Hölder Homeomorphisms and Approximate Nearest Neighbors

Alexandr Andoni
Columbia University
andoni@cs.columbia.edu

Assaf Naor
Princeton University
naor@math.princeton.edu

Aleksandar Nikolov
University of Toronto
anikolov@cs.toronto.edu

Ilya Razenshteyn
Microsoft Research Redmond
ilyaraz@microsoft.com

Erik Waingarten
Columbia University
eaw@cs.columbia.edu

September 10, 2018

Abstract

We study bi-Hölder homeomorphisms between the unit spheres of finite-dimensional normed spaces and use them to obtain better data structures for the 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structural result is a finite-dimensional quantitative version of the following theorem of Daher (1993) and Kalton (unpublished).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.

Contents

1 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).

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 [Nao17] of the work of Daher [Dah93] which is itself an extension of a landmark contribution of Odell and Schlumprecht [OS94] (in unpublished work, Kalton independently obtained the result of [Dah93]; see [BL00, page 216] or the MathSciNet review of [Dah93]). Our proof of Theorem 1 is an adaptation of the proof of the corresponding qualitative infinite-dimensional result that appears in [BL00, 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 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 1.2.

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

[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 Section 2.4 below.

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.

1.1 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 [AI17, AIR18] and the thesis [Raz17] 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) [IM98, AI06] and its data-dependent counterparts [AINR14, AR15, ALRW17]. 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 [Joh48], which states that any d-dimensional norm can be approximated 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 [ANN+18] 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 [Yao81, Mil99]. 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 [ANN+18] formally:

Theorem 2 ([ANN+18])**.** *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 [ANN+18] 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.

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* $n^{O(1)} \cdot d^{O(1)}$*.*

Both the preprocessing and query procedures access the norm through an oracle, which, given a vector $x \in \mathbb{C}^d$*, computes* $\|x\|_X$*.*

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 [NR06, BG18] (see Section 1.3).

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 now near-linear in n. This is achieved by substituting our algorithmic version of Theorem 1 into the framework of [ANN+18], which relies on nonlinear spectral gaps. We will sketch later in the introduction (Section 1.4) 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* $c \leqslant \exp\Big(O\Big(\sqrt{\log d} \cdot \max\Big\{\sqrt{\log\log d}, \frac{\log(1/\varepsilon)}{\sqrt{\log\log d}}\Big\}\Big)\Big)$*;*

[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 [ANN+18] had dependence $d^{O(p)}$, which is undesirable for $p \gg 1$.

- *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* $n^{O(1)} \cdot d^{O(d)}$*.*

Both the preprocessing and query procedures access the norm through 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 [ANN+17] for a formalization and proof of this statement.

1.2 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 [Cal64]. 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 [LSV17]). 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 on 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:

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

for a fixed $\varepsilon > 0$, M and L, and variables are $v_k \in \mathbb{C}^d$. 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, one needs to upper bound M and the magnitudes of v_k. This can be done 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 [Kat04]. To bound the magnitudes of v_k, we need a statement similar to the Paley–Wiener theorem [Kat04]. Finally, to address the issue that the norm in $\mathcal{F}$ is defined as a supremum over the 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.

1.3 The embedding approach: proof of Theorem 3

The first application 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 depend 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 [IM98, KOR00].

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 [NR06, BG18] in the context of designing ANN data structures for ℓ_p spaces, where instead of Theorem 1, the Mazur map [Maz29] was used. We may resolve the issue of large distances in the same way as [NR06, BG18]: in particular, [BG18] 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.

1.4 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 [ANN$^+$18]. In a sentence, the outline of the proof is in the spirit of what has been done in [ANN$^+$18] for the Schatten-p norm, while using Theorem 1 instead of the estimates on the noncommutative Mazur map from [Ric15].

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 inequality using the map from Theorem 1.

Let us now explain why in Theorem 4 we do not obtain efficient preprocessing. The main obstacle is the exponential in d size of graphs embedded in X, in which we would like to find sparse cuts. Another issue is that the argument for the existence of sparse cuts proceeds using 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 a greater detail.

1.4.1 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 the work [Nao17]. 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 ([Nao17])**.** *Let $0 < \varepsilon < 1$. 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 1.1, the proof of Theorem 2 proceeds in two steps:

- First, we use a version of the minimax theorem to convert Lemma 1.1 to the following randomized partitioning procedure, which can be seen as a version of data-dependent hashing (in spirit of [AINR14, AR15, ALRW17]).

  Lemma 1.2 ([ANN$^+$18])**.** *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:*

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

- *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:*

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

- Then, we apply Lemma 1.2 recursively to build a 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 [Ind01, AR15, ALRW17].

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 1.1. 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 1.1 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 1.2 we invoke Lemma 1.1 for graphs of size *exponential in d*, so we cannot afford to store the resulting sparse cuts *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 1.2), 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 [AHK12]. 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 1.1 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 1.3. *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 $\{v \in V \mid \operatorname{Re}\Phi(f(v) - w)_i \leqslant \tau\}$ or $\{v \in V \mid \operatorname{Im}\Phi(f(v) - w)_i \leqslant \tau\}$ 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 1.3, we use Theorem 1 crucially. Namely, the map Φ in Lemma 1.3 is a radial extension of the map φ from Theorem 1.

Let us remark that for $R \lesssim \sqrt{d}/\varepsilon$, the analog of Lemma 1.3 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 1.3 are more complicated (yet we can work with them efficiently), but this complication allows us to get a much better bound of $R = \exp\Big(\widetilde{O}_\varepsilon\big(\sqrt{\log d}\big)\Big)$.

1.4.2 Nonlinear Rayleigh quotient inequalities and Lemma 1.3

Let $A = (a_{ij})$ be a non-negative symmetric $n \times n$ matrix with $\sum_{i,j=1}^n a_{ij} = 1$. Denote $\mathsf{\rho}_A(i) = \sum_{j=1}^n a_{ij}$. 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:

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

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 [Che69] 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 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 1.1 or Lemma 1.3, 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 1.1 readily follows from the following inequality proved in [Nao17]:

Theorem 5 ([Nao17], 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 $\boldsymbol{y}$; in particular, a priori it does not have to be related to $\boldsymbol{x}$ at all. This is exactly the reason why in Lemma 1.1 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 1.3 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 of Matoušek from [Mat96]. In [Mat96], a nonlinear Rayleigh quotient inequality for ℓ_p norms was proved, but we show that the argument in fact is much more versatile. 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 [Mat96], 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).

1.5 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 [IM98, KOR00, AI06, AINR14, AR15, ALRW17]. Another common target space is ℓ_∞^d which can be handled with $O(\log\log d)$-approximation using the algorithm in [Ind01]. 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 [Ind02, Ind04, AIK09, And09], the Ulam metric [AIK09], the Earth-Mover's distance (EMD) [Cha02, IT03], the edit distance [OR07], the Frechét distance [Ind02], and symmetric normed spaces [ANN+17].

Another class of metric spaces studied assume low intrinsic dimension, and efficient ANN algorithms in this setting are known for *any metric space* [Cla99, KR02, KL04, BKL06]. 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 $\omega(\log n) \leqslant d \leqslant n^{o(1)}$), the dependence on d must be polynomial.

1.6 Organization of the paper

We present the necessary background to our results in Section 2. We formulate the Hölder homeomorphism from Theorem 1 in Section 3. In Section 4, we define the approximate Hölder homeomorphism used in the applications to ANN. We assume two algorithms in Section 4 which we give in Section 5. After that, the next three sections (Sections 6, 7, and 8) give the applications of the approximate Hölder homeomorphism from Section 4 to ANN. Specifically, Section 6 and Section 7 give the proof of Theorem 4, and Section 8 gives the proof of Theorem 3.

Readers eager for the applications of Theorem 1 to ANN may find a summary of the properties of the approximate Hölder homeomorphism in Section 4.3 and proceed to Section 7 and Section 8 for the ANN algorithms.

1.7 Acknowledgments

We thank Sébastien Bubeck, Yin Tat Lee and Yuval Peres for useful discussions.

2 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 $f : \Omega \to X$ such that

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

we define the norm to be:

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

Sometimes, we omit Ω in the notation if it is clear from the context (or unimportant).

2.1 Computational model for general normed spaces

Throughout this work, we deal with computational aspects of ANN defined over general normed spaces, in particular $X = (\mathbb{R}^d, \|\cdot\|_X)$. We work with the standard computational models for convex sets over $\mathbb{R}^d$. In particular, we may assume the following about X:

- There exists an oracle which, given $x \in \mathbb{R}^d$, computes $\|x\|_X$;
- The unit ball of X satisfies $B_X \subseteq B_2 \subseteq \mathsf{d}B_X$ for $\mathsf{d} = \operatorname{poly}(d)$.

The second assumption is essentially without loss of generality. Indeed, if one assumes B_X is contained within the unit Euclidean ball and contains a small Euclidean ball of radius $r = \exp(-\text{poly}(d))$, then, by the reductions of [GLS12], we may design a separation oracle for B_X, and as noted in Section 1.1 of [KLS97], this means we can transform B_X to be in a position such that the second assumptions holds.

2.2 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:

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

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 [Wid61] for further properties of the kernel $P(\cdot,\cdot)$.

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

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

Claim 2.1. *For any $z \in \partial\mathfrak{S}$ and $\theta_1, \theta_2 \in (0,1)$,*

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

Proof. First, consider the case $z = i\tau$ when $\tau \in \mathbb{R}$. Then by the first case of (5),

$$\begin{aligned} \frac{P(\theta_1, i\tau)}{P(\theta_2, i\tau)} &= \frac{\sin(\pi\theta_1)}{\cosh(\pi\tau) - \cos(\pi\theta_1)} \cdot \frac{\cosh(\pi\tau) - \cos(\pi\theta_2)}{\sin(\pi\theta_2)} \\ &\lesssim \theta_1 \left(\frac{1}{\theta_2} + \frac{1}{1-\theta_2}\right) \cdot \frac{\cosh(\pi\tau) - \cos(\pi\theta_2)}{\cosh(\pi\tau) - \cos(\pi\theta_1)} \\ &\lesssim \theta_1 \left(\frac{1}{\theta_2} + \frac{1}{1-\theta_2}\right) \cdot \left(1 + \frac{1}{\theta_1^2}\right), \end{aligned}$$

where in the first line, we use the fact that $\sin(\pi\theta) \approx \theta$ when $\theta \approx 0$ and $\sin(\pi\theta) \approx 1 - \theta$ when $\theta \approx 1$, and in the second line, we use the fact that $\cosh(\pi\tau) \geqslant 1$, and the fact that $1 - \cos(\pi\theta) \gtrsim \frac{1}{\theta^2}$. By the second case of (5), when $z = 1 + i\tau$ for $\tau \in \mathbb{R}$,

$$\begin{aligned} \frac{P(\theta_1, 1 + i\tau)}{P(\theta_2, 1 + i\tau)} &= \frac{\sin(\pi\theta_1)}{\cosh(\pi\tau) + \cos(\pi\theta_1)} \cdot \frac{\cosh(\pi\tau) + \cosh(\pi\theta_2)}{\sin(\pi\theta_2)} \\ &\lesssim (1-\theta_1)\left(\frac{1}{\theta_2} + \frac{1}{1-\theta_2}\right) \cdot \left(1 + \frac{1}{(1-\theta_1)^2}\right), \end{aligned}$$

where we now use the fact that $1 + \cos(\pi\theta) \gtrsim \frac{1}{(1-\theta)^2}$. □

2.3 Harmonic and holomorphic functions on $\mathfrak{S}$

Lemma 2.2 ([Wid61]). *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*

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

if finite for $j = 0, 1$ and some $w \in \mathfrak{S}$. Then for every $w \in \mathfrak{S}$, one has:

