paper, the notation $A \lesssim B$ means that there exists an absolute constant C, independent of all other parameters, such that $A \le CB$.

Theorem 2 gives a form of LSH: we can think of points inside S as being mapped to 1, and points outside mapped to 0. We still have the p_1 condition: close points get mapped to the same value with probability at least $1 - \varepsilon$. Rather than guaranteeing that far points are less likely to be mapped to the same value, we have a data-dependent condition: if the data set contains no dense clusters of points, then most pairs of points are mapped to different values. It is not hard to construct a randomized decision tree using Theorem 2, and [6] showed how to use it to give a data structure for (c, r)-NNS over ℓ_p^d with approximation $c \lesssim \frac{p}{\varepsilon}$, space $O(n^{1+\varepsilon} \cdot \mathrm{poly}(d))$ and query time $O(n^\varepsilon \cdot \mathrm{poly}(d))$. Similar results were also proved for the Schatten-p norms, which extend the ℓ_p norms to matrices, and for arbitrary norms (with appropriate access to the norm ball) in [7].

Theorem 2, however, has a significant shortcoming: it does not guarantee that the distribution over axis-aligned boxes can be sampled efficiently given P as input. Indeed, the proof of the theorem in [6], as well as the proofs of other similar results in [6, 7], rely on a duality argument and yield sampling algorithms with running time exponential in the dimension. For this reason, the resulting data structures have pre-processing time that is also exponential in the dimension. These works thus raise an intriguing open problem: *can we sample a distribution such as the one in Theorem 2 in time polynomial in n and d?*

1.1 Our Results on Near Neighbor Search

In this work, we resolve the open problem above (also posed explicitly in [6]), and prove the following theorem.

▶ **Theorem 3.** *Let $\varepsilon \in (0, 1]$, $r > 0$, $p \ge 2$. For some $c \lesssim \frac{p}{\varepsilon}$, there exists a data structure for the (c, r)-NNS problem over n-point sets in ℓ_p^d with*

- *pre-processing time* $\mathrm{poly}(nd)$*;*
- *space* $O(n^{1+\varepsilon} \log(n) \cdot \mathrm{poly}(d))$*;*
- *query time* $O(n^\varepsilon \log(n) \cdot \mathrm{poly}(d))$*.*

We note that the only previous NNS data structure over ℓ_p^d (for $p > 2$) with pre-processing time $\mathrm{poly}(nd)$ could only achieve approximation on the order of 2^p, and used polynomial rather than nearly linear space [28, 9].

We further extend this result to the Schatten-p norms, which are a natural extension of ℓ_p^d to matrices. For a $d \times d$ symmetric real matrix X and $p \in [1, \infty]$, the Schatten-p norm $\|X\|_{C_p}$ of X is defined as the ℓ_p norm of the eigenvalues of X. In addition to their intrinsic interest [16, 17, 18, 27], the Schatten-p spaces are an interesting first step when extending geometric and analytic results from the ℓ_p spaces to more general norms: while Schatten-p shares many properties with ℓ_p, extending proofs and algorithms from ℓ_p to Schatten-p requires finding coordinate-free and often, more natural arguments. Here, we partially succeed in extending Theorem 3 to Schatten-p spaces, and show the following two theorems.

▶ **Theorem 4.** *Let $\varepsilon \in (0, 1]$, $r > 0$, $1 \le p \le 2$. For some $c \lesssim \frac{1}{\varepsilon^{2/p}}$, there exists a data structure for the (cr, r)-NNS problem over n-point sets of $d \times d$ symmetric matrices with respect to the Schatten-p norm with*

- *pre-processing time* $\mathrm{poly}(nd)$*;*
- *space* $O(n^{1+\varepsilon} \log(n) \cdot \mathrm{poly}(d))$*;*
- *query time* $O(n^\varepsilon \log(n) \cdot \mathrm{poly}(d))$*.*

The only previously known NNS data structures for Schatten-p with constant approximation and $\mathrm{poly}(nd)$ pre-processing time have polynomial rather than nearly linear space complexity. While not explicitly described there, such data structures follow from the techniques in [28, 9], in combination with the results in [29].

▶ **Theorem 5.** *Let $\varepsilon \in (0,1]$, $r > 0$, $p \geq 2$. For some $c \lesssim \frac{1}{\varepsilon}$, there exists a data structure for the (cr, r)-NNS problem over n-point sets of $d \times d$ symmetric matrices with respect to the Schatten-p norm with*

- *pre-processing time* $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(d)}$*;*
- *space* $O(n^{1+\varepsilon} \log(n) \cdot \mathrm{poly}(d))$*;*
- *query time* $O(n^{\varepsilon} \log(n) \cdot \mathrm{poly}(d))$*.*

In this theorem, the pre-processing time is exponential in the dimension as it is, as well, in [6]. Nevertheless, the data structure in Theorem 5 has the benefit that it has query time polynomial in d, rather than polynomial in d^p, as in the data structure in [6].

1.2 Techniques and Results on Average Distortion Embeddings

To prove Theorems 3, 4, and 5, we develop a new approach for proving partitioning statements such as Theorem 2 that relies on the notion of embeddings with average distortion, defined below (this definition is taken from [26]).

