2018 IEEE 59th Annual Symposium on Foundations of Computer Science

Hölder Homeomorphisms and Approximate Nearest Neighbors

Alexandr Andoni*, Assaf Naor†, Aleksandar Nikolov‡, Ilya Razenshteyn§ and Erik Waingarten*
* *Columbia University, New York, NY*
Email: {andoni, eaw}@cs.columbia.edu
† *Princeton University, Princeton, NJ*
Email: naor@math.princeton.edu
‡ *University of Toronto, Toronto, Ontario, Canada*
Email: anikolov@cs.toronto.edu
§ *Microsoft Research, Redmond, WA*
Email: ilyaraz@microsoft.com

***Abstract*—We study bi-Hölder homeomorphisms between the unit spheres of finite-dimensional normed spaces and use them to obtain better data structures for high-dimensional Approximate Near Neighbor search (ANN) in general normed spaces.**

Our main structural result is a finite-dimensional quantitative version of the following theorem of Daher (1993) and Kalton (unpublished). Every d-dimensional normed space X admits a small perturbation Y such that there is a bi-Hölder homeomorphism with good parameters between the unit spheres of Y and Z, where Z is a space that is close to ℓ_2^d. Furthermore, the bulk of this article is devoted to obtaining an algorithm to compute the above homeomorphism in time polynomial in d. Along the way, we show how to compute efficiently the norm of a given vector in a space obtained by the *complex interpolation* between two normed spaces.

We demonstrate that, despite being much weaker than bi-Lipschitz embeddings, such homeomorphisms can be efficiently utilized for the ANN problem. Specifically, we give two new data structures for ANN over a general d-dimensional normed space, which for the first time achieve approximation $d^{o(1)}$, thus improving upon the previous general bound $O(\sqrt{d})$ that is directly implied by John's theorem.

***Keywords*-near neighbor search; complex interpolation; John's theorem**

I. Introduction

Fix $d \in \mathbb{N}$. Below, the unit ball and unit sphere of a (complex[1]) normed space $X = (\mathbb{C}^d, \|\cdot\|_X)$ are denoted $B_X = \{x \in \mathbb{C}^d : \|x\|_X \leqslant 1\}$ and $S_X = \{x \in \mathbb{C}^d : \|x\|_X = 1\}$, respectively. The main geometric contribution of the present work is the following statement, as well as a (quite intricate) derivation of its algorithmic counterpart. Beyond its intrinsic interest, we will demonstrate the utility of this result by showing how it leads to major progress on the Approximate Nearest Neighbor Search problem (ANN).

[1]It is convenient and most natural to carry out the ensuing geometric and analytic considerations for normed spaces over the complex scalars $\mathbb{C}$, but all of their applications that we obtain here hold also for normed spaces over the real scalars $\mathbb{R}$ through a standard complexification procedure which is recalled in the full version.

Theorem 1 (Existence of a Hölder homeomorphism between spheres of perturbed spaces)**.** *Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a normed space and fix $\alpha, \beta, \gamma \in (0, \frac{1}{2}]$. Suppose that the inradius and outradius of B_X are $r > 0$ and $R > 0$, respectively, i.e., $rB_{\ell_2^d} \subseteq B_X \subseteq RB_{\ell_2^d}$. Then there are normed spaces $Y = (\mathbb{C}^d, \|\cdot\|_Y)$ and $Z = (\mathbb{C}^d, \|\cdot\|_Z)$, and a bijection $\varphi \colon S_Y \to S_Z$, with the following properties.*

1) $r^{2\alpha+\beta(1-2\alpha)} B_Y \subseteq B_X \subseteq R^{2\alpha+\beta(1-2\alpha)} B_Y$.
2) $r^{\gamma(1-2\alpha)} B_{\ell_2^d} \subseteq B_Z \subseteq R^{\gamma(1-2\alpha)} B_{\ell_2^d}$.
3) $\|\varphi(y_1) - \varphi(y_2)\|_Z \lesssim \frac{1}{\sqrt{\beta\gamma}} \|y_1 - y_2\|_Y^{\alpha}$ *for all* $y_1, y_2 \in S_Y$.
4) $\|\varphi^{-1}(z_1) - \varphi^{-1}(z_2)\|_Y \lesssim \frac{1}{\sqrt{\beta\gamma}} \|z_1 - z_2\|_Z^{\alpha}$ *for all* $z_1, z_2 \in S_Z$.

In the applications of Theorem 1 obtained in this paper, the parameters α, β, γ are chosen to be small, in which case the first two assertions of Theorem 1 mean that Y and Z are relatively small perturbations of X and ℓ_2^d, respectively. The last two assertions of Theorem 1 state that the mapping φ is a homeomorphism between the unit spheres of these perturbed spaces with quite good continuity properties. There is tension between the smallness of α, β, γ (thus, the extent to which the initial geometries of X and ℓ_2^d were deformed) and the quality of the continuity of φ and φ^{-1}; the parameters will eventually be set to appropriately balance these competing features.

Theorem 1 is a finite-dimensional quantitative refinement in the spirit of [1] of the work of Daher [2] which is itself an extension of a landmark contribution of Odell and Schlumprecht [3] (in unpublished work, Kalton independently obtained the result of [2]; see [4, page 216] or the MathSciNet review of [2]). Our proof of Theorem 1 is an adaptation of the proof of the corresponding qualitative infinite-dimensional result that appears in [4, Chapter 9], i.e., our contribution towards Theorem 1 is mainly the idea that such a formulation should hold true via an application of known insights (and that it is useful, as we shall soon see). However, this is only the conceptual starting point of the present investigation, because Theorem 1 is merely

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00024

an existential statement which is insufficient for the ensuing algorithmic application. Making Theorem 1 algorithmic raises a number of challenges whose resolution is interesting in its own right; this constitutes the bulk of the present work and an overview of what it entails appears later in Subsection I-B.

The mapping φ of Theorem 1 has several drawbacks in comparison to more traditional bi-Lipschitz embeddings that are used ubiquitously for algorithmic purposes. These drawbacks include the fact that one first deforms the initial space X of interest to obtain a new space Y, that φ is defined only on the sphere of Y rather than on all of Y, and that $\varphi\colon S_Y \to S_Z$ and $\varphi^{-1} : S_Z \to S_Y$ are Hölder continuous rather than Lipschitz. In addition, φ takes values in a normed space Z which is a perturbation of ℓ_2^d, so the image of the embedding does not have the "vanilla" Euclidean structure. We will later see how to overcome all of these drawbacks, and demonstrate that to a certain extent the "curse of dimensionality" is not present for the Approximate Nearest Neighbor Search problem in arbitrary normed spaces.

A. Approximate near neighbors

Given $c > 1$ and $r > 0$, the *c-Approximate Near Neighbor Search* (c-ANN) problem is defined as follows. Given an n-point dataset $P \subseteq X$ lying in a metric space (X, d_X), we want to preprocess P to answer *approximate near neighbor queries* quickly. Namely, given a query point $q \in X$ such that there is a data point $p \in P$ with $d_X(q,p) \leqslant r$, the algorithm should return a data point $\widetilde{p} \in X$ with $d_X(q,\widetilde{p}) \leqslant cr$. We refer to c as the *approximation* and r as the *distance scale*; both parameters are known during the preprocessing. The main quantities to optimize are: the time it takes to build the data structure for a given set of points (*preprocessing time*); the *space* the data structure occupies, and the time it takes to answer a query (*query time*). In addition to being an indispensable tool for data analysis, ANN data structures have spawned two decades of influential theoretical developments (see, e.g., the surveys [5], [6] and the thesis [7] for an overview).

The best-studied metrics in the context of ANN are the ℓ_1^d (Hamming/Manhattan) and the ℓ_2^d (Euclidean) distances on $\mathbb{R}^d$. Both ℓ_1^d and ℓ_2^d are very common in applications and admit efficient algorithms based on *randomized space partitions*; in particular, Locality-Sensitive Hashing (LSH) [8], [9] and its data-dependent counterparts [10], [11], [12]. Hashing-based algorithms for ANN over ℓ_1^d and ℓ_2^d have now been the subject of a long line of work, leading to a comprehensive understanding of the respective time–space trade-offs.

Beyond ℓ_1^d and ℓ_2^d, our understanding of the ANN problem is much more limited. For example, if a metric of interest is given by a *norm* on $\mathbb{R}^d$ or $\mathbb{C}^d$, then the best known general approximation bound for the ANN problem is $c \lesssim \sqrt{d}$ if we require space to be polynomial in n and d and query time to be sublinear in n and polynomial in d. This follows from John's theorem [13], which states that any d-dimensional norm can be approximated (after a linear transformation) by ℓ_2^d within a factor of $\sqrt{d}$, combined with any ANN data structure for ℓ_2^d which has constant approximation.