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

Corollary 2.3. *Let $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d$ be a continuous function which is holomorphic in $\mathfrak{S}$. Moreover, suppose that*

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

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

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

Proof. This follows from Lemma 2.2 and the fact that the real and the imaginary part of a holomorphic function are harmonic. □

2.4 Complexification

Let $X = (\mathbb{R}^d, \|\cdot\|_X)$ be a normed space over the vector space $\mathbb{R}^d$. The *complexification* of X, denoted by $X^{\mathbb{C}}$, is a normed space over $\mathbb{C}^d$ defined as follows. Elements of $X^{\mathbb{C}}$ are formal sums $u + iv$ for $u, v \in X$. Given $u + iv, w + iy \in X^{\mathbb{C}}$, addition of $(u + iv) + (w + iy)$ is given by $(u+iv)+(w+iy) = (u+w)+i(v+y)$. Given $u+iv \in X^{\mathbb{C}}$ and $\alpha = p+iq \in \mathbb{C}$, scalar multiplication $\alpha(u + iv)$ is given by, $\alpha(u + iv) = (pu - qv) + i(pv + qu)$. Finally, the norm on $X^{\mathbb{C}}$ is defined as:

$$\|u + iv\|_{X^{\mathbb{C}}}^2 = \frac{1}{\pi} \int\limits_0^{2\pi} \|u \cos\varphi - v \sin\varphi\|_X^2 \, d\varphi. \tag{7}$$

The space $X^{\mathbb{C}} = (\mathbb{C}^d, \|\cdot\|_{X^{\mathbb{C}}})$ defined above is indeed a complex normed space and, moreover, X embeds into $X^{\mathbb{C}}$ isometrically via the map $u \mapsto u + i \cdot 0$. In addition, consider the d-dimensional space $(\ell_2^d)^{\mathbb{C}} = (\mathbb{C}^d, \|\cdot\|_{(\ell_2^d)^{\mathbb{C}}})$ given by the complexification of the space $\ell_2^d = (\mathbb{R}^d, \|\cdot\|_2)$. We note that:

$$\|u + iv\|_{(\ell_2^d)^{\mathbb{C}}}^2 = \frac{1}{\pi}\int_0^{2\pi} \|u\cos\varphi - v\sin\varphi\|_2^2 d\varphi = \frac{1}{\pi}\int_0^{2\pi} \sum_{i=1}^d (u_i \cos\varphi - v_i \sin\varphi)^2 d\varphi = \|u\|_2^2 + \|v\|_2^2,$$

which implies that $(\ell_2^d)^{\mathbb{C}}$ is isometric to $\ell_2^{2d} = (\mathbb{R}^{2d}, \|\cdot\|_2)$, where we consider splitting the real and imaginary parts of each coordinate, and interpreting these as real numbers. We note that by a simple calculation, if $W_0 = (\mathbb{R}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{R}^d, \|\cdot\|_{W_1})$ are real normed spaces with $B_{W_0} \subseteq B_{W_1} \subseteq \mathsf{d} \cdot B_{W_0}$, then $B_{W_0{}^{\mathbb{C}}} \subseteq B_{W_1{}^{\mathbb{C}}} \subseteq \mathsf{d} \cdot B_{W_0{}^{\mathbb{C}}}$.

2.5 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 [Cal64], see also the book [BL76]. 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:

$$\|x\|_{[W_0,W_1]_\theta} = \inf_{\substack{f\in\mathcal{F}:\\ f(\theta)=x}} \|f\|_{\mathcal{F}}. \tag{8}$$

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 [SS03].

Fact 2.4. *For every $\theta \in [0,1]$, $[W_0, W_1]_\theta = [W_1, W_0]_{1-\theta}$.*

Fact 2.5 (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.

Fact 2.6 ([Cal64, BL76])**.** *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 $T\colon \mathbb{C}^d \to \mathbb{C}^{d'}$ be 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 2.7. *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:*

$$\frac{1}{\mathsf{d}_1} \cdot \|x\|_{W_1} \leqslant \|x\|_{W_0} \leqslant \mathsf{d}_2 \cdot \|x\|_{W_1}. \tag{9}$$

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

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

Proof. This follows from Fact 2.6 applied to the identity map. □

Fact 2.8 ([Cal64, BL76])**.** *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 $[{W_0}^*, {W_1}^*]_\theta$.*

2.6 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 [BCL94].

Definition 2.9. *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 $x, y \in W$:*

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

Definition 2.10. *The infimum of such K is called the p-convexity constant of W and is denoted by $K_p(W)$.*

Claim 2.11. *One always has $K_\infty(W) = 1$, and for a Hilbert space, one has: $K_2(\ell_2^d) = 1$.*

Claim 2.12. *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 $K_p(W_0 \oplus_p W_1) \leqslant \max\{K_p(W), K_p(W_1)\}$. $\square$

Lemma 2.13 ([Nao12])**.** *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 2.14 ([Nao17])**.** *For every $2 \leqslant p_1, p_2 \leqslant \infty$ and every $0 \leqslant \theta \leqslant 1$, one has:*

$$K_{\frac{p_1p_2}{\theta p_1 + (1-\theta)p_2}}\left([W_0, W_1]_\theta\right) \leqslant K_{p_1}(W_0)^{1-\theta} K_{p_2}(W_1)^\theta.$$

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

Now we define another space related to $\mathcal{F}$. This definition appears in [Cal64], see also [BL00]. First, for $0 < \theta < 1$, let us consider the normed space $\mathcal{G}(\theta)$ of continuous functions $f : \overline{\mathfrak{S}} \to \mathbb{C}^d$, which are holomorphic in $\mathfrak{S}$, and

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

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

$$\|f\|_{\mathcal{G}(\theta)}^2 = \int_{\mathrm{Re}(z)=0} \|f(z)\|_{W_0}^2 \, d\mu_\theta(z) + \int_{\mathrm{Re}(z)=1} \|f(z)\|_{W_1}^2 \, d\mu_\theta(z). \tag{10}$$

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 $f : \overline{\mathfrak{S}} \to \mathbb{C}^d$ 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 similar to (10):

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

Fact 2.15. *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 2.3. 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 pointwise in $\mathfrak{S}$ and in L_2 in $\partial\mathfrak{S}$. $\square$

The following gives an alternative definition of an interpolated norm, which should be compared with the original definition (8).

Fact 2.16 ([BL00])**.** *For every $x \in \mathbb{C}^d$, one has:*

$$\|x\|_{[W_0,W_1]_\theta} = \inf_{\substack{f\in\mathcal{F}_2(\theta):\\ f(\theta)=x}} \|f\|_{\mathcal{F}_2(\theta)}. \tag{11}$$

Claim 2.17. *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:

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

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 2.12, and the third step is due to Lemma 2.13. $\square$

Lemma 2.18. *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 2.1. $\square$

3 Hölder homeomorphisms: an existential argument

In this section we show the proof of Theorem 1 making the exposition of the result from [Dah93] in [BL00] quantitative. We make the construction of the map algorithmic in Section 4 and Section 5. 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 2.16. 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 3.1 ([BL00])**.** *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 $F^*_{\theta x} \in \mathcal{F}_2(\theta)$.

Below statement shows that minimizers $F^*_{\theta x}$ have very special structure.

Fact 3.2 ([BL00])**.** *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 $0 < \widetilde{\theta} < 1$, $\|F^*_{\theta x}(\widetilde{\theta})\|_{[X,\ell_2^d]_{\widetilde{\theta}}} = \|x\|_{[X,\ell_2^d]_\theta}$.*

Below lemma is the core of the overall argument.

Lemma 3.3 (a quantitative version of a statement from [BL00])**.** *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 2.17, 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,

$$\left\|F^*_{\theta x_1} + F^*_{\theta x_2}\right\|_{\mathcal{F}_2(\theta)} \geqslant \|x_1 + x_2\|_{[X,\ell_2^d]_\theta} \geqslant 2 - \varepsilon, \tag{12}$$

where the first step follows from Fact 2.16, 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 2.9) and the fact that the minimizers are unit, we have:

$$\left\|F^*_{\theta x_1} + F^*_{\theta x_2}\right\|^p_{\mathcal{F}_2(\theta)} + \frac{\left\|F^*_{\theta x_1} - F^*_{\theta x_2}\right\|^p_{\mathcal{F}_2(\theta)}}{K_p(\mathcal{F}_2(\theta))^p} \leqslant \frac{\|2F^*_{\theta x_1}\|^p_{\mathcal{F}_2(\theta)} + \|2F^*_{\theta x_2}\|^p_{\mathcal{F}_2(\theta)}}{2} = 2^p. \tag{13}$$

Combining (12) and (13), we get:

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

Finally, we get:

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

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 3.2, 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, again by Fact 3.2, 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 3.4 (a quantitative version of the statement from [BL00]). *For $x_1, x_2 \in S_{[X,\ell_2^d]_{\theta_1}}$, one has:*

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

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 2.16, the third step is due to Lemma 2.18, and the last step is due to Lemma 3.3. □

The below theorem 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:*

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

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 2.7. The third inequality follows from Lemma 3.4 combined with the estimate $\Lambda(\beta, \gamma) \lesssim \frac{1}{\sqrt{\beta\gamma}}$. The fourth inequality is shown similar 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:*

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

Proof. Denote $A = [X, \ell_2^d]_{2\alpha}$. By Lemma 2.14, 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 2.5, 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 2.7 yields the first two items. The third and fourth items follow from Theorem 7 applied to A. □

4 Approximate Hölder homeomorphisms

In this section, we give an *approximate* version of the exact map given in Section 3. Recall that we have a complex normed space $X = (\mathbb{C}^d, \|\cdot\|_X)$ and the Hilbert space $H = (\ell_2^d)^{\mathbb{C}}$ satisfying $B_X \subseteq B_H \subseteq \mathsf{d} \cdot B_X$ for $\mathsf{d} = \mathrm{poly}(d)$. For a parameters $\alpha, \beta \in (0, 1)$ (which we set later), we define the complex normed spaces

$$A = [X, H]_{\alpha} \qquad Y = [A, H]_{\beta} \qquad \text{and} \qquad Z = [A, H]_{1-\beta}.$$

We will consider an approximate homeomorphism $\Phi\colon Y \to Z$ in order to make the embedding of Theorem 8 in Section 3 algorithmic, and extend it to the whole space (rather than just between unit spheres). In particular, the relaxed conditions will allow us to give a polynomial time (in d) algorithm which can compute this embedding.

In order to see (one of the reasons) we need an approximate guarantee, consider the task of computing the norm of a vector in $Y = [A, H]_\beta$. By definition, $\|x\|_Y$ is the result of an optimization over a subset of $\mathcal{F}$ (i.e., a subset over holomorphic functions defined on $\mathfrak{S}$). In order to optimize over this subset algorithmically, we will optimize over a particular discretization of $\mathcal{F}$ to obtain a $(1+\varepsilon)$-approximation to $\|x\|_Y$ for a parameter $\varepsilon > 0$ which we will chose to be small enough. We leave the specific details of the algorithm computing the interpolations to Section 5, but we now show how an approximation algorithm for computing interpolations will yield an approximate Hölder homeomorphism.

The argument proceeds by first designing a map between thin shells (done in Subsection 4.1), and the extending this map to the whole space (done in Subsection 4.2). In Subsection 4.3, we summarize the discussion of this section and state the necessary properties of the approximate Hölder homeomorphism used in subsequent sections.