▶ **Definition 6.** *Given two metric spaces $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ and $(\mathcal{N}, \mathrm{d}_{\mathcal{N}})$, and an n-point set $P \subseteq \mathcal{M}$, we say a function $f : \mathcal{M} \to \mathcal{N}$ is an embedding of $\mathcal{M}$ into $\mathcal{N}$ with q-average distortion D (with respect to P) if*

$$\sum_{x \in P} \sum_{y \in P} \mathrm{d}_{\mathcal{N}}(f(x), f(y))^q \geq \frac{\|f\|_{\mathrm{Lip}}^q}{D^q} \sum_{x \in P}^{n} \sum_{y \in P}^{n} \mathrm{d}_{\mathcal{M}}(x, y)^q.$$

As before, $\|f\|_{\mathrm{Lip}}$ is the Lipschitz constant of f, i.e., $\|f\|_{\mathrm{Lip}} = \sup_{x \neq y, x, y \in \mathcal{M}} \frac{\mathrm{d}_{\mathcal{N}}(f(x), f(y))}{\mathrm{d}_{\mathcal{M}}(x,y)}$. When the embedding has 1-average distortion D, we simply say it has average distortion D. If for every integer n and any n point set in $\mathcal{M}$, there exists an embedding into $\mathcal{N}$ with average distortion D, then we say that $\mathcal{M}$ embeds into $\mathcal{N}$ with average distortion D.

Somewhat informally, we show that if a metric space $\mathcal{M}$ embeds into ℓ_1^d with average distortion D via an embedding f that can be efficiently computed from P, and efficiently evaluated, then $\mathcal{M}$ supports NNS data structures with approximation $c \lesssim \frac{D \log D}{\varepsilon}$, space $O(n^{1+\varepsilon})$, query time $O(n^{\varepsilon})$, and efficient pre-processing. (Precise statements follow from Lemmas 14 and 19, and Theorem 20 below.) This connection strengthens the reductions via bi-Lipschitz embeddings mentioned above. Moreover, this connection between NNS and average distortion embeddings is closely related to the connection between NNS and the cutting modulus from prior work [6]. In particular, bounds on the cutting modulus in [6] were proved by utilizing comparison inequalities between non-linear spectral gaps, in the sense of [25]. Such comparison inequalities were shown in [25] to be equivalent to the existence of average distortion embeddings. However, the connection between the cutting modulus and NNS data structures in [6] is not algorithmic: it involves a non-algorithmic duality argument that only yields data structures with exponential pre-processing, even when the cutting modulus bound is witnessed by an efficiently computable average distortion embedding. By contrast, our new connection between average distortion embeddings and NNS data structures is an efficient reduction: if the embedding is computationally efficient, so is the data structure, including the pre-processing.

To reduce constructing efficient NNS data structures to finding efficient average distortion embeddings, we first formalize the type of data dependent LSH family implicit in Theorem 2 and in other similar results. As mentioned above, these data-dependent LSH families relax the p_2 requirement of the standard LSH definition, by requiring that it holds empirically for the input point set P. I.e., we require that each hash function in the family maps at least $1 - p_2$ fraction of the pairs of points in P to different values (see Definition 9 for the precise requirement). As noted above, such a data-dependent LSH family is still sufficient to design an NNS data structure with similar running time and space guarantees as given by standard LSH (Lemma 14). Moreover, it is also not hard to construct a data-dependent LSH family using a standard LSH family when the point set P is dispersed, i.e., when no ball of radius cr contains more than half of P (Lemma 15).

So far these results just give a different perspective on standard NNS data structures using LSH. The benefit of using data-dependent LSH, however, is that the data-dependent requirement allows using a larger class of embeddings in reductions. While the existence of a standard LSH family for, e.g., ℓ_1^d, is inherited by all metrics that have a bi-Lipschitz embedding into ℓ_1^d with small distortion,[3] the existence of a data-dependent LSH family for ℓ_1^d is inherited by metrics $\mathcal{M}$ that have, for any dispersed point set $P \subseteq \mathcal{M}$, an embedding f into ℓ_1^d which (1) does not expand distances too much, and (2) does not map a dispersed point set P into a point set that is not dispersed. We formally define this class of embeddings, which we call embeddings with weak average distortion, in Definition 6 below. To finish the connection between NNS and average distortion embeddings, we prove that the existence of (computationally efficient) average distortion embeddings implies the existence of (computationally efficient) weak average distortion embeddings. The proof of this fact uses ideas previously used to relate embeddings with p- and q-average distortion (see Section 5.1 in [26]).

Finally, in order to utilize this general connection between average distortion embeddings and NNS data structures, we need to construct explicit, efficiently computable average distortion embeddings into ℓ_1^d or ℓ_2^d. Naor has shown that the existence of average distortion embeddings of a metric space $\mathcal{M}$ into ℓ_2 is equivalent to proving a certain inequality between non-linear spectral gaps, and, using this equivalence, he showed that when $p \geq 2$ ℓ_p^d embeds into ℓ_2 with average distortion $D \lesssim p$ [25, 26]. The connection between average distortion embeddings and spectral gap inequalities, however, uses a duality argument, and does not provide explicit, efficiently computable embeddings. In fact, an explicit construction of an embedding of ℓ_p^d into ℓ_2 is given as an open problem in [25]. Here we resolve (a variant of) this open problem. In the theorem below, the functions $M_{p,1}, \widetilde{M}_{p,1} : \ell_p^d \to \ell_1^d$ are defined by

$$M_{p,1}(x) = (\mathrm{sign}(x_i)|x_i|^p)_{i=1}^d \qquad \widetilde{M}_{p,1}(x) = \|x\|_p M_{p,1}\left(\frac{x}{\|x\|_p}\right) = \|x\|_p^{1-p} M_{p,1}(x),$$

with $\widetilde{M}_{p,1}(0) = 0$.