The recent work of the authors [14] made the first progress on ANN for arbitrary normed spaces beyond the use of John's theorem. The approximation has been improved from $\sqrt{d}$ to $\log d$, however the data structure is only implementable in the *cell-probe* model of computation [15], [16]. Recall that in the cell-probe model, data structures are only charged for the number of cells used (space), and the number of cells probed during a query procedure; however, the *time* of the query procedure may be *unbounded*. We now state the main result of [14] formally:

Theorem 2 ([14]). *Let $0 < \varepsilon < 1$ and $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a d-dimensional normed space. There exists a randomized data structure for c-ANN over X with the following guarantees:*

- *The approximation is $c \lesssim \frac{\log d}{\varepsilon^2}$;*
- *The query procedure probes $n^{\varepsilon} \cdot d^{O(1)}$* words *in memory, where each word has $O(\log n)$ bits[2];*
- *The space used by the data structure is $n^{1+\varepsilon} \cdot d^{O(1)}$.*

The work [14] was able to make the data structure of Theorem 2 *time-efficient* for two special cases, ℓ_p and Schatten-p spaces[3], however the pressing question of getting a time-efficient ANN data structure for a general normed space with approximation $o(\sqrt{d})$ was left open. In this paper, we answer this question by showing two new ANN data structures, which rely heavily on (an algorithmic counterpart of) Theorem 1. The two data structures (to be presented below as Theorem 3 and Theorem 4) use the Hölder homeomorphism in two different ways: Theorem 3 proceeds by the "embedding" approach, and Theorem 4 proceeds by the "spectral" approach of [14].

Theorem 3. *Suppose that $X = (\mathbb{C}^d, \|\cdot\|_X)$ is a d-dimensional normed space. Then there exists a randomized data structure for c-ANN over X with the following guarantees:*

- *The approximation is*
  $$c \leqslant \exp\Big(O\Big((\log d)^{\frac{2}{3}}(\log\log d)^{\frac{1}{3}}\Big)\Big);$$
- *The query procedure takes $d^{O(1)} \cdot (\log n)^{O(1)}$* time*;*
- *The space used by the data structure is $n^{O(1)} \cdot d^{O(1)}$;*
- *The preprocessing time is $n^{O(1)} \cdot d^{O(1)}$.*

The norm is specified by an oracle, which, given a vector $x \in \mathbb{C}^d$, computes $\|x\|_X$.

[2]We assume that all the coordinates of the dataset and query points as well as r can be stored in $O(\log n)$ bits.

[3]For the case of Schatten-p spaces, the space and time of the data structure of [14] had dependence $d^{O(p)}$, which is undesirable for $p \gg 1$.

Theorem 3 is the first ANN data structure with approximation $d^{o(1)}$ that works for an arbitrary norm, but its virtue is not only its great generality: there are concrete norms of interest, such as the operator norm on d-by-d matrices, or more generally Schatten-p spaces when $p \gg 1$, for which it yields the first data structure of this type. The proof of Theorem 3 is achieved by substituting our (yet to be stated) algorithmic version of Theorem 1 into an appropriate adaptation of the ANN framework of [17], [18] (see Section I-C).

If one is allowed to drop the requirement that the preprocessing time is polynomial, then we have the following result that yields both improved approximation, and space that is near-linear in n. This is achieved by substituting our algorithmic version of Theorem 1 into the framework of [14], which relies on nonlinear spectral gaps. We will sketch later in the introduction (Section I-D) why this requires us to sacrifice the polynomial preprocessing time.

Theorem 4. *Let $0 < \varepsilon < 1$ and $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a d-dimensional normed space. Then there exists a randomized data structure for c-ANN over X with the following guarantees:*

- *The approximation is* $c \leqslant \exp\left(O\left(\sqrt{\log d}\cdot \max\left\{\sqrt{\log\log d}, \frac{\log(1/\varepsilon)}{\sqrt{\log\log d}}\right\}\right)\right)$;
- *The query procedure takes $n^{\varepsilon}\cdot d^{O(1)}$ time;*
- *The space used by the data structure is $n^{1+\varepsilon}\cdot d^{O(1)}$;*
- *The preprocessing time is $n^{O(1)}\cdot d^{O(d)}$.*

The norm is specified by an oracle, which, given a vector $x \in \mathbb{C}^d$, computes $\|x\|_X$.

The new bounds on the approximation c cannot possibly be obtained by designing a (linear) low-distortion bi-Lipschitz embedding of X into ℓ_1, ℓ_2, or any fixed (universal) $d^{O(1)}$-dimensional normed space, even if the embedding is randomized; see [19] for a formalization and proof of this statement.

B. Algorithmic version of Theorem 1

For algorithmic applications, we would like to compute the mapping φ from Theorem 1 efficiently at any given input point in $\mathbb{C}^d$. The main ingredient in the construction of F is the notion of *complex interpolation* between normed spaces, which was introduced in [20]. For two d-dimensional normed spaces U and V, complex interpolation provides a one-parameter family of d-dimensional normed spaces $[U,V]_\theta$ indexed by $\theta \in [0,1]$, such that $[U,V]_0 = U$, $[U,V]_1 = V$ and $[U,V]_\theta$ depends, in a certain sense, smoothly on θ. In particular, we need to compute the norm of a vector in $[U,V]_\theta$ given suitable oracles for the norm computation in U and V. This is a non-trivial task since the norm in $[U,V]_\theta$ is defined as the minimum of a certain functional on an infinite-dimensional space of holomorphic functions. We show how to properly "discretize" this optimization problem using harmonic and complex analysis, and ultimately solve it using convex programming (more specifically, the "robust" ellipsoid method [21]). We expect that the resulting algorithmic version of complex interpolation will have further applications.

More specifically, for $x \in \mathbb{C}^d$ the interpolated norm $\|x\|_{[U,V]_\theta}$ is defined as follows. First, we consider the space $\mathcal{F}$ of functions $F\colon \overline{\mathfrak{S}} \to \mathbb{C}^d$, where $\overline{\mathfrak{S}} = \{z \in \mathbb{C} \mid 0 \leqslant \operatorname{Re} z \leqslant 1\}$ is a strip in the complex plane, such that:

- F is bounded and continuous;
- F is holomorphic on the interior of $\overline{\mathfrak{S}}$.

The norm $\|F\|_{\mathcal{F}}$ in the space $\mathcal{F}$ is defined as follows:

$$\|F\|_{\mathcal{F}} = \max\left\{\sup_{\operatorname{Re} z=0}\|F(z)\|_U, \sup_{\operatorname{Re} z=1}\|F(z)\|_V\right\}.$$

Finally, for $x \in \mathbb{C}^d$, we define:

$$\|x\|_{[U,V]_\theta} = \inf_{\substack{F\in\mathcal{F}:\\ F(\theta)=x}} \|F\|_{\mathcal{F}}. \tag{1}$$

A priori, it is not clear how to solve (1), since the space $\mathcal{F}$ is infinite-dimensional. However, we are able to show that one can search for an approximately optimal $F \in \mathcal{F}$ of the following form:

$$F(z) = e^{\varepsilon z^2}\cdot \sum_{|k|\leqslant M} v_k e^{\frac{kz}{L}},$$

where $\varepsilon > 0$, M and L are fixed parameters, and the vectors $v_k \in \mathbb{C}^d$ are the variables. This turns (1) into a finite-dimensional convex program, which we might hope to solve. However, in order for the optimization procedure to be efficient, we need upper bounds on M and the magnitudes of v_k. These can be established by taking an approximately optimal (in terms of (1)) function F, smoothing it by convolving with an appropriate Gaussian, and finally considering its Fourier expansion, whose convergence we can control using the classical Fejér's theorem [22]. To bound the magnitudes of v_k, we need a statement similar to the Paley–Wiener theorem [22]. Finally, to address the issue that the norm in $\mathcal{F}$ is defined as a supremum over an infinite set (the boundary of the strip $\overline{\mathfrak{S}}$), we show how to discretize and truncate the boundary so that the maximum over the discretization is not too far from the true supremum. This is again possible due to the bounds on the magnitudes of ε, v_k and M we are able to show.

C. The embedding approach: proof of Theorem 3

The first application of Theorem 1 to ANN for general normed spaces (Theorem 3) follows the "embedding" approach. Suppose we want to design an efficient data structure for ANN over a metric space (W_0, d_{W_0}), and we have an efficient data structure for ANN over another metric space (W_1, d_{W_1}). Then, if we have an embedding $W_0 \to W_1$ at our disposal, a data structure for (W_0, d_{W_0}) could be obtained by applying the embedding and employing the known data structure for (W_1, d_{W_1}). The approximation guarantee one obtains depends on how well the embedding preserves the geometry of W_0.

The key to Theorem 3 is to use Theorem 1 as an embedding of Y into Z. Recall that $Y = (\mathbb{C}^d, \|\cdot\|_Y)$ and $Z = (\mathbb{C}^d, \|\cdot\|_Z)$ are small perturbations of the spaces $X = (\mathbb{C}^d, \|\cdot\|_X)$ and ℓ_2^d, respectively. At a high level, an ANN data structure for ℓ_2^d gives a data structure for Z, a data structure for Z gives a data structure for Y via the embedding, and a data structure for Y gives a data structure for X. The initial step in this chain (giving efficient ANN data structures for ℓ_2^d) is accomplished by any of the efficient data structures known for ℓ_2^d, specifically, we use the data structure of [8], [23].