4.1 Maps between thin shells

This argument is the constructive version of Theorem 8 in Section 3. We will write the map with respect to a very small parameter $\varepsilon > 0$ which will dictate the approximation guarantee. We note that since $K_\infty(X) \leqslant 1$ and $K_2(H) = 1$, so by Lemma 2.14, the space $A = (\mathbb{C}^d, \|\cdot\|_A)$ is uniformly convex normed space with p-convexity constant $K = K_p(A) \leqslant 1$ for $p = \frac{1}{\alpha} \geqslant 1$. We also note that $B_A \subseteq B_H \subseteq \mathsf{d} B_A$. For a parameter $\beta \in (0,1)$, we will build a map $\varphi = \varphi_\varepsilon\colon Y \to Z$ where $Y = [A, H]_\beta$ and $Z = [A, H]_{1-\beta}$ where $0 \leqslant \varepsilon \lesssim \frac{\beta^2}{p^3}$.

We first specify a map $\varphi_\varepsilon\colon \mathrm{Sh}(Y, \varepsilon) \to \mathrm{Sh}\left(Z, O(\frac{\varepsilon^{1/3}}{\beta^{2/3}})\right)$, where for a normed space X over $\mathbb{C}^d$, the set $\mathrm{Sh}(X, \varepsilon)$ specifies a thin shell around the unit sphere of X:

$$\mathrm{Sh}(X, \varepsilon) = \{x \in \mathbb{C}^d : 1 - \varepsilon \leqslant \|x\|_X \leqslant 1 + \varepsilon\}.$$

Looking ahead, the algorithm for computing φ_ε in Section 5 will run in time polynomial with d and $\frac{1}{\varepsilon}$. For our desired application to ANN, we encourage the reader to think of the setting of parameters as $\beta \approx \frac{1}{\log \mathsf{d}}$ and $p \approx \sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}$. Therefore, while we specify the required restrictions on the parameter settings, we think of these restrictions as easy to attain due to the flexibility in setting ε to be the inverse of a sufficiently high degree polynomial in d.

The map φ_ε is defined as the concatenation of three maps:

$$\varphi_\varepsilon\colon \mathrm{Sh}(Y, \varepsilon) \xrightarrow{F_\varepsilon} \mathcal{F}_2(\beta) \xrightarrow{I} \mathcal{F}_2(1-\beta) \xrightarrow{E} \mathrm{Sh}\left(Z, O(\tfrac{\varepsilon^{1/3}}{\beta^{2/3}}))\right),$$

(see Subsection 2.7 for the formal definition of $\mathcal{F}_2(\beta)$) where:

1. The map $F_\varepsilon\colon \mathrm{Sh}(Y, \varepsilon) \to \mathcal{F}_2(\beta)$ is promised to output for each $x \in \mathrm{Sh}(Y, \varepsilon)$, a function $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}_2(\beta)$ satisfying:

- $\|f(\beta) - x\|_Y \leqslant \varepsilon$, and
- $\max\{\sup_{t\in\mathbb{R}} \|f(it)\|_A, \sup_{t\in\mathbb{R}} \|f(1+it)\|_H\} \leqslant 1+\varepsilon$.

2. The map $I\colon \mathcal{F}_2(\beta) \to \mathcal{F}_2(1-\beta)$ is the identity map which interprets $f \in \mathcal{F}_2(1-\beta)$.

3. The map $E\colon \mathcal{F}_2(1-\beta) \to Z$ is the evaluation map, which on input $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}_2(1-\beta)$, outputs $f(1-\beta) \in Z$.[5]

Specifically, the map $F_\varepsilon\colon \mathrm{Sh}(Y,\varepsilon) \to \mathcal{F}_2(\beta)$ is defined as the output of the algorithm from Section 5, and our arguments in will only use the properties of F_ε listed above. The goal of this subsection is to prove the following lemma.

Lemma 4.1. *The map $\varphi_\varepsilon\colon \mathrm{Sh}(Y,\varepsilon) \to \mathrm{Sh}\left(Z, \frac{200\varepsilon^{1/3}}{\beta^{2/3}}\right)$ satisfies the following two inequalities for every $x, y \in \mathrm{Sh}(Y,\varepsilon)$:*

$$\|\varphi_\varepsilon(x) - \varphi_\varepsilon(y)\|_Z \lesssim \frac{1}{\beta} \cdot \left(\|x-y\|_Y + 5\varepsilon\right)^{1/p}, \tag{14}$$

$$\|x-y\|_Y - 2\varepsilon \lesssim \frac{1}{\beta} \cdot \left(\|\varphi_\varepsilon(x) - \varphi_\varepsilon(y)\|_Z + \frac{1000\varepsilon^{1/3}}{\beta^{2/3}}\right)^{1/p}. \tag{15}$$

Before proving Lemma 4.1 (which we give in Subsection 4.1.1), we give the following claim, which bounds the p-convexity constants for $\mathcal{F}_2(\beta)$ and $\mathcal{F}_2(1-\beta)$.

Claim 4.2. *The spaces $\mathcal{F}_2(\beta)$ and $\mathcal{F}_2(1-\beta)$ are uniformly convex, with p-convexity constants $K_p(\mathcal{F}_2(\beta)), K_p(\mathcal{F}_2(1-\beta)) \leqslant 20$.*

Proof. Let W be the normed space over $\mathbb{C}^d \times \mathbb{C}^d$ given by:

$$\|(x,y)\|_W = \left(\|x\|_A^p + \|y\|_H^p\right)^{1/p}.$$

Note that $\mathcal{F}_2(\beta)$ is isometric to a subspace of $L_2(\mu_\beta, W)$. The isometric embedding proceeds as follows: for any $f\colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}_2(\beta)$, we let $g\colon \partial\overline{\mathfrak{S}} \to \mathbb{C}^d \times \mathbb{C}^d \in L_2(\mu_\beta, W)$ given by

$$g(z) = \begin{cases} (f(z), 0) & \mathrm{Re}(z) = 0 \\ (0, f(z)) & \mathrm{Re}(z) = 1 \end{cases}.$$

Thus, we may use Corollary 6.4 in [MN13] to conclude

$$K_p(\mathcal{F}_2(\beta)) \leqslant K_p(L_2(\mu_\beta, W)) \leqslant \left(\frac{10p}{p-1}\right)^{1-1/p} K_p(W) \leqslant \left(\frac{10p}{p-1}\right)^{1-1/p} \leqslant 20,$$

where the third inequality follows from the fact that $K_p(W) \leqslant 1$ if $K_p(A) \leqslant 1$ and $K_p(H) \leqslant 1$, and the last inequality follows from the fact $p > 1$. Similarly, $\mathcal{F}_2(1-\beta)$ is isometric to a subspace of $L_2(\mu_{1-\beta}, W)$, giving the same bound on the p-convexity constant. □

We will think of φ_ε becoming closer to the Hölder homeomorphism from Theorem 8 mapping unit spheres as $\varepsilon \to 0$. In fact, we may consider the case when $\varepsilon = 0$; while the definition of $F_0\colon S(Y) \to \mathcal{F}_2(\beta)$ can no longer be the output of the algorithm of Section 5, we let F_0 denote the value of the *exact* optimization over $\mathcal{F}$. Using F_0 to specify the map $\varphi_0\colon S(Y) \to S(Z)$ recovers the map from Theorem 8.

[5] We will show that $1 - \frac{200\varepsilon^{1/3}}{\beta^{2/3}} \leqslant \|f(1-\beta)\|_Z \leqslant 1+\varepsilon$.

4.1.1 Proof of Lemma 4.1

We first focus on showing (14). The proof will follow from combining a sequence of claims, which we will combine to obtain the string of inequalities in (18). After that, we turn our attention to (15).

The following two claims follow from the definition of the complex interpolations $Y = [A, H]_\beta$ and $Z = [A, H]_{1-\beta}$, as well as the fact that complex interpolation may be defined as an optimization in $\mathcal{F}$ or $\mathcal{F}_2(\beta)$ (see Subsection 2.5 and Subsection 2.7, specifically, Fact 2.16).

Claim 4.3. *Let $x, y \in \mathbb{C}^d$.*

- *If $f_x, f_y \in \mathcal{F}_2(\beta)$ satisfy $f_x(\beta) = x$ and $f_y(\beta) = y$, then $\|x - y\|_Y \leqslant \|f_x - f_y\|_{\mathcal{F}_2(\beta)}$.*
- *If $f_x, f_y \in \mathcal{F}_2(1-\beta)$ satisfy $f_x(1-\beta) = x$ and $f_y(1-\beta) = y$, then $\|x-y\|_Z \leqslant \|f_x - f_y\|_{\mathcal{F}_2(1-\beta)}$.*

Proof. The claim follows from the definitions of complex interpolation, as well as Fact 2.16. We have that $f_x - f_y \in \mathcal{F}_2(\beta)$ satisfies $(f_x - f_y)(\beta) = x - y$, therefore, $\|x - y\|_Y \leqslant \|f_x - f_y\|_{\mathcal{F}_2(\beta)}$. The same observation occurs for Z and $\mathcal{F}_2(1 - \beta)$. □

Claim 4.4. *For any f in $\mathcal{F}_2(1-\beta)$ and $\mathcal{F}_2(\beta)$, $\|f\|_{\mathcal{F}_2(1-\beta)} \lesssim \frac{1}{\beta}\|f\|_{\mathcal{F}_2(\beta)}$ and $\|f\|_{\mathcal{F}_2(\beta)} \lesssim \frac{1}{\beta}\|f\|_{\mathcal{F}_2(1-\beta)}$.*

Proof. This follows from Lemma 2.18 and the fact that $\Lambda(\beta, 1-\beta) \lesssim \frac{1}{\beta}$. □

Claim 4.5. *For $\varepsilon < \frac{1}{4p}$, let $x, y \in \mathrm{Sh}(Y, \varepsilon)$ and $f_x, f_y \in \mathrm{Sh}(\mathcal{F}, \varepsilon)$ where $\|f_x(\beta) - x\|_Y, \|f_y(\beta) - y\|_Y \leqslant \varepsilon$. Then, $\|f_x - f_y\|^p_{\mathcal{F}_2(\beta)} \leqslant 60^p p \cdot (\|x - y\|_Y + 5\varepsilon)$.*

Proof. By Claim 4.2 and the definition of the p-convexity constant $K_p(\mathcal{F}_2(\beta))$,

$$\begin{aligned}
\frac{\|f_x - f_y\|^p_{\mathcal{F}_2(\beta)}}{20^p} &\leqslant \frac{\|2f_x\|^p_{\mathcal{F}_2(\beta)} + \|2f_y\|^p_{\mathcal{F}_2(\beta)}}{2} - \|f_x + f_y\|^p_{\mathcal{F}_2(\beta)} \\
&\overset{(4.3)}{\leqslant} \frac{\|2f_x\|^p_{\mathcal{F}_2(\beta)} + \|2f_y\|^p_{\mathcal{F}_2(\beta)}}{2} - \|f_x(\beta) + f_y(\beta)\|^p_Y \\
&\leqslant \frac{\|2f_x\|^p_{\mathcal{F}_2(\beta)} + \|2f_y\|^p_{\mathcal{F}_2(\beta)}}{2} - (\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y - \|f_x(\beta) - f_y(\beta)\|_Y)^p \qquad (16)
\end{aligned}$$

where the third inequality follows from the fact that $\|f_x(\beta) + f_y(\beta)\|_Y + \|f_x(\beta) - f_y(\beta)\|_Y \geqslant \|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y$ by applying the triangle inequality twice. We write:

$$\Gamma = 2^{p-1}\left(\|f_x\|^p_{\mathcal{F}_2(\beta)} + \|f_y\|^p_{\mathcal{F}_2(\beta)}\right) \qquad \gamma = \frac{(\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y)^p}{\Gamma}.$$