▶ **Theorem 7.** *For any $p \geq 1$, and any n-point set P in $\mathbb{R}^d$, for $t \in \mathbb{R}^d$ so that*

$$\forall i \in [d] : \quad |\{x \in P : x_i < t_i\}| = |\{x \in P : x_i > t_i\}|,$$

the map $g : \ell_p^d \to \ell_1^d$ defined by $g(x) = \widetilde{M}_{p,1}(x - t)$ has average distortion $D \lesssim p$.

We note that Naor's open problem was defined for embeddings into ℓ_2 and 2-average distortion. Our techniques can be extended to give similar results for such embeddings as well.

[3] In fact a weaker notion of randomized embedding suffices [4].

Above, $M_{p,1}$ is the classical Mazur map from ℓ_p^d to ℓ_1^d. This map sends the unit sphere in ℓ_p^d to the unit sphere in ℓ_1^d, and its restriction to the sphere has Lipschitz constant bounded by p up to absolute constants. The Mazur map was previously used by Matoušek to prove bounds on non-linear spectral gaps in ℓ_p^d [20]. As mentioned before, such bounds are closely related to the existence of average distortion embeddings, via Naor's duality argument in [25]. The Mazur map itself, however, cannot be used directly to give an average distortion embedding, since its Lipschitz constant is unbounded over all of ℓ_p^d. Our main technical contribution in the proof of Thereom 7 is the observation that the rescaled Mazur map $\widetilde{M}_{p,1}$ has Lipschitz constant $\lesssim p$ everywhere, and that the machinery of Matoušek's spectral gap argument can be then used to prove a bound on the average distortion directly, without going through a duality argument.

Our technique for constructing explicit average distortion embeddings in fact extends to every pair of normed spaces X and Y for which we have a Hölder-continuous homeomorphism f between the unit spheres of X and Y. We can then show that this homeomorphism can be extended to a function $\widetilde{f}$ which is Hölder-continuous on all of X, and that there is a shift $t \in X$ so that the map $g : X \to Y$ defined by $g(x) = \widetilde{f}(x - t)$ is an average distortion embedding of (a snowflake of) X into Y. One can then use a variety of known homemorphisms between spheres and construct reasonably explicit average distortion embeddings. We do so for the Schatten-p spaces, and one can also use the homeomorphism between finite dimensional normed spaces in [7] to give results for general normed spaces, too. Except for some special cases like ℓ_p^d and Schatten-p for $1 \le p \le 2$, however, one aspect of these embeddings is still not fully explicit, and in particular, not computationally efficient. Namely, the argument showing that there exists a good shift t, which was first given in [6], uses the theory of topological degree that is also used in textbook proofs of Brouwer's fixed point theorem, and does not suggest an efficient algorithm for computing t. We leave finding such an algorithm, even for the case of Schatten-p norms with $p \ge 2$, as an open problem.

2 Weak Average Distortion Embeddings Imply NNS

We are going to assume that the metric spaces $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ we deal with are endowed with a dimension $\dim(\mathcal{M})$, which we use to quantify running times of basic tasks, e.g., evaluating distances. We will mostly deal with metric spaces defined by a norm on $\mathbb{R}^d$, in which case $\dim(\mathcal{M}) = d$. We will assume that a point $x \in \mathcal{M}$ can be represented by $\mathrm{poly}(\dim(\mathcal{M}))$ bits, and that the distance $\mathrm{d}_{\mathcal{M}}(x, y)$ can be computed in time $\mathrm{poly}(\dim(\mathcal{M}))$, as well.

In this section we first introduce a formalization of the data-dependent LSH families we are going to use. We show how to use such LSH families to construct a data structure for the NNS problem, by generalizing the randomized decision tree data structure from [6]. Then we introduce the notion of weak average distortion embedding, and show that it can weak average distortion embeddings into ℓ_1^d or ℓ_2^d imply the existence of data-dependent LSH.

2.1 Data-dependent LSH Families Imply NNS

The following definition is standard.

▶ **Definition 8.** *Let $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ be a metric space and fix a scale $r > 0$, approximation factor $c > 1$, and range Ω. Then a probability distribution $\mathcal{H}$ over maps from $\mathcal{M}$ to Ω is called (r, cr, p_1, p_2)-sensitive if*

$$\mathrm{d}_{\mathcal{M}}(x, y) \le r \implies \mathbb{P}_{h\sim\mathcal{H}}[h(x) = h(y)] \ge p_1,$$
$$\mathrm{d}_{\mathcal{M}}(x, y) > cr \implies \mathbb{P}_{h\sim\mathcal{H}}[h(x) = h(y)] < p_2.$$

Our data structures are based on the following, in a sense, weaker definition which allows $\mathcal{H}$ to depend on the point set, and defines p_2 in terms of the how hash functions spread the points among the bins they are hashed to.