One caveat to the plan set forth above is that Theorem 1 gives an embedding only for the unit sphere of Y. It can be extended to the whole space, but the resulting map distorts large distances prohibitively. This challenge already comes up in [17], [18] in the context of designing ANN data structures for ℓ_p spaces, where instead of Theorem 1, the Mazur map [24] was used. We resolve the issue of large distances in the same way as [17], [18]: in particular, [18] gives a clean reduction from the general ANN problem to a special case, when all the points lie in a small ball. Our final approximation guarantee in Theorem 3 is the result of balancing the parameters α, β and γ in Theorem 1.

D. The spectral approach: proof of Theorem 4

We now sketch the proof of Theorem 4. For this we use the framework based on nonlinear spectral gaps developed in [14]. In a sentence, the outline of the proof is in the spirit of what has been done in [14] for the Schatten-p norm, while using Theorem 1 instead of the estimates on the noncommutative Mazur map from [25].

The proof of Theorem 4 consists of a few steps. The data structure for a normed space X relies on a randomized space partition of X, which by duality is equivalent to the existence of sparse cuts in graphs embedded into X. The latter follows from a nonlinear Rayleigh quotient inequality, which refines the nonlinear spectral gap inequality used to prove Theorem 2. Finally, we show how to obtain the desired nonlinear Rayleigh quotient inequality using the map from Theorem 1.

Let us now explain why in Theorem 4 we do not obtain efficient preprocessing. The main obstacle that in order to construct a randomized space partition of X, we need to find sparse cuts in graphs embedded in X that have size exponential in the dimension d. Another issue is that the argument for the existence of sparse cuts uses a fixed-point argument similar to the Brouwer's fixed point theorem, and it is unclear how to make it algorithmically efficient.

Now let us describe the proof of Theorem 4 in greater detail.

1) Sparse cuts in embedded graphs: We first recall the outline of the proof of Theorem 2. The starting point is a space partitioning statement, which readily follows from [1]. Recall that for a k-regular graph $G = (V, E)$ the conductance of a cut $(S, \overline{S})$ is defined as:

$$\frac{E(S, \overline{S})}{k \cdot \min\{|S|, |\overline{S}|\}}.$$

Lemma 1 ([1])**.** *Let $0 < \varepsilon < 1$, and suppose that $X = (\mathbb{C}^d, \|\cdot\|_X)$ is a d-dimensional normed space. Let $G = (V, E)$ be a regular undirected graph with n vertices. Suppose that $f\colon V \to X$ is an arbitrary map such that for every edge $\{u, v\} \in E$ one has $\|f(u) - f(v)\|_X \leqslant 1$. Then,*

- *Either there exists a ball[4] of radius $R \lesssim \frac{\log d}{\varepsilon^2}$, which contains $\Omega(n)$ images of the vertices V under f;*
- *Or there exists a cut in G with conductance at most ε.*

Equipped with Lemma 1, the proof of Theorem 2 proceeds in two steps:

- First, we use a version of the minimax theorem to convert Lemma 1 to the following randomized partitioning procedure, which can be seen as a version of data-dependent hashing (in spirit of [10], [11], [12]).

  Lemma 2 ([14])**.** *Let $0 < \varepsilon < 1$. Suppose that $X = (\mathbb{C}^d, \|\cdot\|_X)$ is a d-dimensional normed space. Let $P \subseteq X$ be a dataset of n points. Then:*

  - *Either there exists a ball of radius $R \lesssim \frac{\log d}{\varepsilon^2}$, which contains $\Omega(n)$ points from P;*
  - *Or there exists a distribution $\mathcal{D}$ over "reasonable" sets (see below for a clarification of what "reasonable" means here) $A \subseteq X$ such that:*
    - $\Pr_{A \sim \mathcal{D}}\Big[\Omega(n) \leqslant |A \cap P| \leqslant (1 - \Omega(1)) \cdot n\Big] = 1$*;*
    - *For every $x_1, x_2 \in X$ with $0 < \|x_1 - x_2\|_X \leqslant 1$, one has:*

      $$\Pr_{A \sim \mathcal{D}}\Big[\big|A \cap \{x_1, x_2\}\big| = 1\Big] < \varepsilon.$$

- Then, we apply Lemma 2 recursively to build the desired $O\left(\frac{\log d}{\varepsilon^2}\right)$-ANN data structure, which concludes the proof of Theorem 2. This step is by now standard and is similar to what was done in [26], [11], [12].

Let us now explain why Theorem 2 requires the cell-probe model. In the resulting data structure, a query point is tested against a sequence of cuts guaranteed by Lemma 1. Thus, it is crucial to be able to check efficiently, which side of the cut a given vertex of the graph G belongs to. However, the main issue is that Lemma 1 gives us *no control* on the promised sparse cut in G. In particular, a cut does not have to be induced by a geometrically nice subset of the ambient space $\mathbb{C}^d$. This is a serious problem, since in the proof of Lemma 2 we invoke Lemma 1 for graphs of size *exponential in d*, so we cannot afford to store the resulting sparse cuts

[4]In the metric induced by the norm $\|\cdot\|_X$.

explicitly. Nevertheless, there is a way to store cuts from the support of $\mathcal{D}$ in space $\mathrm{poly}(d)$ (this is exactly what we mean by "reasonable" in the statement of Lemma 2), but the argument for this is quite delicate: we need to perform the minimax argument in a careful way using the (nested) Multiplicative Weights Update algorithm [27]. This yields Theorem 2, but the query procedure is grossly inefficient in terms of time, since in order to test a point against a cut, one has to spend time exponential in d to re-compute the cut from its succinct description.

Thus, in order to prove Theorem 4, we need a version of Lemma 1 which gives a sparse cut that we are able to not only store efficiently, but also to test against in time $\mathrm{poly}(d)$. We accomplish this by showing the following lemma.

Lemma 3. *Suppose that $X = (\mathbb{C}^d, \|\cdot\|_X)$ is a d-dimensional normed space. There exists a map $\Phi\colon \mathbb{C}^d \to \mathbb{C}^d$, which one can compute efficiently for a given input point, such that the following holds. Suppose that $0 < \varepsilon < 1$ and let $G = (V, E)$ be a regular undirected graph with n vertices. Suppose that $f\colon V \to X$ is an arbitrary map such that for every edge $\{u, v\} \in E$ one has $\|f(u) - f(v)\|_X \leqslant 1$. Then,*

- *either there exists a ball of radius $R = \exp\Big(\widetilde{O}_\varepsilon\big(\sqrt{\log d}\big)\Big)$, which contains $\Omega(n)$ images of the vertices V under f;*
- *Or there exists a vector $w = w(G, f) \in \mathbb{C}^d$, an index $i = i(G, f) \in [d]$, and a threshold $\tau = \tau(G, f) \in \mathbb{R}$ such that at least one of the cuts $\{v \in V \mid \mathrm{Re}\, \Phi(f(v) - w)_i \leqslant \tau\}$ or $\{v \in V \mid \mathrm{Im}\, \Phi(f(v) - w)_i \leqslant \tau\}$ in G has conductance at most ε.*

Now we can store a cut by simply storing w, i, τ and whether we test real or imaginary part, and, moreover, one can test, on which side of the cut a given point lies, since the map Φ is efficiently computable (and depends only on the norm). To prove Lemma 3, we use Theorem 1 crucially. Namely, the map Φ in Lemma 3 is a radial extension of the map φ from Theorem 1. More details of the proof of Lemma 3 are given in the following section.

Let us remark that for $R \lesssim \sqrt{d}/\varepsilon$, the analog of Lemma 3 holds with cuts induced by the sets $\{v \in V \mid \mathrm{Re}(Tf(v))_i \leqslant \tau\}$ and $\{v \in V \mid \mathrm{Im}(Tf(v))_i \leqslant \tau\}$, where $T\colon \mathbb{C}^d \to \mathbb{C}^d$ is a fixed linear map. This is an easy corollary of Cheeger's inequality and John's theorem. The cuts guaranteed by Lemma 3 are more complicated (yet we can work with them efficiently), but this complication allows us to get a much better bound of $R = \exp\Big(\widetilde{O}_\varepsilon\big(\sqrt{\log d}\big)\Big)$.