We note that $\Gamma \leqslant 3^p$, since $f_x, f_y \in \mathrm{Sh}(\mathcal{F}, \varepsilon)$ and $\|f_x(\beta) - x\|_Y, \|f_y(\beta) - y\|_Y \leqslant \varepsilon$ so $f_x, f_y \in \mathrm{Sh}(\mathcal{F}_2(\beta), \varepsilon)$. Additionally, $f_x(\beta), f_y(\beta) \in \mathrm{Sh}(Y, 2\varepsilon)$ since $x, y \in \mathrm{Sh}(Y, \varepsilon)$ and $\|f_x(\beta) - x\|_Y \leqslant \varepsilon$ and $\|f_y(\beta) - y\|_Y \leqslant \varepsilon$ by the definition of F_ε; hence,

$$\gamma \geqslant \frac{2^p(1-2\varepsilon)^p}{2^{p-1}(2(1+\varepsilon)^p)} \geqslant (1 - 3\varepsilon)^p \geqslant 1 - 3\varepsilon p \qquad (17)$$

by Bernoulli's inequality since $3\varepsilon \leqslant 1$, and $\gamma > 0$ since $3\varepsilon p \leqslant 1$. Thus, combining (16) with the bound $\Gamma \leqslant 3^p$ and (17) we have

$$\begin{aligned}\frac{\|f_x - f_y\|^p_{\mathcal{F}_2(\beta)}}{20^p} &\leqslant \Gamma\left(1 - \gamma\left(1 - \frac{\|f_x(\beta) - f_y(\beta)\|_Y}{\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y}\right)^p\right) \\ &\leqslant 3^p\left(1 - (1 - 3\varepsilon p)\left(1 - p \cdot \frac{\|f_x(\beta) - f_y(\beta)\|_Y}{\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y}\right)\right),\end{aligned}$$

where again we used Bernoulli's inequality. Thus, we have:

$$\frac{\|f_x - f_y\|^p_{\mathcal{F}_2(\beta)}}{20^p} \leqslant 3^p\left(p \cdot \frac{\|f_x(\beta) - f_y(\beta)\|_Y}{\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y} + 3\varepsilon p\right)$$

By the triangle inequality, $\|f_x(\beta) - f_y(\beta)\|_Y \leqslant \|x - y\|_Y + 2\varepsilon$. Therefore,

$$\frac{\|f_x - f_y\|^p_{\mathcal{F}_2(\beta)}}{20^p} \leqslant 3^p\left(p \cdot \frac{\|x - y\|_Y}{\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y} + 5\varepsilon p\right) \leqslant 60^p p\left(\|x - y\|_Y + 5\varepsilon\right),$$

since $\|f_x(\beta)\|_Y + \|f_y(\beta)\|_Y \geqslant 2 - 4\varepsilon \geqslant 1$ since $\varepsilon < \frac{1}{4}$. □

Proof of (14) in Lemma 4.1. Combining Claims 4.3, 4.4, and 4.5, we have that for any $x, y \in \mathrm{Sh}(Y, \varepsilon)$,

$$\begin{aligned}\|\varphi(x) - \varphi(y)\|_Z &\overset{(4.3)}{\leqslant} \|f_x - f_y\|_{\mathcal{F}_2(1-\beta)} \overset{(4.4)}{\lesssim} \frac{1}{\beta} \cdot \|f_x - f_y\|_{\mathcal{F}_2(\beta)} \\ &\overset{(4.5)}{\lesssim} \frac{1}{\beta} \cdot (60^p \cdot p)^{1/p}\left(\|x - y\|_Y + 5\varepsilon\right)^{1/p} \lesssim \frac{1}{\beta} \cdot \left(\|x - y\|_Y + 5\varepsilon\right)^{1/p}.\end{aligned} \tag{18}$$

□

We now turn our attention to (15) from Lemma 4.1. We follow a similar approach as above, and apply Claims 4.3 and 4.4 to say:

$$\|x - y\|_Y \leqslant \|f_x(\beta) - f_y(\beta)\|_Y + 2\varepsilon \overset{(4.3)}{\leqslant} \|f_x - f_y\|_{\mathcal{F}_2(\beta)} + 2\varepsilon \overset{(4.4)}{\lesssim} \frac{1}{\beta} \cdot \|f_x - f_y\|_{\mathcal{F}_2(1-\beta)} + 2\varepsilon, \tag{19}$$

where in the first inequality, we used the triangle inequality and fact that $\|f_x(\beta) - x\|_Y \leqslant \varepsilon$ and $\|f_y(\beta) - y\|_Y \leqslant \varepsilon$. It remains to deduce an analogous statement to Claim 4.5 for the spaces $\mathcal{F}_2(1 - \beta)$ and Z. In order to do so, we first show the following claim, which shows that the map φ has co-domain in $\mathrm{Sh}(Z, \frac{200\varepsilon^{1/3}}{\beta^{2/3}})$.

Claim 4.6. *Let $x \in \mathrm{Sh}(Y, \varepsilon)$ and $f_x = F_\varepsilon(x)$. Then,*

$$1 - \frac{200\varepsilon^{1/3}}{\beta^{2/3}} \leqslant \|f_x(1 - \beta)\|_Z \leqslant 1 + \varepsilon.$$

Proof. In order to show the upper bound, note that

$$\|f_x(1-\beta)\|_Z \leqslant \max\left\{\sup_{t\in\mathbb{R}} \|f_x(it)\|_A, \sup_{t\in\mathbb{R}} \|f_x(1+it)\|_H\right\} \leqslant 1+\varepsilon,$$

where the last inequality follows from the properties of F_ε. For the lower bound, let A^* and H^* denote the dual spaces of A and H, respectively, and let $Y^* = [A^*, H^*]_\beta$ and $Z^* = [A^*, H^*]_{1-\beta}$, where we denote $\mathcal{F}^\#$ and $\mathcal{F}_2(\beta)^\#$ as the analogous spaces to $\mathcal{F}$ and $\mathcal{F}_2(\beta)$ for the interpolation Y^*, respectively. We note that Y^* is the dual space to Y and Z^* the dual space to Z.

Recall that $z = f_x(\beta) \in \mathrm{Sh}(Y, 2\varepsilon)$ by the properties of F_ε, so let $z^* \in S(Y^*)$ be the corresponding dual certificate, where $|\langle z, z^*\rangle| \geqslant 1-2\varepsilon$. Let $g\colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}^\#$ satisfy $g(\beta) = z^*$ and $\|g\|_{\mathcal{F}^\#} = 1$ (which exists since $z^* \in S(Y^*)$). We note that for any $t \in \mathbb{R}$, we have $\|g(it)\|_{A^*} \leqslant 1$ and $\|g(1+it)\|_{H^*} \leqslant 1$, and $\|g(1-\beta)\|_{Z^*} \leqslant \|g\|_{\mathcal{F}^\#} = 1$.

Consider the function $h\colon \overline{\mathfrak{S}} \to \mathbb{C}$ given by $h(z) = \langle f_x(z), g(z)\rangle$, and note that h is continuous and holomorphic on $\mathfrak{S}$. Since $\|g(j+it)\| \leqslant 1$ for $j \in \{0,1\}$ and $t \in \mathbb{R}$ with $\|\cdot\| = \|\cdot\|_{A^*}$ when $j = 0$ and $\|\cdot\| = \|\cdot\|_{H^*}$ when $j = 1$,

$$|h(j+it)| \leqslant \left\{ \begin{array}{cc} \|f_x(it)\|_A & j=0 \\ \|f_x(1+it)\|_H & j=1 \end{array} \right. ,$$

so by definition of the space $\mathcal{F}_2(\beta)$,

$$\sum_{j=0,1} \int_{-\infty}^{\infty} |h(j+it)| P(\beta, j+it) dt \leqslant \|f_x\|_{\mathcal{F}_2(\beta)} \leqslant 1+\varepsilon. \tag{20}$$

Since $\|g(1-\beta)\|_{Z^*} \leqslant 1$, we have:

$$\|f_x(1-\beta)\|_Z \geqslant |\langle f_x(1-\beta), g(1-\beta)\rangle| = |h(1-\beta)|.$$

We have that $|h(\beta)| = |\langle z, g(\beta)\rangle| \geqslant 1-2\varepsilon$ by definition, and $|h(z)| \leqslant 1+\varepsilon$ for $z \in \partial\mathfrak{S}$. Thus, consider the sets

$$L = \{t \in \mathbb{R} : |h(it) - h(\beta)| > \delta\} \qquad \text{and} \qquad R = \{t \in \mathbb{R} : |h(1+it) - h(\beta)| > \delta\},$$

for $\delta = \frac{100\varepsilon^{1/3}}{\beta^{2/3}}$. Let $a = \frac{h(\beta)}{|h(\beta)|} \in \mathbb{C}$, and for any $t \in \mathbb{R}$ and $j \in \{0,1\}$, consider the quantity

$$\pi(j,t) = \mathrm{Re}(a) \cdot \mathrm{Re}(h(j+it)) + \mathrm{Im}(a) \cdot \mathrm{Im}(h(j+it)),$$

which measures the magnitude of the projection of $h(j+it)$ onto the direction of $h(\beta)$, interpreted as vectors in $\mathbb{R}^2$. Then, since h is continuous and holomorphic on $\mathfrak{S}$, and (20) holds, we may apply Corollary 2.3 to conclude

$$\sum_{j=0,1} \int_{-\infty}^{\infty} h(j+it) P(\beta, j+it) dt = h(\beta),$$

and therefore, the projections satisfy

$$\sum_{j=0,1} \int_{-\infty}^{\infty} \pi(j,t) P(\beta, j+it) dt = |h(\beta)|.$$

Additionally, since $|h(j+it)| \leqslant 1+\varepsilon$ and $1-2\varepsilon \leqslant |h(\beta)| \leqslant 1+\varepsilon$, by Lemma 4.8, if $t \in L$, then $\pi(t,0) \leqslant 1-\delta^2/4$ and if $t \in R$, then $\pi(t,1) \leqslant 1-\delta^2/4$. Thus, if we let:

$$\alpha = \int_L P(\beta, it)dt + \int_R P(\beta, 1+it)dt,$$

we have

$$1-2\varepsilon \leqslant \alpha(1-\delta^2/4) + (1-\alpha)(1+\varepsilon),$$

and thus $\alpha \leqslant \frac{12\varepsilon}{\delta^2}$. Finally, we have:

$$\begin{aligned}
|h(1-\beta) - h(\beta)| &\leqslant \sum_{j=0,1} \int_{-\infty}^{\infty} |h(j+it) - h(\beta)| P(1-\beta, j+it)dt \\
&\leqslant \delta + \int_L |h(it) - h(\beta)| P(1-\beta, t)dt + \int_R |h(1+it) - h(\beta)| P(1-\beta, t)dt \\
&\overset{(4.4)}{\leqslant} \delta + \left(\frac{2}{\pi^2\beta^2}\right) \cdot 3 \left(\int_L P(\beta, it)dt + \int_R P(\beta, 1+it)dt\right) \\
&\leqslant \delta + \frac{2}{\pi^2\beta^2} \cdot 3 \cdot \alpha \leqslant \tfrac{150\varepsilon^{1/3}}{\beta^{2/3}},
\end{aligned}$$

where the first inequality follows from Corollary 2.3, and we used the fact that $|h(it) - h(\beta)| \leqslant 3$ and $|h(1+it) - h(\beta)| \leqslant 3$ for all $t \in \mathsf{R}$. In other words, we have that $|h(1-\beta)| \geqslant 1-2\varepsilon - \frac{150\varepsilon^{1/3}}{\beta^{2/3}} \geqslant 1 - \frac{200\varepsilon^{1/3}}{\beta^{2/3}}$, which gives the desired lower bound. □

We now note that we may follow the same argument of Claim 4.5 with $\varepsilon < \frac{\beta^2}{800p^3}$, or equivalently, $\frac{200\varepsilon^{1/3}}{\beta^{2/3}} < \frac{1}{4p}$ to conclude the following claim.