▶ **Definition 9.** *Let $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ be a metric space and fix a scale $r > 0$, and range Ω. Let $P \subseteq \mathcal{M}$ be set of cardinality $|P| = n$. Then a probability distribution $\mathcal{H}(P)$ over maps from $\mathcal{M}$ to Ω is called (r, p_1, p_2)-empirically sensitive for P if*

$$\mathrm{d}_{\mathcal{M}}(x, y) \leq r \implies \underset{h \sim \mathcal{H}}{\mathbb{P}}[h(x) = h(y)] \geq p_1,$$

$$\forall \omega \in \Omega, \forall h \in \mathrm{supp}(\mathcal{H}): \quad |\{x \in P : h(x) = \omega\}| \leq p_2 n.$$

We call families of hash functions $\mathcal{H}(P)$ as in Definition 9 *data-dependent locality sensitive hash functions*, because p_2 constrains the data-dependent distribution of the points in P among the bins defined by the hash function.

▶ **Definition 10.** *Let $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ be a metric space and fix a scale $t > 0$. A set P of n points in a metric space $\mathcal{M}$ is called (t, β)-dispersed if, for all $x \in \mathcal{M}$, $|P \cap B_{\mathcal{M}}(x, t)| \leq (1 - \beta)n$.*

The following is a straightforward observation.[4]

▶ **Lemma 11.** *Suppose for a set P of n points in a metric space $\mathcal{M}$, for every $x_0 \in P$ we have $|P \cap B_{\mathcal{M}}(x_0, 2t)| \leq (1 - \beta)n$. Then P is (t, β)-dispersed.*

▶ **Definition 12.** *For a metric space $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$, and a (multi-)set $P \subset \mathcal{M}$ of size n, we use the notation*

$$\Psi_{\mathcal{M}}(P, t) = \frac{|\{(x, y) \in P \times P : \mathrm{d}_{\mathcal{M}}(x, y) > t\}|}{n^2}.$$

The following lemma relates the notion of being (r, β)-dispersed and the function $\Psi_{\mathcal{M}}(P, t)$.

▶ **Lemma 13.** *Let $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ be a metric space, and let P be a set of n points in $\mathcal{M}$. Then, if P is (t, β)-dispersed, then $\Psi_{\mathcal{M}}(P, t) \geq \beta$. Conversely, P is (t, β)-dispersed for $\beta = \frac{1}{2}\Psi_{\mathcal{M}}(P, 2t)$.*

The next lemma shows that data-dependently LSH families imply the existence of efficient NNS data structures.

▶ **Lemma 14.** *Let $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ be a metric space and let $r > 0$, and $c > 1$. Suppose that for every $(cr, \frac{1}{2})$-dispersed m-point set $Q \subseteq \mathcal{M}$, there exists a $(r, p_1(Q), p_2(Q))$-empirically sensitive $\mathcal{H}(Q)$ such that $\frac{\log(1/p_1(Q))}{\log(1/p_2(Q))} \leq \rho$ where all $p_2(Q) \leq p_2$ for some $p_2 \in (0, 1)$. Define $b = \max(\frac{1}{2}, p_2)$. Further, suppose that any h in the support of $\mathcal{H}(Q)$ can be stored in space and evaluted in time polynomial in $\dim(\mathcal{M})$, that $h \sim \mathcal{H}(Q)$ can be sampled in time $T_s(m)$, for a non-decreasing function $T_s(m)$ and the range Ω of $\mathcal{H}(Q)$ has size at most $\exp(\dim(\mathcal{M}))$. Then there exists a data structure for the $(O(c), r)$-NNS problem over n-point sets in $\mathcal{M}$ with*
- *pre-processing time $O(\mathrm{poly}(n \log_{1/b}(n) \dim(\mathcal{M})) \cdot T_s(n))$;*
- *space $O(n^{1+\rho} \log_{1/b}(n) \cdot \mathrm{poly}(\dim(\mathcal{M})))$;*
- *query time $O(n^{\rho} \log_{1/b}(n) \cdot \mathrm{poly}(\dim(\mathcal{M})))$;*

The terminology used in its proof is largely adopted from [6]. The data structure is a randomized decision tree similar to the one in [6] with a couple of generalizations: we allow the hash function to split space into more than two parts, and, more importantly, we allow the parameters of the data-dependent LSH family to change from one node of the decision tree to the next.

[4] The proofs of the statements in this section can be found in the full version of this paper.

2.2 Data-dependent LSH Families from LSH Families

We show that the existence of an LSH family implies the existence of a data-dependent LSH family.

▶ **Lemma 15.** *Let $(\mathcal{M}, \mathrm{d}_\mathcal{M})$ be a metric space and let $r > 0$, and $c > 1$. Suppose there exists a (r, cr, p_1, p_2)-sensitive $\mathcal{H}$ where $p_2 \leq 1/2$. Let $P \subseteq \mathcal{M}$ be a (cr, β)-dispersed n-point set. Then, for*

$$p_2' = \sqrt{1 - \beta(1 - 2p_2)}$$

the event $\mathcal{E} = \{\max_{\omega \in \Omega} |\{x \in P : h(x) = \omega\}| \leq p_2' n\}$ occurs with probability at least $1/2$, and $\mathcal{H}$ conditioned on $\mathcal{E}$ is $(r, 2p_1 - 1, p_2')$-empirically sensitive for P.