2) Nonlinear Rayleigh quotient inequalities and Lemma 3: Let $A = (a_{ij})$ be a non-negative symmetric $n \times n$ matrix with $\sum_{i,j=1}^n a_{ij} = 1$. Denote $\rho_A(i) = \sum_{j=1}^n a_{ij}$. For a metric space (X, d_X), $q > 0$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in X^n$, where not all x_i's are the same, we define the *nonlinear Rayleigh quotient* $\mathsf{R}(\boldsymbol{x}, A, d_X^q)$ as follows:

$$\mathsf{R}(\boldsymbol{x}, A, d_X^q) = \frac{\sum_{i,j=1}^n a_{ij} \cdot d_X(x_i, x_j)^q}{\sum_{i,j=1}^n \rho(i)\rho(j) \cdot d_X(x_i, x_j)^q}.$$

Let G be a regular undirected graph with n vertices, and denote by A its normalized adjacency matrix. On the one hand, Cheeger's inequality [28] states that if for some $\boldsymbol{x} \in (\mathbb{C}^d)^n$, one has

$$\mathsf{R}(\boldsymbol{x}, A, \|\cdot\|_{\ell_2^d}^2) \leqslant \frac{\varepsilon^2}{10}, \tag{2}$$

then there exists a cut in G with conductance at most ε. Moreover, up to the exact dependence on ε, the condition (2) for *some* $\boldsymbol{x}$ is necessary to have a sparse cut. One the other hand, suppose that $X = (\mathbb{C}^d, \|\cdot\|)$ is a normed space, and $f\colon V \to X$ is a map such that for every edge $(u, v) \in E$ one has $\|f(u) - f(v)\|_X \leqslant 1$. If there is no ball of radius D, which contains $\Omega(n)$ images of the vertices V under f, then the definition of nonlinear Rayleigh quotient directly implies that:

$$\mathsf{R}(\boldsymbol{x}, A, \|\cdot\|_X^2) \lesssim \frac{1}{D^2},$$

where $\boldsymbol{x}_v = f(v)$. Thus, in order to prove Lemma 1 or Lemma 3, we need statements that relate nonlinear Rayleigh quotients with respect to the Euclidean geometry and the geometry given by X, a normed space of interest.

In light of the above discussion, Lemma 1 readily follows from the following inequality proved in [1]:

Theorem 5 ([1], reformulation)**.**

$$\inf_{\boldsymbol{y} \in (\mathbb{C}^d)^n} \mathsf{R}(\boldsymbol{y}, A, \|\cdot\|_{\ell_2^d}^2) \lesssim (\log d) \cdot \inf_{\boldsymbol{x} \in (\mathbb{C}^d)^n} \mathsf{R}(\boldsymbol{x}, A, \|\cdot\|_X^2)^{\frac{1}{2}}. \tag{3}$$

The standard proof of Cheeger's inequality shows that if $\mathsf{R}(\boldsymbol{y}, A, \|\cdot\|_{\ell_2^d}^2)$ is small, then there exists a sparse cut induced by a *coordinate cut* of $\boldsymbol{y}$. More formally, there exist $i \in [d]$ and $\tau \in \mathbb{C}$ such that one of the cuts $\{v \in V \mid \mathrm{Re}(\boldsymbol{y}_v)_i \leqslant \tau\}$ or $\{v \in V \mid \mathrm{Im}(\boldsymbol{y}_v)_i \leqslant \tau\}$ is sparse. However, Theorem 5 gives *no control* over how $\boldsymbol{y}$ is related to $\boldsymbol{x}$. This is the reason why in Lemma 1 we cannot guarantee that the desired sparse cut is induced by a geometrically nice subset of $\mathbb{C}^d$.

In this work, we prove a refinement of Theorem 5, which implies Lemma 3 similarly to the above argument.

Theorem 6. *For every $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in (\mathbb{C}^d)^n$ such that not all x_i's are equal, there exists $w = w(\boldsymbol{x}, A) \in \mathbb{C}^d$ such that:*

$$\mathsf{R}(\boldsymbol{\Phi}_w(\boldsymbol{x}), A, \|\cdot\|_{\ell_2^d}^2) \lesssim \log^2 d \cdot \mathsf{R}(\boldsymbol{x}, A, \|\cdot\|_X^2)^{\Omega\left(\sqrt{\frac{\log\log d}{\log d}}\right)},$$

where:

$$\boldsymbol{\Phi}_w(x_1, x_2, \ldots, x_n) = (\Phi(x_1 - w), \Phi(x_2 - w), \ldots, \Phi(x_n - w)),$$

and Φ is a radial extension of the map φ from Theorem 1.

The proof of Theorem 6 is a combination of two ingredients. The first is an argument used by Matoušek in [29] to prove a nonlinear Rayleigh quotient inequality for ℓ_p norms. We show that this argument is in fact much more versatile, and in particular, coupled with Theorem 1, it implies Theorem 6. The vector $w = w(\boldsymbol{x}, A) \in \mathbb{C}^d$ in Theorem 6 is such that:

$$\sum_i \rho(i)\Phi(x_i - w) = 0. \tag{4}$$

And here comes the second ingredient. In the argument from [29], the counterpart of (4) easily follows from the intermediate value theorem, since $\|\cdot\|_{\ell_p}^p$ is additive over the coordinates. However, finding w such that (4) holds is more delicate. For this we use tools from algebraic topology (related to the Brouwer's fixed point theorem).

E. Related work

Most efficient ANN data structures in high-dimensional spaces beyond ℓ_1 and ℓ_2 have proceeded via the embedding approach. The typical target spaces are ℓ_1 and ℓ_2, since these admit very efficient ANN algorithms [8], [23], [9], [10], [11], [12]. Another common target space is ℓ_∞^d which can be handled with $O(\log\log d)$-approximation using the algorithm in [26]. A growing body of work has added to the list of "tractable" spaces by designing low-distortion embeddings. These include the ℓ_p-direct sums [30], [31], [32], [33], the Ulam metric [32], the Earth-Mover's distance (EMD) [34], [35], the edit distance [36], the Frechét distance [30], and symmetric normed spaces [19].

Another class of metric spaces studied assume low intrinsic dimension, and efficient ANN algorithms in this setting are known for *any metric space* [37], [38], [39], [40]. The dimensionality of these spaces is assumed to be $d = o(\log n)$, so efficient algorithms may depend *exponentially* on d. In this paper, we deal with the *high-dimensional* regime (when $\omega(\log n) \leqslant d \leqslant n^{o(1)}$), hence the dependence on d must be polynomial.

F. Organization of the paper

We present the necessary background to our results in Section II. We formulate the Hölder homeomorphism from Theorem 1 in Section III. We show the algorithmic version of Theorem 1 and its applications to the ANN problem in the full version.

G. Acknowledgments

We thank Sébastien Bubeck, Yin Tat Lee and Yuval Peres for useful discussions. The first-named author was supported in part by the Simons Foundation (#491119), NSF grants CCF-1617955, CCF-1740833, and Google Research Award. The second-named author was supported in part by the NSF grant CCF-1412958, the Packard Foundation and the Simons Foundation. The third author was supported by an NSERC Discovery grant (application number RGPIN-2016-06333), and part of this work was done while he was visiting the Simons Institute for the Theory of Computing semester on Bridging Continuous and Discrete Optimization, partially supported through NSF grant #CCF-1740425. The fifth-named author was supported in part by the NSF grants CCF-1563155, CCF-1420349, CCF-1149257, CCF-1423100), and the NSF Graduate Research Fellowship (DGE-16-44869).

II. Preliminaries

Given two quantities $a, b > 0$, the notation $a \lesssim b$ and $b \gtrsim a$ means $a \leqslant Cb$ for some universal constant $C > 0$. In this work we use some tools from complex analysis. Denote $\mathfrak{S} = \{z \in \mathbb{C} \mid 0 < \operatorname{Re} z < 1\} \subseteq \mathbb{C}$ the unit open strip on the complex plane, let $\partial\mathfrak{S} = \{z \in \mathbb{C} \mid \operatorname{Re} z \in \{0,1\}\}$ be its boundary, and, finally, let $\overline{\mathfrak{S}} = \mathfrak{S} \cup \partial\mathfrak{S}$ be the corresponding closed strip. Given a normed space X defined over a (real or complex) vector space V, the subset $B_X \subseteq V$ is the unit ball of X, i.e., $B_X = \{x \in V : \|x\|_X \leqslant 1\}$. For a measure space (Ω, μ) and a Banach space X we denote $L_p(\Omega, \mu, X)$ the Banach space of measurable functions $f \colon \Omega \to X$ such that

$$\int_\Omega \|f\|_X^p \, d\mu < +\infty;$$

we define the norm to be:

$$\|f\|_{L_p(\Omega,\mu,X)}^p = \int_\Omega \|f\|_X^p \, d\mu.$$

Function that agree almost everywhere are identified. Sometimes, we omit Ω in the notation if it is clear from the context (or unimportant).

A. The Poisson kernel for the strip $\mathfrak{S}$

For $w \in \mathfrak{S}$ and $z \in \partial\mathfrak{S}$, the Poisson kernel $P(w, z)$ for $\mathfrak{S}$ is defined as follows:

$$P(w,z) = \begin{cases} \frac{1}{2} \cdot \frac{\sin \pi u}{\cosh \pi(\tau - v) - \cos \pi u}, \\ \text{for } w = u + iv \text{ and } z = i\tau, \\ \frac{1}{2} \cdot \frac{\sin \pi u}{\cosh \pi(\tau - v) + \cos \pi u}, \\ \text{for } w = u + iv \text{ and } z = 1 + i\tau. \end{cases} \tag{5}$$

For every $w \in \mathfrak{S}$, and every $z \in \partial\mathfrak{S}$, one has $P(w, z) \geqslant 0$. In addition, for every $w \in \mathfrak{S}$,

$$\int_{\partial\mathfrak{S}} P(w,z)\, dz = 1,$$

which allows us to denote μ_w the measure on $\partial\mathfrak{S}$ with the density $P(w, \cdot)$. We refer the reader to [41] for further properties of the kernel $P(\cdot, \cdot)$.