Claim 4.7. *For $\varepsilon < \frac{\beta^2}{800p^3}$, let $u, v \in \mathrm{Sh}(Z, \frac{200\varepsilon^{1/3}}{\beta^{2/3}})$ and $f_u, f_v \in \mathrm{Sh}\left(\mathcal{F}_2(1-\beta), \frac{200\varepsilon^{1/3}}{\beta^{2/3}}\right)$, with $f_u(1-\beta) = u$ and $f_v(1-\beta) = v$. Then, $\|f_u - f_v\|_{F_2(1-\beta)}^p \leqslant 60^p p \cdot \left(\|u-v\|_Z + \frac{1000\varepsilon^{1/3}}{\beta^{2/3}}\right)$.*

Combining (19) with Claim 4.7, we conclude:

$$\|x-y\|_Y \overset{(19)}{\lesssim} \frac{1}{\beta} \cdot \|f_x - f_y\|_{\mathcal{F}_2(1-\beta)} + 2\varepsilon \overset{(4.7)}{\lesssim} \frac{1}{\beta} \cdot \left(\|\varphi(x) - \varphi(y)\|_Z + \frac{1000\varepsilon^{1/3}}{\beta^{2/3}}\right)^{1/p} + 2\varepsilon,$$

which gives (15) in Lemma 4.1.

4.1.2 An auxiliary lemma

Lemma 4.8. *For $0 < \varepsilon, \delta < \frac{1}{10}$ with $\delta^2 \geqslant 10\varepsilon$, let $z \in \mathbb{C}$ have $1-\varepsilon \leqslant |z| \leqslant 1+\varepsilon$ and $a = \frac{z}{|z|}$. Suppose $z' \in \mathbb{C}$ with $|z'| \leqslant 1+\varepsilon$ satisfies $|z-z'| \geqslant \delta$, then,*

$$\mathrm{Re}(a) \cdot \mathrm{Re}(z') + \mathrm{Im}(a) \cdot \mathrm{Im}(z') \leqslant 1 - \delta^2/4.$$

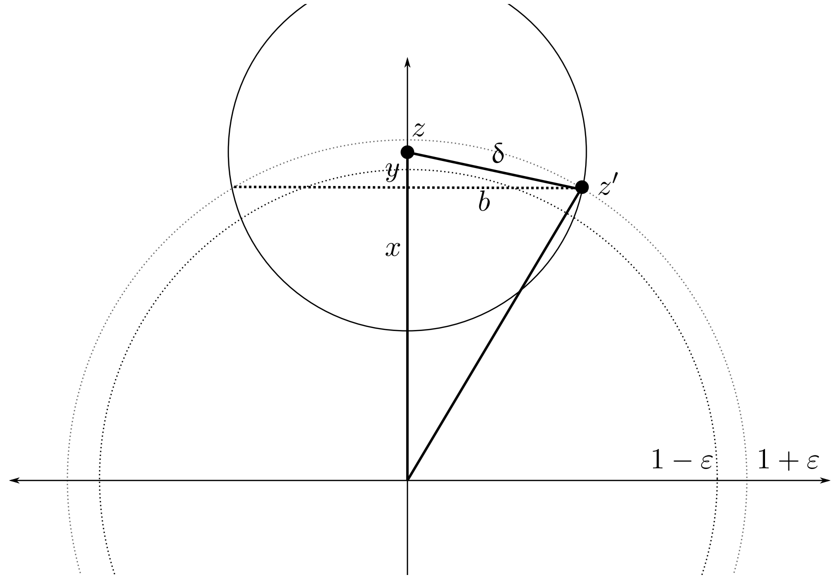

Figure 1: Figure corresponding to the proof of Lemma 4.8. The number $z \in \mathbb{C}$ is shown with $1-\varepsilon \leqslant |z| \leqslant 1+\varepsilon$. Then, the number $z' \in \mathbb{C}$ lies at a distance at least δ from z with $|z'| \leqslant 1+\varepsilon$.

Proof. Consider the scalar $z' \in \mathbb{C}$ with $|z'| \leqslant 1+\varepsilon$, $|z-z'| = \delta$ which maximizes $x = \mathrm{Re}(a)\,\mathrm{Re}(z') + \mathrm{Im}(a)\,\mathrm{Im}(z') \in \mathbb{R}$, and let $y = |z| - x \in \mathbb{R}$ and $b = z' - x \cdot a \in \mathbb{C}$ (see Figure 4.1.2). Note that by the Pythagorean theorem, we have:

$$x^2 + |b|^2 = |z'|^2 \qquad \text{and} \qquad y^2 + |b|^2 = \delta^2.$$

Thus, simplifying the above two equalities, we have $x^2 + \delta^2 - y^2 = (x+y)(x-y) + \delta^2 = |z'|^2$, and therefore,

$$2x = \frac{|z'|^2 - \delta^2}{|z|} + |z| \leqslant \frac{(1+\varepsilon)^2 - \delta^2}{1-\varepsilon} + (1+\varepsilon) \leqslant 2 + \frac{4\varepsilon - \delta^2}{1-\varepsilon} + \varepsilon \leqslant 2 - \frac{\delta^2}{2}$$

□

4.2 Extension to the whole space

From Lemma 4.1, we obtain an approximate Hölder homeomorphism φ mapping a thin shell of a uniformly convex space to another thin shell of a uniformly convex space. We now show how such a map may be extended to the whole space by an approximate radial extension. Similarly to Subsection 4.1, we define the radial extension assuming a map $\ell\colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$ which approximately computes the norm of a vector in an interpolated space. We defer the details of how to compute ℓ to Section 5. The resulting extended map will have similar Hölder guarantees, and in Corollary 4.10, we apply the lemma below to φ_ε.

Lemma 4.9. *Let* $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ *and* $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ *be complex normed spaces,* $C > 0$, $0 \leqslant \varepsilon_X, \varepsilon_1 \leqslant \frac{1}{6}$, *and* $0 < \alpha \leqslant 1$. *Suppose there exists a map* $\varphi\colon \mathrm{Sh}(W_0, \varepsilon_0) \to \mathrm{Sh}(W_1, \varepsilon_1)$ *satisfying:*

$$\begin{aligned} \|\varphi(x) - \varphi(y)\|_{W_1} &\leqslant C \cdot (\|x-y\|_{W_0} + 5\varepsilon_0)^\alpha \\ \|x-y\|_{W_0} - 2\varepsilon_0 &\leqslant C \cdot (\|\varphi(x) - \varphi(y)\|_{W_1} + 5\varepsilon_1)^\alpha \end{aligned}$$

for all $x, y \in \mathrm{Sh}(W_0, \varepsilon_0)$. *Then, for any real* $r \geqslant \alpha$, φ *can be extended to* $\Phi \colon W_0 \to W_1$ *so that* $\|\Phi(x)\|_{W_1} \approx (1 \pm \varepsilon_0)^r (1 \pm \varepsilon_1) \|x\|_{W_0}^r$ *and*

$$\|\Phi(x) - \Phi(y)\|_{W_1} \leqslant \left(2^r C + 2 \cdot 3^{r+1} \max\{1, r\}\right) \cdot (\|x - y\|_{W_0} + 5R\varepsilon_0)^\alpha \left(\|x\|_{W_0}^{r-\alpha} + \|y\|_{W_0}^{r-\alpha}\right), \quad (21)$$

$$\|x - y\|_{W_0} - 2\varepsilon_0 R \leqslant \left(4C + 12 \max\{1, \tfrac{1}{r}\}\right) \cdot (\|\Phi(x) - \Phi(y)\|_{W_1} + 5R^r \varepsilon_1)^\alpha \left(\|x\|_{W_0}^{1-r\alpha} + \|y\|_{W_0}^{1-r\alpha}\right), \quad (22)$$

with $R = \max\{\|x\|_{W_0}, \|y\|_{W_0}\}$.

Proof. We define the map Φ as an approximate radial extension of φ. In particular, consider some vector $x \in \mathbb{C}^d$, and let $\ell(x) \colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$ be a function satisfying

$$(1 - \varepsilon_0/2)\|x\|_{W_0} \leqslant \ell(x) \leqslant (1 + \varepsilon_0/2)\|x\|_{W_0},$$

and note that $\frac{x}{\ell(x)} \in \mathrm{Sh}(W_0, \varepsilon_0)$. We let the function $\Phi \colon W_0 \to W_1$ be:

$$\Phi(x) = \ell(x)^r \cdot \varphi\left(\frac{x}{\ell(x)}\right).$$

For $x, y \in \mathbb{C}^d$, let $a = \frac{x}{\ell(x)}$ and $b = \frac{y}{\ell(y)}$ and let $t = \frac{\ell(y)}{\ell(x)} \leqslant 1$ without loss of generality. By the triangle inequality, we have:

$$\begin{aligned}
\frac{\|\Phi(x) - \Phi(y)\|_{W_1}}{\ell(x)^r} &\leqslant \|\varphi(a) - \varphi(b)\|_{W_1} + \|\varphi(b) - t^r \varphi(b)\|_{W_1} \\
&\leqslant \|\varphi(a) - \varphi(b)\|_{W_1} + (1 + \varepsilon_1)(1 - t^r) \\
&\leqslant C\left(\|a - b\|_{W_0} + 5\varepsilon_0\right)^\alpha + (1 + \varepsilon_1)\max\{1, r\}(1 - t), \qquad (23)
\end{aligned}$$

where the last inequality is by the assumption on φ, and the fact that $(1 - t^r) \leqslant \max\{1, r\}(1 - t)$ when $t \leqslant 1$. In addition,

$$\|a - tb\|_{W_0} \geqslant \|a\|_{W_0} - t\|b\|_{W_0} \geqslant 1 - \varepsilon_0 - t(1 + \varepsilon_0) = (1 - t)(1 + \varepsilon_0) - 2\varepsilon_0, \qquad (24)$$

and therefore,

$$\|a - b\|_{W_0} \leqslant \|a - tb\|_{W_0} + \|tb - b\|_{W_0} \leqslant \|a - tb\|_{W_0} + (1 - t)(1 + \varepsilon_0) \overset{(24)}{\leqslant} 2\|a - tb\|_{W_0} + 2\varepsilon_0. \quad (25)$$

Plugging in (24) and (25) into (23) and recalling that $\varepsilon_1 \leqslant \frac{1}{6}$ implies

$$\begin{aligned}
\frac{\|\Phi(x) - \Phi(y)\|_{W_1}}{\ell(x)^r} &\overset{(25)}{\leqslant} C\left(2\|a - tb\|_{W_0} + 7\varepsilon_0\right)^\alpha + 2\max\{1, r\}(1 - t) \\
&\overset{(24)}{\leqslant} C\left(2\|a - tb\|_{W_0} + 7\varepsilon_0\right)^\alpha + 2\max\{1, r\}\left(\|a - tb\|_{W_0} + 2\varepsilon_0\right) \\
&\leqslant \left(2^\alpha \cdot C + 2\max\{1, r\}\left(\|a - tb\|_{W_0} + 4\varepsilon_0\right)^{1-\alpha}\right)\left(\|a - tb\|_{W_0} + 4\varepsilon_0\right)^\alpha \\
&\leqslant \left(2^\alpha \cdot C + 2 \cdot 3^{1-\alpha} \cdot \max\{1, r\}\right)\left(\|a - tb\|_{W_0} + 4\varepsilon_0\right)^\alpha,
\end{aligned}$$

where we use the fact that $\|a - tb\|_{W_0} + 4\varepsilon_0 \leqslant 1 + \varepsilon_0 + t(1+\varepsilon_0) + 4\varepsilon_0 \leqslant 3$ since $t \leqslant 1$ and $\varepsilon_0 \leqslant \frac{1}{6}$, as well as the fact that $\alpha \leqslant 1$. Finally, since $\ell(x) \leqslant \max\{(1+\varepsilon_0/2)\|x\|_{W_0}, (1+\varepsilon_0/2)\|y\|_{W_0}\}$:

$$\|\Phi(x) - \Phi(y)\|_{W_1} \leqslant \left(2^\alpha C + 2 \cdot 3^{1-\alpha} \max\{1, r\}\right) (\|x - y\|_{W_0} + 4\varepsilon_0 \ell(x))^\alpha \cdot (1 + \tfrac{\varepsilon_0}{2})^{r-\alpha} \left(\|x\|_{W_0}^{r-\alpha} + \|y\|_{W_0}^{r-\alpha}\right),$$

which gives (21), since $(1 + \frac{\varepsilon_0}{2}) \leqslant 2$. For the reverse inequality, we follow in a similar fashion:

$$\begin{aligned}\frac{\|x-y\|_{W_0}}{\ell(x)} &\leqslant \|a - b\|_{W_0} + (1-t)\|b\|_{W_0} \\ &\leqslant C\left(\|\varphi(a) - \varphi(b)\|_{W_1} + 5\varepsilon_1\right)^\alpha + (1-t)(1+\varepsilon_0) + 2\varepsilon_0,\end{aligned} \tag{26}$$