We instantiate this lemma with the LSH for ℓ_1^d due to Indyk and Motwani [15] (see also [13]).

▶ **Lemma 16.** *For any $r > 0$, any $c > 1$, and any $\Delta \geq 1$, the space ℓ_1^d restricted to $[-\Delta, \Delta]^d$ has a (r, cr, p_1, p_2)-sensitive $\mathcal{H}$ with $p_1 = 1 - \frac{r}{2d\Delta}$ and $p_2 = 1 - \frac{cr}{2d\Delta}$. Moreover, $h \sim \mathcal{H}$ can be sampled and evaluated in constant time.*

Together with Lemma 15, Lemma 16 implies the following corollary.

▶ **Corollary 17.** *Let $r > 0$, $c > 6$, $\Delta \geq 1$, and let P be a (cr, β)-dispersed n-point set in ℓ_1^d restricted to $[-\Delta, \Delta]^d$. There exists a $(r, 1 - \frac{8}{c}, 1 - \frac{\beta}{4})$-empirically sensitive $\mathcal{H}_{\ell_1^d}(P)$ for P. Moreover, a function $h \sim \mathcal{H}(P)$ can be sampled in $O(n \operatorname{poly}(d\Delta/r))$ time, evaluated in $\operatorname{poly}(d\Delta/r)$ time, and stored using $\operatorname{poly}(d\Delta/r)$ bits.*

2.3 Weak Average Distortion Embeddings

Below is our definition of weak average distortion embedding. We will use such embeddings to construct data-dependent LSH families for spaces other than ℓ_1^d.

▶ **Definition 18.** *Let $(\mathcal{M}, \mathrm{d}_\mathcal{M})$ and $(\mathcal{N}, \mathrm{d}_\mathcal{N})$ be two metric spaces, and let $P \subseteq \mathcal{M}$ be an n-point set. A function $f : \mathcal{M} \to \mathcal{N}$ is an embedding with weak average distortion D with respect to P if we have*

$$\sup_{t \geq 0} t\Psi_\mathcal{N}(f(P), t) \geq \frac{\|f\|_{\mathrm{Lip}}}{D} \sup_{t \geq 0} t\Psi_\mathcal{M}(P, t).$$

The name "weak average distortion" originates from the fact that $\sup_{t \geq 0} t\Psi_\mathcal{M}(P, t)$ is the weak-L_1 norm of $\mathrm{d}_\mathcal{M}$ with respect to the uniform measure over $f(P) \times f(P)$. So, a weak average distortion embedding is required to not expand distances too much while also not decreasing the weak-L_1 norm of the pairwise distances. The analogous notion of q-average distortion, where we instead take the L_q norm of $\mathrm{d}_\mathcal{M}$ with respect to the same measure (see Definition 6), has been studied before. The definition of weak average distortion embedding appears to be new. It can be extended in the natural way to more general probability measures, but we will not pursue this here.

In this subsection, we show that a weak average distortion embeddings of $\mathcal{M}$ into ℓ_1^d imply, via Corollary 17, a data-dependent LSH family for $\mathcal{M}$.

▶ **Lemma 19.** *Let $r > 0$, $D \geq 1$, and $\Delta > 0$. Fix an approximation factor $c \geq 48D$. Suppose that P is a $(cr, \frac{1}{2})$-dispersed set of n points in a metric space $(\mathcal{M}, \mathrm{d}_\mathcal{M})$, and let Δ be the diameter of P. If $f : \mathcal{M} \to \ell_1^d$ (or $f : \mathcal{M} \to \ell_2^d$) is an embedding with weak average distortion D with respect to P, then there exists a (r, p_1, p_2)-empirically sensitive $\mathcal{H}(P)$ for P with $\frac{\log(1/p_1)}{\log(1/p_2)} \lesssim \frac{D}{c}$ and $p_2 \geq 1 - \frac{cr}{16D\Delta}$.*

Moreover, assume that $d \in \mathrm{poly}(\dim(\mathcal{M}))$*, and that* f *can be computed from* P *in time* T*, and then stored in* $\mathrm{poly}(\dim(\mathcal{M}))$ *bits, and evaluated in* $\mathrm{poly}(\dim(\mathcal{M}))$ *time. Then a function* $h \sim \mathcal{H}(P)$ *can be sampled in time* $O(T + \mathrm{poly}(n \dim(\mathcal{M})\Delta/r))$*, evaluated in* $\mathrm{poly}(\dim(\mathcal{M})\Delta/r)$ *time, and stored using using* $\mathrm{poly}(\dim(\mathcal{M})\Delta/r)$ *bits.*

3 From Average to Weak Average Distortion Embeddings

We show the following theorem connecting average and weak average distortion embeddings.