For $\theta_1, \theta_2 \in (0, 1)$, we let

$$\Lambda(\theta_1, \theta_2) \stackrel{\text{def}}{=} \sqrt{\left(\frac{1}{\theta_1} + \frac{1}{1-\theta_1}\right)\left(\frac{1}{\theta_2} + \frac{1}{1-\theta_2}\right)}, \tag{6}$$

The following claim is proved in the full version.

Claim 4. *For any $z \in \partial\mathfrak{S}$ and $\theta_1, \theta_2 \in (0,1)$,*

$$\frac{P(\theta_1, z)}{P(\theta_2, z)} \lesssim \Lambda(\theta_1, \theta_2)^2.$$

B. Harmonic and holomorphic functions on $\mathfrak{S}$

Lemma 5 ([41])**.** *Let $f\colon \overline{\mathfrak{S}} \to \mathbb{R}$ be a continuous function which is harmonic (as a function of two real variables) in $\mathfrak{S}$. Moreover, suppose that the integral*

$$\int_{\partial\mathfrak{S}} |f(z)|\, d\mu_w(z)$$

is finite for some $w \in \mathfrak{S}$. Then for every $w \in \mathfrak{S}$, one has:

$$f(w) = \int_{\partial\mathfrak{S}} f(z)\, d\mu_w(z).$$

Corollary 6. *Let $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d$ be a continuous function which is holomorphic in $\mathfrak{S}$. Moreover, suppose that*

$$\int_{\partial\mathfrak{S}} \big\|f(z)\big\|\, d\mu_w(z) < \infty$$

for some $w \in \mathfrak{S}$. Then for every $w \in \mathfrak{S}$, one has:

$$f(w) = \int_{\partial\mathfrak{S}} f(z)\, d\mu_w(z).$$

Proof: This follows from Lemma 5 and the fact that the real and the imaginary part of a holomorphic function are harmonic. ∎

C. Complex interpolation between normed spaces

Let $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ be two d-dimensional complex normed spaces. We will now define a family of spaces $[W_0, W_1]_\theta = (\mathbb{C}^d, \|\cdot\|_{[W_0,W_1]_\theta})$ for $0 \leqslant \theta \leqslant 1$ that, in a sense we will make precise later, interpolate between W_0 and W_1. This definition appeared for the first time in [20]; see also the book [42]. Let us first define an auxiliary (infinite-dimensional) normed space $\mathcal{F}$ as the space of bounded continuous functions $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d$, which are holomorphic in $\mathfrak{S}$. The norm on $\mathcal{F}$ is defined as follows:

$$\|f\|_{\mathcal{F}} = \max\left\{\sup_{\mathrm{Re}(z)=0} \|f(z)\|_{W_0}, \sup_{\mathrm{Re}(z)=1} \|f(z)\|_{W_1}\right\}.$$

Now we can define the interpolation norm $\|\cdot\|_{[W_0,W_1]_\theta}$ on $\mathbb{C}^d$ as follows:

$$\|x\|_{[W_0,W_1]_\theta} = \inf_{\substack{f \in \mathcal{F}:\\ f(\theta)=x}} \|f\|_{\mathcal{F}}. \tag{7}$$

The fact that $\|x\|_{[W_0,W_1]_\theta}$ is a norm is straightforward to check modulo the property "$\|x\|_{[W_0,W_1]_\theta} = 0$ implies $x = 0$". The latter is a consequence of the Hadamard three-lines theorem [43]. The following facts can be found in [42].

Fact 7. *For every $\theta \in [0,1]$, $[W_0, W_1]_\theta = [W_1, W_0]_{1-\theta}$.*

Fact 8 (Reiteration theorem)**.** *For every $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant 1$ and $0 \leqslant \theta_3 \leqslant 1$, one has:*

$$\Big[[W_0, W_1]_{\theta_1}, [W_0, W_1]_{\theta_2}\Big]_{\theta_3} = [W_0, W_1]_{(1-\theta_3)\theta_1 + \theta_3\theta_2}.$$

Below is arguably the most useful statement about complex interpolation. We recall that for a linear map $T\colon \mathbb{C}^d \to \mathbb{C}^{d'}$ and normed spaces $X = (\mathbb{C}^d, \|\cdot\|_X)$ and $Y = (\mathbb{C}^d, \|\cdot\|_Y)$, $\|T\|_{X\to Y}$ denotes the operator norm, equal to $\sup_{x\in\mathbb{C}^d} \frac{\|Tx\|_Y}{\|x\|_X}$.

Fact 9. *Let $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ be d-dimensional complex normed spaces, and let $U_0 = (\mathbb{C}^{d'}, \|\cdot\|_{U_0})$ and $U_1 = (\mathbb{C}^{d'}, \|\cdot\|_{U_1})$ be a couple of d'-dimensional ones. Suppose that $T\colon \mathbb{C}^d \to \mathbb{C}^{d'}$ is a linear map. Then, for every $0 \leqslant \theta \leqslant 1$, one has:*

$$\|T\|_{[W_0,W_1]_\theta \to [U_0,U_1]_\theta} \leqslant \|T\|_{W_0\to U_0}^{1-\theta} \cdot \|T\|_{W_1\to U_1}^{\theta}.$$

Corollary 10. *Let $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ be complex normed spaces such that for some $\mathsf{d}_1, \mathsf{d}_2 \geqslant 1$ and every $x \in \mathbb{C}^d$, the following holds:*

$$\frac{1}{\mathsf{d}_1} \cdot \|x\|_{W_1} \leqslant \|x\|_{W_0} \leqslant \mathsf{d}_2 \cdot \|x\|_{W_1}. \tag{8}$$

Then, for every $0 \leqslant \theta \leqslant 1$ and every $x \in \mathbb{C}^d$, one has:

$$\frac{1}{\mathsf{d}_1^\theta} \cdot \|x\|_{[W_0,W_1]_\theta} \leqslant \|x\|_{W_0} \leqslant \mathsf{d}_2^\theta \cdot \|x\|_{[W_0,W_1]_\theta}$$

and

$$\frac{1}{\mathsf{d}_1^{1-\theta}} \cdot \|x\|_{W_1} \leqslant \|x\|_{[W_0,W_1]_\theta} \leqslant \mathsf{d}_2^{1-\theta} \cdot \|x\|_{W_1}.$$

Proof: This follows from Fact 9 applied to the identity map. ∎

Fact 11. *Let $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ be complex normed spaces, and let ${W_0}^* = (\mathbb{C}^d, \|\cdot\|_{{W_0}^*})$ and ${W_1}^* = (\mathbb{C}^d, \|\cdot\|_{{W_1}^*})$ be the dual spaces, respectively. For any $\theta \in [0,1]$, the dual space to $[W_0, W_1]_\theta$, given by $[W_0, W_1]_\theta^* = (\mathbb{C}^d, \|\cdot\|_{[W_0,W_1]_\theta^*})$ is isometric to the space $[{W_0}^*, {W_1}^*]_\theta$.*

D. Uniform convexity

Let $W = (\mathbb{C}^d, \|\cdot\|_W)$ be a complex normed space. We give necessary definitions related to the notion of uniform convexity. For a thorough overview, see [44].

Definition 12. *For $2 \leqslant p \leqslant \infty$, the space W has modulus of convexity of power type p iff there exists $K \geqslant 1$ such that for every $x, y \in W$:*

$$\left(\|x\|_W^p + \frac{1}{K^p}\|y\|_W^p\right)^{1/p} \leqslant$$

$$\leqslant \left(\frac{\|x+y\|_W^p + \|x-y\|_W^p}{2}\right)^{1/p}.$$

The infimum of such K is called the p-convexity constant of W and is denoted by $K_p(W)$.

Claim 13. *One always has $K_\infty(W) = 1$, and for a Hilbert space, one has: $K_2(\ell_2^d) = 1$.*

Given two normed spaces $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^{d'}, \|\cdot\|_{W_1})$, and $1 \leqslant p < \infty$, we use the notation $W_0 \oplus_p W_1$ for the direct sum of W_0 and W_1 endowed with the norm $\|(x_0, x_1)\|_{W_0 \oplus_p W_1} = (\|x_0\|_{W_0}^p + \|x_1\|_{W_1}^p)^{1/p}$.

Claim 14. *One has $K_p(W_0 \oplus_2 W_1) \lesssim \max\{K_p(W), K_p(W_1)\}$ for every $p \geqslant 2$.*

Proof: The claim follows from $W_0 \oplus_2 W_1$ being isomorphic to $W_0 \oplus_p W_1$ and the fact that $K_p(W_0 \oplus_p W_1) \leqslant \max\{K_p(W), K_p(W_1)\}$. ■

Lemma 15 ([45]). *One has $K_p(L_2(\mu, W)) \lesssim K_p(W)$ for every $p \geqslant 2$.*

The following lemma shows how the p-convexity constant interacts with complex interpolation.