In addition, we have:

$$\|\varphi(a) - t^r \varphi(b)\|_{W_1} \geqslant (1 - \varepsilon_1) - t^r(1+\varepsilon_1) \geqslant (1 - t^r)(1+\varepsilon_1) - 2\varepsilon_1,$$

So,

$$\begin{aligned}\|\varphi(a) - \varphi(b)\|_{W_1} &\leqslant \|\varphi(a) - t^r\varphi(b)\|_{W_1} + \|t^r\varphi(b) - \varphi(b)\|_{W_1} \leqslant \|\varphi(a) - t^r\varphi(b)\|_{W_1} + (1-t^r)(1+\varepsilon_1) \\ &\leqslant 2\|\varphi(a) - t^r\varphi(b)\|_{W_1} + 2\varepsilon_1.\end{aligned} \tag{27}$$

Combining (26) and (27),

$$\begin{aligned}\frac{\|x-y\|_{W_0}}{\ell(x)} &\leqslant C\left(2\|\varphi(a) - t^r\varphi(b)\|_{W_1} + 7\varepsilon_1\right)^\alpha + (1-t)(1+\varepsilon_0) + 2\varepsilon_0 \\ &\leqslant C\left(2\|\varphi(a) - t^r\varphi(b)\|_{W_1} + 7\varepsilon_1\right)^\alpha + 2\max\{1, \tfrac{1}{r}\}\left(\|\varphi(a) - t^r\varphi(b)\|_{W_1} + 2\varepsilon_1\right) + 2\varepsilon_0 \\ &\leqslant \left(2^\alpha C + 2 \cdot 3^{1-\alpha}\max\{1, \tfrac{1}{r}\}\right)\left(\|\varphi(a) - t^r\varphi(b)\|_{W_1} + 4\varepsilon_0\right)^\alpha + 2\varepsilon_0.\end{aligned}$$

Finally, we have:

$$\|x - y\|_{W_0} \leqslant \left(4C + 12\max\{1, \tfrac{1}{r}\}\right)\left(\|\Phi(x) - \Phi(y)\|_{W_1} + 4\varepsilon_0\ell(x)^r\right)^\alpha \left(\|x\|_{W_0}^{1-r\alpha} + \|y\|_{W_0}^{1-r\alpha}\right) + 2\varepsilon_0\ell(x).$$

□

4.3 Summary and necessary subroutines

We now combine Lemma 4.1 with Lemma 4.9 in order to obtain an approximate homeomorphism which will later be used in the design and analysis of ANN algorithms. Recall that $A = (\mathbb{C}^d, \|\cdot\|_A)$ is a uniformly convex normed space with p-convexity constant $K_p(A) \leqslant 1$, and $H = (\mathbb{C}^d, \|\cdot\|_H)$ is the complex Hilbert space $(\ell_2^d)^{\mathbb{C}}$ with $B_A \subseteq B_H \subseteq \mathsf{d} \cdot B_A$, where $\mathsf{d} = \mathrm{poly}(d)$. We define the normed spaces given by the complex interpolation of A and H with parameter $\beta \in (0,1)$, so

$$Y = [A, H]_\beta \qquad \text{and} \qquad Z = [A, H]_{1-\beta}.$$

We summarize the discussion of an approximate homeomorphism $\Phi_\varepsilon \colon Y \to Z$, built by combining Lemma 4.1 with Lemma 4.9 in the following corollary. After that, we collect the necessary assumptions made about existence of two maps which are needed in order to compute Φ_ε. In Section 5, we will show how to implement the necessary steps in an efficient manner.

Corollary 4.10. *For any $R > 0$, $0 \leqslant \varepsilon \leqslant \frac{1}{6p}$, and $r \geqslant \frac{1}{p}$, there exists a map $\Phi_\varepsilon \colon Y \to Z$ satisfying the following conditions:*

- *For every $x \in \mathbb{C}^d$, we have $(1-\varepsilon)^{r+1}\|x\|_Y^r \leqslant \|\Phi_\varepsilon(x)\|_Z \leqslant (1+\varepsilon)^{r+1}\|x\|_Y^r$.*
- *For every $x, y \in \mathbb{C}^d$ with $\|x\|_Y, \|y\|_Y \leqslant R$ and $\|x-y\|_Y \geqslant 5R(\frac{\varepsilon}{\beta})^{1/100}$,*

$$\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z \lesssim \frac{4^r}{\beta} \cdot \|x-y\|_Y^{1/p}\left(\|x\|_Y^{r-1/p} + \|y\|_Y^{r-1/p}\right),$$

additionally, if $\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z \geqslant 5R^r(\frac{\varepsilon}{\beta})^{1/100}$,

$$\|x-y\|_Y \lesssim \left(\frac{1}{\beta} + \frac{1}{r}\right) \cdot \|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z^{1/p}\left(\|\Phi_\varepsilon(x)\|_Z^{\frac{1}{r}-\frac{1}{p}} + \|\Phi_\varepsilon(y)\|_Z^{\frac{1}{r}-\frac{1}{p}}\right).$$

Before presenting the proof, which sets appropriate parameters in order to apply Lemma 4.1 and Lemma 4.9, let us take a moment to comment on the parameters R, ε and r from the corollary above. Recall that p will be high enough so the p-convexity constant $K_p(A) \leqslant 1$, specifically, for our applications, $p \approx \sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}$. In addition, the parameter $\beta \in (0,1)$ defining the interpolation Y and Z will have $\beta \approx \frac{1}{\log \mathsf{d}}$. We encourage the reader to think of $r = 1$ and $R = \mathsf{d}^4$, as these will be the parameters used for our applications. At a high level, we think of R as a large enough radius containing all the points of interest. Then, we set $\varepsilon > 0$ to be the inverse of a large enough polynomial in d, so that $R(\varepsilon/\beta)^{1/100}$ and $R^p(\varepsilon/\beta)^{1/100}$ are small enough numbers. These specific parameter settings allow us to only consider pairs of points $x, y \in \mathbb{C}^d$ whose distance $\|x-y\|_Y$ and $\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z$ is large enough to overcome the additive error terms from Lemma 4.1.

Proof. Let the parameters $\varepsilon_0, \varepsilon_1 > 0$ and $\alpha \in (0,1]$ be given by:

$$\varepsilon_0 = \frac{\varepsilon^3\beta^2}{200^3} \qquad \varepsilon_1 = \varepsilon \qquad \alpha = \frac{1}{p},$$

so that $\varepsilon_0 \leqslant \frac{\beta^2}{800p^3}$, and $\varepsilon_1 = \frac{\varepsilon_0^{1/3}}{200\beta^{2/3}}$. These parameter settings allow us to invoke Lemma 4.9, and we denote $\varphi_{\varepsilon_0} \colon \mathrm{Sh}(Y, \varepsilon_0) \to \mathrm{Sh}(Z, \varepsilon_1)$, as the map satisfying the inequalities from Lemma 4.1. Then, we let $\Phi_\varepsilon \colon Y \to Z$ be the map defined as,

$$\Phi_\varepsilon(x) = \ell(x)^r \varphi_{\varepsilon_0}\left(\frac{x}{\ell(x)}\right)$$

from Lemma 4.9. Since $\varepsilon \geqslant \varepsilon_0$, we have Φ_ε satisfies:

$$(1-\varepsilon)^{r+1}\|x\|_Y^r \leqslant \|\Phi_\varepsilon(x)\|_Z \leqslant (1+\varepsilon)^{r+1}\|x\|_Y^r.$$

In addition, whenever $\|x\|_Y, \|y\|_Y \leqslant R$ and $\|x-y\|_Y \geqslant 5R(\frac{\varepsilon}{\beta})^{1/100}$, then $\|x-y\|_Y \geqslant 5R\varepsilon_0$, and when $\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z \geqslant 5R^r(\frac{\varepsilon}{\beta})^{1/100}$, $\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z \geqslant 5R^p\varepsilon_1$. We obtain the desired inequalities the properties of Φ_ε given in Lemma 4.9. $\square$

Finally, the following lemma asserts that the map $\Phi_\varepsilon \colon Y \to Z$ is close to the optimal homeomorphism $\Phi_0 \colon Y \to Z$ obtained by extending $\varphi_0 \colon S(Y) \to S(Z)$ through Lemma 4.9 with perfect access to norm. In particular, we let $\Phi_0 \colon Y \to Z$ be given by:

$$\Phi_0(x) = \|x\|_Y^r \cdot \varphi_0\left(\frac{x}{\|x\|_Y}\right).$$

First, note that by Theorem 8, we may extend that map $\varphi_0 \colon S(Y) \to S(Z)$ to obtain the map $\Phi_0 \colon Y \to Z$, which is a homeomorphism since $\varphi_0 \colon S(Y) \to S(Z)$ is a homeomorphism. Since φ_0 satisfies the conclusions of Lemma 4.1, then Φ_0 satisfies the conclusions of Corollary 4.10. We record this observation as a corollary, which is almost equivalent to Corollary 4.10, except the conditions on the distances are unnecessary, since $\varepsilon = 0$ in this case.

Corollary 4.11. *There exists a map $\Phi_0 \colon Y \to Z$ satisfying the following conditions:*

- *For every $x \in \mathbb{C}^d$, we have $\|x\|_Y^r = \|\Phi_0(x)\|_Z$.*
- *For every $x, y \in \mathbb{C}^d$, we have:*

$$\|\Phi_0(x) - \Phi_0(y)\|_Z \lesssim \frac{4^r}{\beta} \cdot \|x - y\|_Y^{1/p}\left(\|x\|_Y^{r-\frac{1}{p}} + \|y\|_Y^{r-\frac{1}{p}}\right),$$
$$\|x - y\|_Y \lesssim \left(\frac{1}{\beta} + \frac{1}{r}\right) \cdot \|\Phi_0(x) - \Phi_0(y)\|_Z^{1/p}\left(\|\Phi_0(x)\|_Z^{\frac{1}{r}-\frac{1}{p}} + \|\Phi_0(y)\|_Z^{\frac{1}{r}-\frac{1}{p}}\right).$$

Lastly, we give the following lemma, which states $\Phi_\varepsilon(x) \to \Phi_0(x)$ as $\varepsilon \to 0$. This last lemma justifies the fact that the map $\Phi_\varepsilon \colon Y \to Z$ is indeed an approximate Hölder homeomorphism, even though Φ_ε may not be a bijection.