▶ **Theorem 20.** *Suppose that the metric space* $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ *embeds with into a Banach space* $(X, \|\cdot\|)$ *with average distortion* D*. Then, for any* n*-point set* $P \subseteq \mathcal{M}$*, there exists an embedding with weak average distortion* $D' \lesssim D(1 + \log D)$*.*

Moreover, assume that for any point set Q *of* $m \leq n$ *points in* $\mathcal{M}$*, an embedding into* X *with average distortion* D *with respect to* Q *can be can be computed from* Q *in time* T*, and then stored in* $\mathrm{poly}(\dim(\mathcal{M}))$ *bits, and evaluated in* $\mathrm{poly}(\dim(\mathcal{M}))$ *time. Then, for any* n*-point set* $P \subseteq \mathcal{M}$*, the embedding into* X *with weak average distortion* D' *can be computed in time* $\mathrm{poly}(T + n\dim(\mathcal{M}))$*, stored in* $\mathrm{poly}(\dim(\mathcal{M}))$ *bits, and evaluated in time* $\mathrm{poly}(\dim(\mathcal{M}))$*.*

The proof of Theorem 20 is inspired by arguments relating p-average and q-average distortion embeddings for different p and q in [26], and is discussed in complete detail in the full version of this paper.

4 Efficient Average Distortion Embeddings

4.1 Average Distortion Embeddings from Bi-Hölder Homeomorphisms

We first give a general construction of embeddings with bounded average distortion using homeomorphisms between spheres of normed spaces. The main theorems mentioned in the Introduction are then proved using this general construction and the classical and non-commutative Mazur maps as the homeomorphisms. Proofs and some of the additional theorems from this section can be found in the full version of this paper. For a Banach space $(X, \|\cdot\|)$, we will use the notation $S_X = \{x \in X : \|x\| = 1\}$.

We begin by showing that homeomorphisms between spheres can be radially extended to the entire normed spaces while retaining their continuity properties. The lemma below was also shown in [7] but with worse constants.

▶ **Lemma 21.** *Let* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *be Banach spaces, and let* $\alpha, \beta \in (0, 1]$*. Let* $f : S_X \to S_Y$ *be a function that, for any* $x, y \in S_X$ *satisfies*

$$\frac{1}{L}\|x - y\|_X^{1/\beta} \leq \|f(x) - f(y)\|_Y \leq K\|x - y\|_X^{\alpha}.$$

Then the function $\widetilde{f} : X \to Y$ *defined by* $\widetilde{f}(x) = \|x\|_X^{\alpha} f\left(\frac{x}{\|x\|_X}\right)$*, and* $\widetilde{f}(0) = 0$ *satisfies the following for any* $x, y \in X$*:*

$$\|\widetilde{f}(x) - \widetilde{f}(y)\|_Y \leq (1 + 2^{\alpha}K)\|x - y\|_X^{\alpha} \tag{1}$$

$$\|\widetilde{f}^{-1}(x) - \widetilde{f}^{-1}(y)\|_X \leq \left(\frac{1}{\alpha\beta} + 2^{\beta}L^{\beta}\right)\|x - y\|_Y^{\beta} \max\{\|x\|_Y, \|y\|_Y\}^{\frac{1}{\alpha} - \beta} \tag{2}$$

Moreover, $\|\widetilde{f}(x)\|_Y = \|x\|_X^{\alpha}$ *for all* $x \in X$*.*

Recall that a median of a set of n points P in a metric space $\mathcal{M}$ is any point $y \in \mathcal{M}$ that minimizes $\frac{1}{n}\sum_{x\in P} \mathrm{d}_{\mathcal{M}}(x,y)$. The next definition is an approximate version of the median.

▶ **Definition 22.** *We say that a point y in a metric space $\mathcal{M}$ is a (C,ε)-approximate median of a finite point set $P \subseteq \mathcal{M}$ if*

$$\frac{1}{n}\sum_{x\in P} \mathrm{d}_{\mathcal{M}}(x,y) \le C \min_{z\in\mathcal{M}} \frac{1}{n}\sum_{x\in P} \mathrm{d}_{\mathcal{M}}(x,z) + \varepsilon.$$

The next lemma is our main tool for constructing explicit average distortion embeddings. Recall that, for $\alpha \in (0,1]$ the α-snowflake of a metric space $(\mathcal{M}, \mathrm{d}_{\mathcal{M}})$ is the metric space $\mathcal{M}^\alpha$ on the same ground set, with distance function $\mathrm{d}_{\mathcal{M}}(x,y)^\alpha$.

▶ **Lemma 23.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces, and let $\alpha \in (0,1]$. Let $f : S_X \to S_Y$ be a function that, for any $x, y \in S_X$ satisfies*

$$\|f(x) - f(y)\|_Y \le K\|x-y\|_X^\alpha,$$

and let $\widetilde{f}(x) = \|x\|_X^\alpha f\left(\frac{x}{\|x\|_X}\right)$, $\widetilde{f}(0) = 0$. Let $P \subseteq X$ be an n point set, let $t \in X$, and define $g : X^\alpha \to Y$ by $g(x) = \widetilde{f}(x-t)$. Suppose that one of the following conditions is satisfied for some $C \ge 1$ and $\varepsilon \le \frac{1}{4n^2}\sum_{x\in P}\sum_{y\in P}\|x-y\|_X^\alpha$

1. *0 is a (C,ε)-approximate median for $g(P)$;*
2. *$\left\|\frac{1}{n}\sum_{x\in P} g(x)\right\|_Y \le \varepsilon$.*

Then g is an embedding of the α-snowflake X^α into Y with average distortion at most D with respect to P, where $D \lesssim C(1+K)$ if the first condition is satisfied, and $D \lesssim 1+K$ if the second condition is satisfied.