Lemma 16 ([1]). *For every $2 \leqslant p_1, p_2 \leqslant \infty$ and every $0 \leqslant \theta \leqslant 1$, one has:*

$$K_{\frac{p_1 p_2}{\theta p_1 + (1-\theta) p_2}}\Big([W_0, W_1]_\theta\Big) \leqslant K_{p_1}(W_0)^{1-\theta} K_{p_2}(W_1)^\theta.$$

E. The space $\mathcal{F}_2(\theta)$

Now we define another space related to $\mathcal{F}$. This definition appears in [20], see also [4]. First, for $0 < \theta < 1$, let us consider the normed space $\mathcal{G}(\theta)$ of continuous functions $f \colon \overline{\mathfrak{S}} \to \mathbb{C}^d$, which are holomorphic in $\mathfrak{S}$, and

$$\int\limits_{\partial\mathfrak{S}} \big\|f(z)\big\|^2 \, d\mu_\theta(z) < \infty.$$

The norm $\|f\|_{\mathcal{G}(\theta)}$ is defined as follows:

$$\|f\|_{\mathcal{G}(\theta)}^2 = \int\limits_{\mathrm{Re}(z)=0} \big\|f(z)\big\|_{W_0}^2 \, d\mu_\theta(z) + \int\limits_{\mathrm{Re}(z)=1} \big\|f(z)\big\|_{W_1}^2 \, d\mu_\theta(z). \quad (9)$$

Clearly, $\mathcal{F} \subseteq \mathcal{G}(\theta)$. One may naturally view $\mathcal{G}(\theta)$ as a (not closed) subspace of $L_2(\{z \mid \mathrm{Re}(z) = 0\}, \mu_\theta, W_0) \oplus_2 L_2(\{z \mid \mathrm{Re}(z) = 1\}, \mu_\theta, W_1)$. Now we can define the space $\mathcal{F}_2(\theta)$ as the closure of $\mathcal{G}(\theta)$ (in particular, $\mathcal{G}(\theta)$ is dense in $\mathcal{F}_2(\theta)$). An element of $\mathcal{F}_2(\theta)$ can be identified with a function $f \colon \overline{\mathfrak{S}} \to \mathbb{C}^d$ defined *almost everywhere* on $\partial\mathfrak{S}$ and defined everywhere on $\mathfrak{S}$ such that:

- f restricted on $\{z \mid \mathrm{Re}(z) = 0\}$ belongs to $L_2(\{z \mid \mathrm{Re}(z) = 0\}, \mu_\theta, W_0)$;
- f restricted on $\{z \mid \mathrm{Re}(z) = 1\}$ belongs to $L_2(\{z \mid \mathrm{Re}(z) = 1\}, \mu_\theta, W_1)$;
- f is holomorphic in $\mathfrak{S}$;

In this representation, the norm is defined similarly to (9):

$$\|f\|_{\mathcal{F}(\theta)}^2 = \int\limits_{\mathrm{Re}(z)=0} \big\|f(z)\big\|_{W_0}^2 \, d\mu_\theta(z) + \int\limits_{\mathrm{Re}(z)=1} \big\|f(z)\big\|_{W_1}^2 \, d\mu_\theta(z).$$

Claim 17. *For every $f \in \mathcal{F}_2(\theta)$ and $w \in \mathfrak{S}$, one has:*

$$f(w) = \int_{\partial\mathfrak{S}} f(z) \, d\mu_w(z).$$

Proof: This identity is true for $\mathcal{G}(\theta)$ by Corollary 6. Hence it holds for $\mathcal{F}_2(\theta)$, since every element of $\mathcal{F}_2(\theta)$ is a limit of a sequence of elements of $\mathcal{G}(\theta)$, which converges in L_2 in $\partial\mathfrak{S}$ and pointwise in $\mathfrak{S}$. ■

The following gives an alternative definition of an interpolated norm, which should be compared with the original definition (7).

Fact 18 ([4]). *For every $x \in \mathbb{C}^d$, one has:*

$$\|x\|_{[W_0, W_1]_\theta} = \inf_{\substack{f \in \mathcal{F}_2(\theta): \\ f(\theta) = x}} \|f\|_{\mathcal{F}_2(\theta)}. \quad (10)$$

Claim 19. *For every $p \geqslant 2$, one has:*

$$K_p(\mathcal{F}_2(\theta)) \lesssim \max\{K_p(W_0), K_p(W_1)\}.$$

Proof: One has:

$$\begin{aligned} K_p(\mathcal{F}_2(\theta)) &\leqslant K_p\Big(L_2(\{z \mid \mathrm{Re}(z) = 0\}, \mu_\theta, W_0) \oplus_2 \\ &\quad L_2(\{z \mid \mathrm{Re}(z) = 1\}, \mu_\theta, W_1)\Big) \\ &\lesssim \max\Big\{K_p\Big(L_2(\{z \mid \mathrm{Re}(z) = 0\}, \mu_\theta, W_0)\Big), \\ &\quad K_p\Big(L_2(\{z \mid \mathrm{Re}(z) = 1\}, \mu_\theta, W_1)\Big)\Big\} \\ &\lesssim \max\{K_p(W_0), K_p(W_1)\}, \end{aligned}$$

where the first step is due to $\mathcal{F}_2(\theta)$ being a subspace of $L_2(\{z \mid \mathrm{Re}(z) = 0\}, \mu_\theta, W_0) \oplus_2 L_2(\{z \mid \mathrm{Re}(z) = 1\}, \mu_\theta, W_1)$, the second step is due to Claim 14, and the third step is due to Lemma 15. ■

Lemma 20. *For $0 < \theta_1, \theta_2 < 1$, the spaces $\mathcal{F}_2(\theta_1)$ and $\mathcal{F}_2(\theta_2)$ are isomorphic via the identity map. More specifically, for every $f \in \mathcal{F}_2(\theta_1)$ one has:*

$$\|f\|_{\mathcal{F}_2(\theta_2)} \leqslant \Lambda(\theta_1, \theta_2) \cdot \|f\|_{\mathcal{F}_2(\theta_1)};$$

and, similarly, for every $f \in \mathcal{F}_2(\theta_2)$, one has:

$$\|f\|_{\mathcal{F}_2(\theta_1)} \leqslant \Lambda(\theta_1, \theta_2) \cdot \|f\|_{\mathcal{F}_2(\theta_2)}.$$

Proof: This easily follows from the definition of $\mathcal{F}_2(\theta)$ and Claim 4. ■

III. Hölder homeomorphisms: an existential argument

In this section we show the proof of Theorem 1 making the exposition of the result from [2] in [4] quantitative. We make the construction of the map algorithmic in the full version. Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a normed space of interest. For a real normed space, one can consider its complexification, which contains the real version isometrically.

Let us first assume that $K_p(X) < \infty$ for some $2 \leqslant p < \infty$. We start with taking a closer look at Fact 18. Suppose that we interpolate between X and ℓ_2^d and moreover for some $0 < r < R$ one has:

$$rB_{\ell_2^d} \subseteq B_X \subseteq RB_{\ell_2^d}.$$

Let $\mathcal{F}_2(\theta)$ be defined with respect to X and ℓ_2^d.

Fact 21 ([4])**.** *For every $x \in \mathbb{C}^d$, in the optimization problem*

$$\inf_{\substack{F \in \mathcal{F}_2(\theta): \\ F(\theta) = x}} \|F\|_{\mathcal{F}_2(\theta)}$$

the minimum is attained on an element of $\mathcal{F}_2(\theta)$. Moreover, the minimizer is unique, and we denote it by $F^*_{\theta x} \in \mathcal{F}_2(\theta)$.

The fact below shows that the minimizers $F^*_{\theta x}$ have very special structure.

Fact 22 ([4])**.** *Fix $x \in \mathbb{C}^d$ and $0 < \theta < 1$ and consider $F^*_{\theta x} \in \mathcal{F}_2(\theta)$. Then,*

- *For $z \in \mathbb{C}$ such that $\operatorname{Re} z = 0$, $\|F^*_{\theta x}(z)\|_X = \|x\|_{[X, \ell_2^d]_\theta}$ almost everywhere;*
- *For $z \in \mathbb{C}$ such that $\operatorname{Re} z = 1$, $\|F^*_{\theta x}(z)\|_{\ell_2^d} = \|x\|_{[X, \ell_2^d]_\theta}$ almost everywhere;*
- *For every $0 < \widetilde{\theta} < 1$, $\|F^*_{\theta x}(\widetilde{\theta})\|_{[X, \ell_2^d]_{\widetilde{\theta}}} = \|x\|_{[X, \ell_2^d]_\theta}$.*

The lemma below is the core of the overall argument.