Lemma 4.12. *For any $R > 0$ and $0 \leqslant \varepsilon \leqslant \frac{1}{6p}$, for any $x \in \mathbb{C}^d$ with $\|x\|_Y \leqslant R$,*

$$\|\Phi_0(x) - \Phi_\varepsilon(x)\|_Z \lesssim \frac{4^r}{\beta}\left(5 \cdot R\left(\frac{\varepsilon}{\beta}\right)^{1/100}\right)^{\frac{1}{p}}\left(\|x\|_Y^{r-\frac{1}{p}}\right).$$

Proof. Consider the map $\Phi'_\varepsilon \colon Y \to Z$ given by:

$$\Phi'_\varepsilon(z) = \begin{cases} \Phi_\varepsilon(z) & z \neq x \\ \Phi_0(z) & z = x \end{cases}.$$

We note that Φ'_ε also satisfies the conclusions in Corollary 4.10, since we φ'_ε given by $\varphi'_\varepsilon(x) = \varphi_0(x)$ and $\varphi'_\varepsilon(z) = \varphi_\varepsilon(z)$ for $z \neq x$ satisfies Lemma 4.1. Consider $y \in \mathbb{C}^d$ with $\|x - y\|_Y = 5R(\frac{\varepsilon}{\beta})^{1/100}$ and $\|y\|_Y \leqslant \|x\|_Y$, then using the triangle inequality and applying Corollary 4.10 twice, we conclude:

$$\begin{aligned}\|\Phi_0(x) - \Phi_\varepsilon(x)\|_Z &\leqslant \|\Phi'_\varepsilon(x) - \Phi'_\varepsilon(y)\|_Z + \|\Phi_\varepsilon(y) - \Phi_\varepsilon(x)\|_H \\ &\lesssim \frac{4^r}{\beta}\left(5R\left(\frac{\varepsilon}{\beta}\right)^{1/100}\right)^{\frac{1}{p}}\left(\|x\|_Y^{r-\frac{1}{p}} + \|y\|_Y^{r-\frac{1}{p}}\right).\end{aligned}$$

□

Necessary subroutines for computing Φ_ε In order to define $\Phi_\varepsilon \colon Y \to Z$, some assumptions were made on the existence of two maps. We collect these assumptions as algorithmic tasks such that given efficient algorithms for these tasks, we can compute Φ_ε satisfying the properties of Corollary 4.10 and Lemma 4.12.

- (Approximately computing norms) Given approximate oracle access to a normed space $W = (\mathbb{C}^d, \|\cdot\|_W)$, as well as the Hilbert space $H = (\ell_2^d)^{\mathbb{C}}$ with $B_W \subseteq B_H \subseteq \mathsf{d} \cdot B_W$, and parameters $\varepsilon > 0$ and $\theta \in (0,1)$, let $W' = [W, H]_\theta$. For each $x \in \mathbb{C}^d$, output a value $\ell(x)$ satisfying:

$$(1-\varepsilon)\|x\|_{W'} \leqslant \ell(x) \leqslant (1+\varepsilon)\|x\|_{W'}.$$

  The proof of Lemma 4.9 assumed access to this kind of map $\ell \colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$ with $W = A$, $\theta = \beta$ and $\varepsilon = \varepsilon_X/2$.

- (Approximately optimal functions) Given approximate oracle access to a normed space $W = (\mathbb{C}^d, \|\cdot\|_W)$ and the Hilbert space $H = (\ell_2^d)^{\mathbb{C}}$ with $B_W \subseteq B_H \subseteq \mathsf{d} \cdot B_W$, and parameters $\varepsilon > 0$ and $\theta \in (0,1)$, let $W' = [W, H]_\theta$. For each $x \in \mathrm{Sh}(W', \varepsilon)$, output a function $f \colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}_2(\theta)$ satisfying:

  1. $\|f(\theta) - x\|_{W'} \leqslant \varepsilon$, and
  2. $\max\{\sup_{t\in\mathbb{R}} \|f(it)\|_W, \sup_{t\in\mathbb{R}} \|f(1+it)\|_H\} \leqslant 1 + \varepsilon$, and

  The description of the map φ in Subsection 4.1 assumed the existence of such a map F_ε with $W = A$, $\theta = \beta$.

For our applications to ANN over a general normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$ in Section 6 and Section 8, we will instantiate the map by letting $A = [X^{\mathbb{C}}, H]_\alpha$ and $Y = [A, H]_\beta$ and $Z = [A, H]_{1-\beta}$. Therefore, while we have oracle access to computing norms $\|\cdot\|_X$, we will need to solve the algorithmic tasks above for the cases $W = A$ and $W = X^{\mathbb{C}}$. In Section 5, we give an algorithm for accomplishing the two algorithmic tasks set forth above and analyze the complexity in terms of the dimension d, as well as the error parameter ε.

5 Computing approximate Hölder homeomorphisms

5.1 High-level overview

The goal of this section is to give polynomial time algorithms for computing various aspects of complex interpolation which completes the description of the approximate Hölder homeomorphism from Section 4. The data structures for ANN over a real normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$ will compute this map, so we will assume oracle access to computing $\|\cdot\|_X$. In particular, we will provide algorithms for the two tasks specified towards the end of Subsection 4.3. We will consider a complex normed space $W = (\mathbb{C}^d, \|\cdot\|_W)$ and the Hilbert space $H = (\ell_2^d)^{\mathbb{C}}$, i.e., the complexification of ℓ_2^d, and assume

$$B_W \subseteq B_H \subseteq \mathsf{d} \cdot B_W, \tag{28}$$

(note that we may compute $\|x\|_H$ since it is isometrically isomorphic to ℓ_2^{2d} by separating the real and imaginary parts of each coordinate; see Subsection 2.4).

At a high level, our algorithm will express the algorithmic tasks from Subsection 4.3 as the optimums of convex programs, which we then solve using various tools from convex optimization. We first set up some notation. For a convex set $K \subseteq \mathbb{R}^m$, and a real number $\delta > 0$ we let $B(K, \delta) = \{y \in \mathbb{R}^m : x \in K \text{ and } \|x - y\|_2 \leqslant \delta\}$, and we abuse notation slightly by letting $B(y, \delta) = B(\{y\}, \delta)$ for $y \in \mathbb{R}^m$. Then, we let $B(K, -\delta) = \{y \in \mathbb{R}^m : B(y, \delta) \subseteq K\}$. We will frequently interpret convex sets $K \subseteq \mathbb{C}^m$ as convex sets of $\mathbb{R}^{2m}$, by separating the real and imaginary parts of the vectors in $\mathbb{C}^m$.

Definition 5.1 (Membership Oracle ($\texttt{MEM}(K)$) [LSV17])**.** *For a convex set $K \subseteq \mathbb{R}^m$, given a vector $y \in \mathbb{R}^m$ and a real number $\delta > 0$, with probability $1 - \delta$, either assert that $y \in B(K, \delta)$ or assert $y \notin B(K, -\delta)$.*

The goal of this section is to solve the following algorithmic task, which we denote $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$, where we will assume access to $\texttt{MEM}(B_W)$, thus, we will measure the complexity of the algorithm for $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$ in the number of calls to $\texttt{MEM}(B_W)$, as well as the error parameter $\delta > 0$ in these calls. In Subsection 5.5, we show how the subsequent algorithms are used in our applications to ANN.

Definition 5.2 (Algorithmic Task $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$)**.** *For a complex normed space $W = (\mathbb{C}^d, \|\cdot\|_W)$ satisfying*

$$B_W \subseteq B_H \subseteq \mathsf{d} \cdot B_W,$$

we want to solve the following algorithmic task. Given access to $\texttt{MEM}(B_W)$, a parameter $\theta \in (0, 1)$, a vector $x \in \mathbb{C}^d$, and an approximation parameter $\varepsilon > 0$, output the representation of a function $f \colon \overline{\mathfrak{S}} \to \mathbb{C}^d \in \mathcal{F}$ (where $\mathcal{F}$ is defined with respect to the couple (W, H)[6]) such that:

- $\|f(\theta) - x\|_{[W,H]_\theta} \leqslant \varepsilon \|x\|_{[W,H]_\theta}$, *and*
- $\|f\|_{\mathcal{F}} = \max\{\sup_{t \in \mathbb{R}} \|f(it)\|_W, \sup_{t \in \mathbb{R}} \|f(1 + it)\|_H\} \leqslant (1 + \varepsilon_1)\|x\|_{[W,H]_\theta}$.

The representation of the function should also have the property that we may compute a $(1 \pm \varepsilon)$-multiplicative approximation to $\|f\|_{\mathcal{F}}$ in $\mathrm{poly}(d/\varepsilon)$ time, and for any $\theta' \in (0, 1)$ we may compute $f(\theta') \in \mathbb{C}^d$ in $\mathrm{poly}(d/\varepsilon)$ time.

In the following section, we will assume that ε is a small enough parameter, which will later be set to $\frac{1}{\mathrm{poly}(d)}$. Before proceeding to present an algorithm for $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$, we give the following simple consequence of being able to solve $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$, which solves the first algorithmic task from Subsection 4.3.

Corollary 5.3. *For any $x \in \mathbb{C}^d$, $\theta \in (0, 1)$ and $\varepsilon > 0$, we may obtain a $(1 \pm \varepsilon)^2$-multiplicative approximation to $\|x\|_{[W,H]_\theta}$ from a call to $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$.*

[6] see Subsection 2.5 for a formal definition of $\mathcal{F}$

Proof. Given $f \in \mathcal{F}$ as an output of $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$, we note that

$$\|f\|_{\mathcal{F}} \geqslant \|f(\theta)\|_{[W,H]_\theta} \geqslant \|x\|_{[W,H]_\theta} - \|f(\theta) - x\|_{[W,H]_\theta} \geqslant (1-\varepsilon)\|x\|_{[W,H]_\theta},$$

where we first used the definition of interpolation, and then we utilized the triangle inequality, as well as the fact that $f(\theta)$ is close to x. The upper bound follows by definition of $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$. Finally, we note that we may compute a $(1 \pm \varepsilon)$-approximation to $\|f\|_{\mathcal{F}}$ in $\mathrm{poly}(d/\varepsilon)$ time. □

For simplicity, when working with a couple (W, H), where $W = (\mathbb{C}^d, \|\cdot\|_W)$ is a complex normed space and $H = (\ell_2^d)^{\mathbb{C}}$ is a Hilbert space satisfying (28) we denote $W_\theta = (\mathbb{C}^d, \|\cdot\|_\theta)$ as the complex normed space given by $W_\theta = [W, H]_\theta$ and $\|\cdot\|_\theta = \|\cdot\|_{[W,H]_\theta}$.

5.2 Discretization of $\mathcal{F}$

We will now give the discretization of $\mathcal{F}$ we optimize over. Specifically, we show a quantitative version of Lemma 4.2.2 from [BL76]. We will write the discretization of $\mathcal{F}$, where $\mathcal{F}$ is defined with respect to the complex normed spaces $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$. In our applications of this discretization to computing $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$, we will set $W_0 = W$ and $W_1 = H$. Assume that W_0 and W_1 are both close to the Hilbert space $H = (\ell_2^d)^{\mathbb{C}}$, i.e., every $x \in \mathbb{C}^d$ satisfies

$$\|x\|_{W_0} \leqslant \|x\|_H \leqslant \mathsf{d}\|x\|_{W_0} \qquad \text{and} \qquad \|x\|_{W_1} \leqslant \|x\|_H \leqslant \mathsf{d}\|x\|_{W_1}. \tag{29}$$

Recall from Subsection 2.5, the space $\mathcal{F}$, defined with respect to W_0 and W_1, is a space over bounded continuous functions $f \colon \overline{\mathfrak{S}} \to \mathbb{C}^d$ which are holomorphic in $\mathfrak{S}$. We will consider a $x \in \mathbb{C}^d$ such that $\|x\|_{[W_0,W_1]_\theta} = 1$.