Its proof is similar in spirit to bounds on non-linear Rayleigh quotients proved in [7], and is discussed in the full version. In order to use Lemma 23, we need to find some $t \in X$ that satisfies one of the two assumptions in the lemma. A general method for establishing the existence of such a t was proposed in [6, 7, 26], and relies on the following lemma. For a proof, see Lemma 45 from [26].

▶ **Lemma 24.** *For any finite dimensional Banach space $(X, \|\cdot\|)$, and any continuous function $h : X \to X$ such that*

$$\lim_{M\to\infty} \inf_{t:\|t\|\ge M} (\|t\| - \|h(t) - t\|) = \infty,$$

we have that h is surjective.

Using Lemma 24, we can show that there exists a t so that 0 is the mean of $g(P)$. The argument is essentially identical to arguments in [6, 7], but, since the result was not stated in the general form given below, we include a proof of the following lemma in the full version.

▶ **Lemma 25.** *Under the assumptions and notation of Lemma 21, there exists some $t \in X$ such that $\frac{1}{n}\sum_{x\in P} g(x) = 0$. Moreover, any such t must satisfy $\|t\|_X \le \left(\frac{M}{n}\sum_{x\in P}\|x\|_X^\alpha\right)^{\frac{1}{\alpha}}$, for a constant M that only depends on K, L, α, β.*

While Lemma 25 is very general, it does not readily give rise to an efficient algorithm to find t. The proof of Lemma 24 in [6, 26] is existential, and relies on a topological degree argument of the type used to prove Brouwer's fixed point theorem. Identifying general cases in which we can give an algorithmic proof of Lemma 25 is an interesting open problem. In the full version of this paper, we instantiate Lemma 23 with the Mazur map (for embedding ℓ_p^d) and the non-commutative Mazur map (for embedding Schatten-p), and give alternative algorithmic methods for finding a good center t in these special cases. This leads us to the proofs of Theorems 3, 4, and 5 stated in the Introduction.

5 Conclusion and Open Problems

We have constructed data structures for the (c, r)-NNS problem with efficient pre-processing, nearly linear space, and sub-linear query time with approximation $c \lesssim p$ in the case of ℓ_p spaces for all $p \geq 1$, and with $c \lesssim 1$ for Schatten-p spaces for $1 \leq p \leq 2$. Furthermore, we have laid out a general framework for producing such efficient data structures for general metrics: as long as there are (computationally efficient) average distortion embeddings of such metrics into ℓ_1^d or ℓ_2^d, we can produce efficient NNS data structures. This framework is an analogue of the cutting modulus framework from [6], but allows efficient pre-processing.

This connection between NNS data structures and low average distortion embeddings naturally warrants further research into constructing such computationally efficient embeddings of arbitrary metric spaces into ℓ_1^d or ℓ_2^d. A natural first step is to do this for Schatten-p spaces where $p > 2$. As noted earlier, the bottleneck in our construction is the design of an efficient algorithm for computing a center T satisfying the conditions of Lemma 23 for embedding Schatten-p into Schatten-2.

▶ **Problem 1.** *Given a dataset P of n $d \times d$ symmetric matrices in Schatten-p, find a matrix T in $\mathrm{poly}(n, d, 1/\varepsilon)$ time such that either 0 is a (C, ε)-median of $\widetilde{M}_{p,2}(T - P)$, or $\left\| \frac{1}{n} \sum_{X \in P} \widetilde{M}_{p,2}(T - X) \right\|_{C_2} \leq \varepsilon$.*

In our quest to construct efficient NNS data structures for arbitrary finite-dimensional norms, a slightly more ambitious goal is to make the main result of [26] algorithmic, as follows.

▶ **Problem 2.** *Given an n-point dataset P in a d-dimensional normed space $(\mathcal{M}, \|\cdot\|)$, construct an embedding $f : \mathcal{M}^{\frac{1}{2}} \to \ell_2^d$ with 2-average distortion $\lesssim \sqrt{\log d}$ with respect to P such that f can be computed in time $\mathrm{poly}(nd)$, stored in poly(d) bits, and evaluated in time $\mathrm{poly}(d)$.*

A solution to this problem will imply NNS data structures for any d-dimensional norm with polynomial time pre-processing, nearly linear space, sub-linear query time, and approximation poly-logarithmic in the dimension, solving also an open problem in [6]. Note that such data structures are not known even with exponential pre-processing, but it was shown in [6] that they do exist in the cell-probe model.

Finally, on a somewhat different note, it would also be very interesting to further optimize the approximation factor c of our NNS data structures, even in the special case of ℓ_p spaces.

▶ **Problem 3.** *Establish Theorem 3 with $c \lesssim \frac{\log p}{\varepsilon}$.*

A solution to this problem would interpolate between data structures for ℓ_1^d and ℓ_2^d, where constant approximation is possible, and Indyk's data structure for ℓ_∞^d which guarantees an $O(\log \log d)$ approximation [14], and is optimal in several natural models [1].

References

1 Alexandr Andoni, Dorian Croitoru, and Mihai Patrascu. Hardness of Nearest Neighbor under L-infinity. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '2008)*, pages 424–433, 2008.