Lemma 23 (a quantitative version of a statement from [4])**.** *For every $0 < \theta < 1$ and every $x_1, x_2 \in S_{[X, \ell_2^d]_\theta}$, one has:*

$$\left\| F^*_{\theta x_1} - F^*_{\theta x_2} \right\|_{\mathcal{F}_2(\theta)} \lesssim K_p(X) \cdot \|x_1 - x_2\|^{1/p}_{[X, \ell_2^d]_\theta}.$$

Proof: By Claim 19, one has:

$$K_p(\mathcal{F}_2(\theta)) \lesssim \max\{K_p(X), K_p(\ell_2^d)\} \lesssim K_p(X),$$

where the second step follows from $K_p(\ell_2^d) \lesssim 1$. Second, suppose that for $x_1, x_2 \in S_{[X, \ell_2^d]_\theta}$, one has

$$\|x_1 - x_2\|_{[X, \ell_2^d]_\theta} = \varepsilon > 0.$$

Then,

$$\left\| F^*_{\theta x_1} + F^*_{\theta x_2} \right\|_{\mathcal{F}_2(\theta)} \geqslant \|x_1 + x_2\|_{[X, \ell_2^d]_\theta} \geqslant 2 - \varepsilon, \quad (11)$$

where the first step follows from Fact 18, and the second step follows from x_1 and x_2 being unit and the triangle inequality.

Now by the definition of $K_p(\mathcal{F}_2(\theta))$ (Definition 12) and the fact that the minimizers are unit, we have:

$$\left\| F^*_{\theta x_1} + F^*_{\theta x_2} \right\|^p_{\mathcal{F}_2(\theta)} + \frac{\left\| F^*_{\theta x_1} - F^*_{\theta x_2} \right\|^p_{\mathcal{F}_2(\theta)}}{K_p(\mathcal{F}_2(\theta))^p} \leqslant \frac{\|2F^*_{\theta x_1}\|^p_{\mathcal{F}_2(\theta)} + \|2F^*_{\theta x_2}\|^p_{\mathcal{F}_2(\theta)}}{2} = 2^p. \quad (12)$$

Combining (11) and (12), we get:

$$\begin{aligned} \left\| F^*_{\theta x_1} - F^*_{\theta x_2} \right\|^p_{\mathcal{F}_2(\theta)} &\leqslant K_p(\mathcal{F}_2(\theta))^p \cdot (2^p - (2 - \varepsilon)^p) \\ &\leqslant p2^{p-1} \cdot K_p(\mathcal{F}_2(\theta))^p \cdot \varepsilon. \end{aligned}$$

Finally, we get:

$$\begin{aligned} \left\| F^*_{\theta x_1} - F^*_{\theta x_2} \right\|_{\mathcal{F}_2(\theta)} &\lesssim K_p(\mathcal{F}_2(\theta)) \cdot \varepsilon^{1/p} \\ &\lesssim K_p(X) \cdot \varepsilon^{1/p} = K_p(X) \cdot \|x_1 - x_2\|^{1/p}_{[X, \ell_2^d]_\theta} \end{aligned}$$

as desired. ∎

Fix $0 < \theta_1, \theta_2 < 1$. Define the map $U_{\theta_1 \theta_2} \colon S_{[X, \ell_2^d]_{\theta_1}} \to S_{[X, \ell_2^d]_{\theta_2}}$ as follows:

$$x \mapsto F^*_{\theta_1 x}(\theta_2).$$

The map is well-defined, since by Fact 22 for every x with $\|x\|_{[X, \ell_2^d]_{\theta_1}}$, one has $\|F^*_{\theta_1, x}(\theta_2)\|_{[X, \ell_2^d]_{\theta_2}} = 1$. One also has: $U^{-1}_{\theta_1, \theta_2} = U_{\theta_2, \theta_1}$, since, by Fact 22 and the uniqueness of the minimizer (Fact 21), for every $x \in \mathbb{C}^d$, one has:

$$F^*_{\theta_2 F^*_{\theta_1 x}(\theta_2)} = F^*_{\theta_1 x}.$$

In particular, U_{θ_1, θ_2} is a bijection between the unit spheres of $[X, \ell_2^d]_{\theta_1}$ and $[X, \ell_2^d]_{\theta_2}$.

Lemma 24 (a quantitative version of the statement from [4])**.** *For $x_1, x_2 \in S_{[X, \ell_2^d]_{\theta_1}}$, one has:*

$$\begin{aligned} &\|U_{\theta_1 \theta_2}(x_1) - U_{\theta_1 \theta_2}(x_2)\|_{[X, \ell_2^d]_{\theta_2}} \\ &\qquad \lesssim \Lambda(\theta_1, \theta_2) \cdot K_p(X) \cdot \|x_1 - x_2\|^{1/p}_{[X, \ell_2^d]_{\theta_1}}. \end{aligned}$$

Proof: One has:

$$\begin{aligned} &\|U_{\theta_1 \theta_2}(x_1) - U_{\theta_1 \theta_2}(x_2)\|_{[X, \ell_2^d]_{\theta_2}} \\ &= \|F^*_{\theta_1 x_1}(\theta_2) - F^*_{\theta_1 x_2}(\theta_2)\|_{[X, \ell_2^d]_{\theta_2}} \\ &\leqslant \|F^*_{\theta_1 x_1} - F^*_{\theta_1 x_2}\|_{\mathcal{F}_2(\theta_2)} \\ &\leqslant \Lambda(\theta_1, \theta_2) \cdot \|F^*_{\theta_1 x_1} - F^*_{\theta_1 x_2}\|_{\mathcal{F}_2(\theta_1)} \\ &\lesssim \Lambda(\theta_1, \theta_2) \cdot K_p(X) \cdot \|x_1 - x_2\|^{1/p}_{[X, \ell_2^d]_{\theta_1}}, \end{aligned}$$

where the first step is by the definition of $U_{\theta_1 \theta_2}$, the second step is due to Fact 18, the third step is due to Lemma 20, and the last step is due to Lemma 23. ∎

The theorem below summarizes the above discussion.

Theorem 7. *Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a complex normed space such that $K_p(X) < \infty$ for some $2 \leqslant p < \infty$ and for* *some* $0 < r < R$*, one has:* $rB_{\ell_2^d} \subseteq B_X \subseteq RB_{\ell_2^d}$*. Fix* $0 < \beta, \gamma \leqslant 1/2$*. Then there exist two spaces* $Y = (\mathbb{C}^d, \|\cdot\|_Y)$ *and* $Z = (\mathbb{C}^d, \|\cdot\|_Z)$ *and a bijection* $\varphi\colon S_Y \to S_Z$ *such that:*

- $r^{\beta} B_Y \subseteq B_X \subseteq R^{\beta} B_Y$*;*
- $r^{\gamma} B_{\ell_2^d} \subseteq B_Z \subseteq R^{\gamma} B_{\ell_2^d}$*;*
- *for every* $y_1, y_2 \in S_Y$*,*
$$\|\varphi(y_1) - \varphi(y_2)\|_Z \lesssim \frac{K_p(X)}{\sqrt{\beta\gamma}} \cdot \|y_1 - y_2\|_Y^{1/p};$$
- *for every* $z_1, z_2 \in S_Z$*,*
$$\|\varphi^{-1}(z_1) - \varphi^{-1}(z_2)\|_Y \lesssim \frac{K_p(X)}{\sqrt{\beta\gamma}} \cdot \|z_1 - z_2\|_Z^{1/p}.$$

Proof: We set Y and Z to be $[X, \ell_2^d]_\beta$ and $[X, \ell_2^d]_{1-\gamma}$, respectively. Finally, set φ to be $U_{\beta, 1-\gamma}$. Then, the first two inequalities follow from Corollary 10. The third inequality follows from Lemma 24 combined with the estimate $\Lambda(\beta, \gamma) \lesssim \frac{1}{\sqrt{\beta\gamma}}$. The fourth inequality is shown similarly to the third taking into account that $\varphi^{-1} = U_{1-\gamma, \beta}$. ■

Now let us turn to the case when X is not necessarily p-convex.

Theorem 8 (Theorem 1, restated). *Let* $X = (\mathbb{C}^d, \|\cdot\|_X)$ *be a complex normed space such that for some* $0 < r < R$*, one has:* $rB_{\ell_2^d} \subseteq B_X \subseteq RB_{\ell_2^d}$*. Fix* $0 < \alpha, \beta, \gamma \leqslant 1/2$*. Then there exist two spaces* $Y = (\mathbb{C}^d, \|\cdot\|_Y)$ *and* $Z = (\mathbb{C}^d, \|\cdot\|_Z)$ *and a bijection* $\varphi\colon S_Y \to S_Z$ *such that:*

- $r^{2\alpha + \beta(1-2\alpha)} B_Y \subseteq B_X \subseteq R^{2\alpha + \beta(1-2\alpha)} B_Y$*;*
- $r^{\gamma(1-2\alpha)} B_{\ell_2^d} \subseteq B_Z \subseteq R^{\gamma(1-2\alpha)} B_{\ell_2^d}$*;*
- *for every* $y_1, y_2 \in S_Y$*,*
$$\|\varphi(y_1) - \varphi(y_2)\|_Z \lesssim \frac{1}{\sqrt{\beta\gamma}} \cdot \|y_1 - y_2\|_Y^{\alpha};$$
- *for every* $z_1, z_2 \in S_Z$*,*
$$\|\varphi^{-1}(z_1) - \varphi^{-1}(z_2)\|_Y \lesssim \frac{1}{\sqrt{\beta\gamma}} \cdot \|z_1 - z_2\|_Z^{\alpha}.$$

Proof: Denote $A = [X, \ell_2^d]_{2\alpha}$. By Lemma 16, one has $K_{1/\alpha}(A) \leqslant 1$. Let us now apply Theorem 7 to A, which yields two spaces $Y = [A, \ell_2^d]_\beta$ and $Z = [A, \ell_2^d]_\gamma$. By Fact 8, one has: $Y = [X, \ell_2^d]_{2\alpha + \beta(1-2\alpha)}$ and $Z = [X, \ell_2^d]_{2\alpha + (1-\gamma)(1-2\alpha)}$, which together with Corollary 10 yields the first two items. The third and fourth items follow from Theorem 7 applied to A. ■

References

[1] A. Naor, "A Spectral Gap Precludes Low-Dimensional Embeddings," in *Proceedings of the 33rd International Symposium on Computational Geometry (SoCG '2017)*, 2017, pp. 50:1–50:16, available as arXiv:1611.08861.