2 Alexandr Andoni and Piotr Indyk. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS '2006)*, pages 459–468, 2006.

3 Alexandr Andoni, Piotr Indyk, Huy L. Nguyen, and Ilya Razenshteyn. Beyond Locality-Sensitive Hashing. In *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA '2014)*, pages 1018–1028, 2014. Available as arXiv:1306.1547.

4 Alexandr Andoni, Piotr Indyk, and Ilya Razenshteyn. Approximate nearest neighbor search in high dimensions. In *Proceedings of ICM 2018 (to appear)*, 2018.

5 Alexandr Andoni, Thijs Laarhoven, Ilya Razenshteyn, and Erik Waingarten. Optimal Hashing-based Time–Space Trade-offs for Approximate Near Neighbors. In *Proceedings of the 28th ACM-SIAM Symposium on Discrete Algorithms (SODA '2017)*, pages 47–66, 2017. Available as arXiv:1608.03580.

6 Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya P. Razenshteyn, and Erik Waingarten. Data-dependent hashing via nonlinear spectral gaps. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 787–800. ACM, 2018. doi:10.1145/3188745.3188846.

7 Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya P. Razenshteyn, and Erik Waingarten. Hölder homeomorphisms and approximate nearest neighbors. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 159–169. IEEE Computer Society, 2018. doi:10.1109/FOCS.2018.00024.

8 Alexandr Andoni and Ilya Razenshteyn. Optimal Data-Dependent Hashing for Approximate Near Neighbors. In *Proceedings of the 47th ACM Symposium on the Theory of Computing (STOC '2015)*, pages 793–801, 2015. Available as arXiv:1501.01062.

9 Yair Bartal and Lee-Ad Gottlieb. Approximate nearest neighbor search for ℓp-spaces (2<p<∞). *Theor. Comput. Sci.*, 757:27–35, 2019. doi:10.1016/j.tcs.2018.07.011.

10 S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002. doi:10.1109/34.993558.

11 Yoav Benyamini and Joram Lindenstrauss. *Geometric Nonlinear Functional Analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

12 Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S. Mirrokni. Locality-Sensitive Hashing Scheme Based on p-Stable Distributions. In *Proceedings of the 20th ACM Symposium on Computational Geometry (SoCG '2004)*, pages 253–262, 2004.

13 Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Similarity Search in High Dimensions via Hashing. In *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB '1999)*, pages 518–529, 1999.

14 Piotr Indyk. On Approximate Nearest Neighbors under ℓ_∞ Norm. *Journal of Computer and System Sciences*, 63(4):627–638, 2001.

15 Piotr Indyk and Rajeev Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *Proceedings of the 30th ACM Symposium on the Theory of Computing (STOC '1998)*, pages 604–613, 1998.

16 Yi Li, Huy L. Nguyên, and David P. Woodruff. On Sketching Matrix Norms and the Top Singular Vector. In *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA '2014)*, pages 1562–1581, 2014.

17 Yi Li and David P. Woodruff. On Approximating Functions of the Singular Values in a Stream. In *Proceedings of the 48th ACM Symposium on the Theory of Computing (STOC '2016)*, pages 726–739, 2016.

18 Yi Li and David P. Woodruff. Embeddings of Schatten Norms with Applications to Data Streams. In *Proceedings of the 44th International Colloquium on Automata, Languages and Programming (ICALP '2017)*, pages 60:1–60:14, 2017.

19 Richard J. Lipton and Robert Endre Tarjan. Applications of a planar separator theorem. *SIAM J. Comput.*, 9(3):615–627, 1980. doi:10.1137/0209046.

20 Jiří Matoušek. On Embedding Expanders into ℓ_p Spaces. *Israel Journal of Mathematics*, 102:189–197, 1997.

21 Stefan Meiser. Point location in arrangements of hyperplanes. *Inf. Comput.*, 106(2):286–303, 1993. doi:10.1006/inco.1993.1057.

22 Bartlett W. Mel. Seemore: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9(4):777–804, 1997. doi:10.1162/neco.1997.9.4.777.

23 Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119, 2013.

24 Renée J. Miller, Fatemeh Nargesian, Erkang Zhu, Christina Christodoulakis, Ken Q. Pu, and Periklis Andritsos. Making open data transparent: Data discovery on open data. *IEEE Data Eng. Bull.*, 41(2):59–70, 2018. URL: http://sites.computer.org/debull/A18june/p59.pdf.

25 Assaf Naor. Comparison of Metric Spectral Gaps. *Analysis and Geometry in Metric Spaces*, 2:1–52, 2014.

26 Assaf Naor. An average John theorem. *arXiv preprint*, 2019. arXiv:1905.01280.

27 Assaf Naor, Gilles Pisier, and Gideon Schechtman. Impossibility of Dimension Reduction in the Nuclear Norm. In *Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms (SODA '2018)*, 2018.

28 Assaf Naor and Yuval Rabani. On Approximate Nearest Neighbor Search in ℓ_p, $p > 2$. Manuscript, available on request, 2006.

29 Éric Ricard. Hölder Estimates for the Noncommutative Mazur Map. *Archiv der Mathematik*, 104(1):37–45, 2015.

30 Ryan Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theor. Comput. Sci.*, 348(2-3):357–365, 2005. doi:10.1016/j.tcs.2005.09.023.