[2] M. Daher, "Homéomorphismes uniformes entre les sphères unité des espaces d'interpolation," *Comptes Rendus Mathematique*, vol. 316, no. 10, pp. 1051–1054, 1993.

[3] E. Odell and T. Schlumprecht, "The distortion problem," *Acta Math.*, vol. 173, no. 2, pp. 259–281, 1994. [Online]. Available: https://doi.org/10.1007/BF02398436

[4] Y. Benyamini and J. Lindenstrauss, *Geometric Nonlinear Functional Analysis. Vol. 1*, ser. American Mathematical Society Colloquium Publications. Providence, RI: American Mathematical Society, 2000, vol. 48.

[5] A. Andoni and P. Indyk, "Nearest Neighbors in High-Dimensional Spaces," in *Handbook of Discrete and Computational Geometry*, J. E. Goodman, J. O'Rourke, and C. D. Tóth, Eds. CRC Press LLC, 2017, pp. 1133–1153.

[6] A. Andoni, P. Indyk, and I. Razenshteyn, "Approximate nearest neighbor search in high dimensions," in *Proceedings of ICM 2018 (to appear)*, 2018.

[7] I. Razenshteyn, "High-Dimensional Similarity Search and Sketching: Algorithms and Hardness," Ph.D. dissertation, Massachusetts Institute of Technology, 2017.

[8] 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.

[9] A. Andoni and P. 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)*, 2006, pp. 459–468.

[10] A. Andoni, P. Indyk, H. L. Nguyen, and I. Razenshteyn, "Beyond Locality-Sensitive Hashing," in *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA '2014)*, 2014, pp. 1018–1028, available as arXiv:1306.1547.

[11] A. Andoni and I. Razenshteyn, "Optimal Data-Dependent Hashing for Approximate Near Neighbors," in *Proceedings of the 47th ACM Symposium on the Theory of Computing (STOC '2015)*, 2015, pp. 793–801, available as arXiv:1501.01062.

[12] A. Andoni, T. Laarhoven, I. Razenshteyn, and E. 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)*, 2017, pp. 47–66, available as arXiv:1608.03580.

[13] F. John, "Extremum Problems with Inequalities as Subsidiary Conditions," in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*. Interscience Publishers, Inc., New York, N. Y., 1948, pp. 187–204.

[14] A. Andoni, A. Naor, A. Nikolov, I. Razenshteyn, and E. Waingarten, "Data-Dependent Hashing via Nonlinear Spectral Gaps," in *Proceedings of the 50th ACM Symposium on the Theory of Computing (STOC '2018)*, 2018, to appear.

[15] A. C.-C. Yao, "Should tables be sorted?" *Journal of the ACM*, vol. 28, no. 3, pp. 615–628, 1981.

[16] P. B. Miltersen, "Cell Probe Complexity – a Survey," in *Advances in Data Structures*, 1999.

[17] A. Naor and Y. Rabani, "On Approximate Nearest Neighbor Search in ℓ_p, $p > 2$," 2006, unpublished manuscript; available on request.

[18] Y. Bartal and L.-A. Gottlieb, "Approximate Nearest Neighbor Search for ℓ_p-Spaces ($2 < p < \infty$) via Embeddings," in *Proceedings of the 13th Latin American Symposium in Theoretical Informatics (LATIN '2018)*, 2018.

[19] A. Andoni, H. L. Nguyen, A. Nikolov, I. Razenshteyn, and E. Waingarten, "Approximate Near Neighbors for General Symmetric Norms," in *Proceedings of the 49th ACM Symposium on the Theory of Computing (STOC '2017)*, 2017, pp. 902–913, available as arXiv:1611.06222.

[20] A. Calderón, "Intermediate Spaces and Interpolation, the Complex Method," *Studia Mathematica*, vol. 24, no. 2, pp. 113–190, 1964.

[21] Y. T. Lee, A. Sidford, and S. S. Vempala, "Efficient convex optimization with membership oracles," *arXiv:1706.07357*, 2017.

[22] Y. Katznelson, *An introduction to harmonic analysis*. Cambridge University Press, 2004.

[23] E. Kushilevitz, R. Ostrovsky, and Y. Rabani, "Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces," *SIAM Journal on Computing*, vol. 30, no. 2, pp. 457–474, 2000.

[24] S. Mazur, "Une remarque sur l'homéomorphie des champs fonctionels," *Studia Math.*, vol. 1, pp. 83–85, 1929.

[25] E. Ricard, "Hölder Estimates for the Noncommutative Mazur Map," *Archiv der Mathematik*, vol. 104, no. 1, pp. 37–45, 2015.

[26] P. Indyk, "On Approximate Nearest Neighbors under ℓ_∞ Norm," *Journal of Computer and System Sciences*, vol. 63, no. 4, pp. 627–638, 2001.

[27] S. Arora, E. Hazan, and S. Kale, "The Multiplicative Weights Update Method: a Meta-Algorithm and Applications," *Theory of Computing*, vol. 8, no. 1, pp. 121–164, 2012.

[28] J. Cheeger, "A Lower Bound for the Smallest Eigenvalue of the Laplacian," in *Proceedings of the Princeton conference in honor of Professor S. Bochner*, 1969, pp. 195–199.

[29] J. Matoušek, "On the distortion required for embedding finite metric spaces into normed spaces," *Israel Journal of Mathematics*, vol. 93, pp. 333–344, 1996.

[30] P. Indyk, "Approximate Nearest Neighbor Algorithms for Fréchet Distance via Product Metrics," in *Proceedings of the 18th ACM Symposium on Computational Geometry (SoCG '2002)*, 2002, pp. 102–106.

[31] ——, "Approximate Nearest Neighbor under Edit Distance via Product Metrics," in *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms (SODA '2004)*, 2004, pp. 646–650.

[32] A. Andoni, P. Indyk, and R. Krauthgamer, "Overcoming the ℓ_1 Non-Embeddability Barrier: Algorithms for Product Metrics," in *Proceedings of the 20th ACM-SIAM Symposium on Discrete Algorithms (SODA '2009)*, 2009, pp. 865–874.

[33] A. Andoni, "Nearest Neighbor Search: the Old, the New, and the Impossible," Ph.D. dissertation, MIT, 2009.

[34] M. Charikar, "Similarity Estimation Techniques from Rounding Algorithms," in *Proceedings of the 34th ACM Symposium on the Theory of Computing (STOC '2002)*, 2002, pp. 380–388.

[35] P. Indyk and N. Thaper, "Fast Color Image Retrieval via Embeddings," 2003, workshop on Statistical and Computational Theories of Vision (at ICCV).

[36] R. Ostrovsky and Y. Rabani, "Low Distortion Embedding for Edit Distance," *Journal of the ACM*, vol. 54, no. 5, pp. 23:1–23:16, 2007.

[37] K. L. Clarkson, "Nearest neighbor queries in metric spaces," *Discrete and Computational Geometry*, vol. 22, no. 1, pp. 63–93, 1999.

[38] D. R. Karger and M. Ruhl, "Finding Nearest Neighbors in Growth-Restricted Metrics," in *Proceedings of the 34th ACM Symposium on the Theory of Computing (STOC '2002)*, 2002, pp. 741–750.

[39] R. Krauthgamer and J. R. Lee, "Navigating Nets: Simple Algorithms for Proximity Search," in *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms (SODA '2004)*, 2004, pp. 798–807.

[40] A. Beygelzimer, S. Kakade, and J. Langford, "Cover Trees for Nearest Neighbor," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '2006)*, 2006, pp. 97–104.

[41] D. V. Widder, "Functions Harmonic in a Strip," *Proceedings of the American Mathematical Society*, vol. 12, no. 1, pp. 67–72, 1961.

[42] J. Bergh and J. Löfström, *Interpolation spaces: an introduction*. Springer-Verlag, 1976.

[43] E. M. Stein and R. Shakarchi, *Complex analysis. Princeton Lectures in Analysis, II*. Princeton University Press, 2003.

[44] K. Ball, E. A. Carlen, and E. H. Lieb, "Sharp Uniform Convexity and Smoothness Inequalities for Trace Norms," *Inventiones Mathematicae*, vol. 115, no. 1, pp. 463–482, 1994.

[45] A. Naor, "On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon–Roichman graphs," *Combinatorics, Probability and Computing*, vol. 21, no. 4, pp. 623–634, 2012.