Let us first introduce some notation. For $f \in \mathcal{F}$ and $\tau \in \mathbb{R}$, we denote:

$$B_{\mathcal{F}}(f, \tau) \stackrel{\text{def}}{=} \max\{\|f(i\tau)\|_{W_0}, \|f(1+i\tau)\|_{W_1}\}, \qquad \text{and} \tag{30}$$

$$B_\infty(f, \tau) \stackrel{\text{def}}{=} \max_{0 \leqslant u \leqslant 1} \|f(u+i\tau)\|_{\ell_\infty}. \tag{31}$$

In addition, for each $f \in \mathcal{F}$, we may view $g_0 \colon \mathbb{R} \to \mathbb{C}^d$ as $g_0(\tau) = f(i\tau)$ and $g_1(\tau) = f(1+i\tau)$. Given these definitions, when the derivatives $\frac{dg_0}{d\tau}$ and $\frac{dg_1}{d\tau}$ exist, we denote:

$$D_\infty(f, \tau) \stackrel{\text{def}}{=} \max\left\{\left\|\frac{dg_0}{d\tau}(\tau)\right\|_{\ell_\infty}, \left\|\frac{dg_1}{d\tau}(\tau)]\right\|_{\ell_\infty}\right\}.$$

The following lemma is a quantitative version of the classical Fejér's theorem [Kat04].

Lemma 5.4. *Let $f \colon \mathbb{R} \to \mathbb{C}$ to be a differentiable $2\pi L$-periodic function. Consider its Fourier series:*

$$f \mapsto \sum_{n \in \mathbb{Z}} a_n e^{\frac{inx}{L}},$$

where

$$a_n = \frac{1}{2\pi L} \cdot \int_{-\pi L}^{\pi L} f(x) e^{-\frac{inx}{L}}\, dx.$$

Then the Césaro partial sums

$$f_M(x) = \sum_{|n| \leqslant M} \left(1 - \frac{n}{M+1}\right) a_n e^{\frac{inx}{L}}$$

satisfy $\|f_M - f\|_{\ell_\infty} \leqslant \varepsilon$ *for* $0 < \varepsilon < 0.1$ *with*

$$M \lesssim \frac{\|f\|_{\ell_\infty} \cdot \|f'\|_{\ell_\infty}^2 \cdot L^2}{\varepsilon^3}.$$

Proof. We can assume, by rescaling, that $L = 1$. Then, one has:

$$f_M = f * F_M,$$

where

$$F_M = \frac{1}{M} \left(\frac{\sin \frac{Mx}{2}}{\sin \frac{x}{2}} \right)^2$$

is the Fejér's kernel. Thus, for every $\delta > 0$, one has:

$$\begin{aligned} |f_M(x) - f(x)| &= \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x-t) - f(x)) F_M(t)\, dt \right| \\ &\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x-t) - f(x)| F_M(t)\, dt \\ &= \frac{1}{2\pi} \int_{|t| \leqslant \delta} |f(x-t) - f(x)| F_M(t)\, dt + \frac{1}{2\pi} \int_{\delta \leqslant |t| \leqslant \pi} |f(x-t) - f(x)| F_M(t)\, dt \\ &\lesssim \delta \|f'\|_{\ell_\infty} + \frac{\|f\|_{\ell_\infty}}{\delta^2 M}, \end{aligned}$$

where the first step follows from $\frac{1}{2\pi} \int_{-\pi}^{\pi} F_M(t)\, dt = 1$, the second step follows from $F_M(t) \geqslant 0$, and the fourth step follows from $\frac{1}{2\pi} \int_{-\pi}^{\pi} F_M(t)\, dt = 1$, and $F_M(t) \lesssim \frac{1}{\delta^2 M}$ for $\delta \leqslant |t| \leqslant \pi$. Substituting $\delta \sim \left(\frac{\|f\|_{\ell_\infty}}{M \|f'\|_{\ell_\infty}} \right)^{1/3}$, we get

$$|f_M(x) - f(x)| \lesssim \left(\frac{\|f\|_{\ell_\infty} \|f'\|_{\ell_\infty}^2}{M} \right)^{1/3},$$

which implies the desired bound. □

Claim 5.5. *For every* $x \in \mathbb{C}^d$ *with* $\|x\|_{[W_0, W_1]_\theta} = 1$ *and every* $\varepsilon > 0$, *there exists* $f_x \in \mathcal{F}$ *such that:*

- $f_x(\theta) = x$,
- *for every* $\tau \in \mathbb{R}$, $B_{\mathcal{F}}(f_x, \tau) \leqslant 1 + \varepsilon$ *and* $B_\infty(f_x, \tau) \lesssim \mathsf{d}$;

Proof. By definition of $\|x\|_{[W_0,W_1]_\theta}$ from (8), let $f_x \in \mathcal{F}$ be the function with $f_x(\theta) = x$ and $\|f_x\|_{\mathcal{F}} \leqslant \|x\|_{[W_0,W_1]_\theta} + \varepsilon \leqslant 1+\varepsilon$. In addition, since $\|f_x\|_{\mathcal{F}} = \sup_{\tau\in\mathbb{R}} B_{\mathcal{F}}(f_x,\tau)$, we obtain that $B_{\mathcal{F}}(f_x,\tau) \leqslant 1+\varepsilon$. Finally, we note that f_x is holomorphic on $\mathfrak{S}$, continuous on $\overline{\mathfrak{S}}$, and bounded, so by the Hadamard three-lines theorem, $B_\infty(f_x,\tau) \leqslant \sup_{\substack{u\in(0,1)\\ \tau\in\mathbb{R}}} \|f(u+i\tau)\|_H \leqslant \sup_{\tau\in\mathbb{R}} \max\{\|f_x(i\tau)\|_H, \|f_x(1+i\tau)\|_H\} \leqslant \mathsf{d}$. □

Claim 5.6. *For every $x \in \mathbb{C}^d$ with $\|x\|_{[W_0,W_1]_\theta} = 1$ and every $\varepsilon > 0$, there exists $f_x^{(2)} \in \mathcal{F}$ such that:*

- $\|f_x^{(2)}(\theta) - x\|_{\ell_\infty} \lesssim \varepsilon$, *and*
- *for every $\tau \in \mathbb{R}$, $B_{\mathcal{F}}(f_x^{(2)},\tau) \leqslant 1+\varepsilon$, $B_\infty(f_x^{(2)},\tau) \lesssim \mathsf{d}$, and $D_\infty(f_x^{(2)},\tau) \lesssim \frac{\mathsf{d}^2\cdot\Lambda(\theta,\theta)}{\varepsilon}$.*

Proof. Let $f_x \in \mathcal{F}$ be from Claim 5.5 for the vector $x \in \mathbb{C}^d$ and ε. For $\sigma > 0$ let

$$f_x^{(2)}(z) \stackrel{\text{def}}{=} \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} f_x(z+ig). \tag{32}$$

We note the bounds on $B_{\mathcal{F}}(f_x^{(2)},\tau)$ and $B_\infty(f_x^{(2)},\tau)$ are immediate from Jensen's inequality and convexity of $\|\cdot\|\colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$. In order to bound $\|f_x^{(2)}(\theta) - x\|_{\ell_\infty}$, we have

$$\begin{aligned} f_x^{(2)}(\theta) - f_x(\theta) &\stackrel{(2.3)}{=} \int_{\partial\mathfrak{S}} (f_x^{(2)}(z) - f_x(z))\, d\mu_\theta(z) \stackrel{(32)}{=} \int_{\partial\mathfrak{S}} \left(\mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} f_x(z+ig) - f_x(z)\right) d\mu_\theta(z) \\ &= \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} \int_{\partial\mathfrak{S}} (f_x(z+ig) - f_x(z))\, d\mu_\theta(z) \\ &= \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} \int_{\partial\mathfrak{S}} f_x(z)\Big(P(\theta, z-ig) - P(\theta,z)\Big)\, dz, \end{aligned}$$

where we used the definition of $\mu_\theta(z)$ from Subsection 2.2 in the last line. Therefore,

$$\begin{aligned} \|f_x^{(2)}(\theta) - f_x(\theta)\|_{\ell_\infty} &\leqslant \left(\sup_{\tau\in\mathbb{R}} B_\infty(f_x,\tau)\right) \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} \int_{\partial\mathfrak{S}} |P(\theta,z) - P(\theta,z-ig)|\, dz \\ &\lesssim \mathsf{d} \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} \int_{\partial\mathfrak{S}} |P(\theta,z) - P(\theta,z-ig)|\, dz. \end{aligned} \tag{33}$$

Note that for $\theta \in (0,1)$, $P(\theta,\cdot)$ is symmetric around zero and unimodal, when $g \geqslant 0$, we have $P(\theta,z) \geqslant P(\theta,z-ig)$ when $\mathrm{Im}(z) \leqslant \frac{g}{2}$ and $P(\theta,z) \leqslant P(\theta,z-ig)$ when $\mathrm{Im}(z) \geqslant \frac{g}{2}$. So we have that when $g \geqslant 0$,

$$\begin{aligned} \int_{\partial\mathfrak{S}} |P(\theta,z) - P(\theta,z-ig)|dz &= \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\leqslant\frac{g}{2}}} (P(\theta,z) - P(\theta,z-ig))dz - \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\geqslant\frac{g}{2}}} (P(\theta,z) - P(\theta,z-ig))dz \\ &= \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\leqslant\frac{g}{2}}} P(\theta,z)dz - \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\leqslant-\frac{g}{2}}} P(\theta,z)dz - \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\geqslant\frac{g}{2}}} P(\theta,z)dz + \int\limits_{\substack{z\in\partial\mathfrak{S}\\ \mathrm{Im}(z)\geqslant-\frac{g}{2}}} P(\theta,z)dz \\ &= 2\int\limits_{\substack{z\in\partial\mathfrak{S}\\ |\mathrm{Im}(z)|\leqslant\frac{g}{2}}} P(\theta,z)dz. \end{aligned} \tag{34}$$

The case when $g \leqslant 0$ is symmetric, thus, we may combine (33) and (34) to conclude

$$\|f_x^{(2)}(\theta) - f_x(\theta)\|_{\ell_\infty} \lesssim \mathsf{d} \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} \int_{\substack{z\in\partial\mathfrak{S} \\ |\operatorname{Im}(z)|\leqslant\frac{|g|}{2}}} P(\theta, z) \lesssim \mathsf{d}\Lambda(\theta,\theta) \mathop{\mathrm{E}}_{g\sim N(0,\sigma^2)} |g| \lesssim \mathsf{d}\Lambda(\theta,\theta)\sigma \leqslant \varepsilon$$

when $\sigma \lesssim \frac{\varepsilon}{\mathsf{d}\cdot\Lambda(\theta,\theta)}$, where we used the fact that $P(\theta, z) \lesssim \Lambda(\theta,\theta)$ when $z \in \partial\mathfrak{S}$ and $\theta \in (0,1)$. Now let us upper bound $D_\infty(f_x^{(2)}, \tau)$. Denote $p_\sigma(t)$ the p.d.f. of $N(0,\sigma^2)$. In addition, if we let $g_0, g_1, g_0^{(2)}, g_1^{(2)} : \mathbb{R} \to \mathbb{C}^d$ have $g_j(\tau) = f_x(j + i\tau)$ and $g_j^{(2)}(\tau) = f_x^{(2)}(j + i\tau)$ for $j = 0, 1$, we have $g_j^{(2)} = g_j * p_\sigma(\tau)$. Thus, we have $\frac{dg_j^{(2)}}{d\tau}(\tau) = g_j * p'_\sigma(\tau)$,

$$\left\| \frac{d}{d\tau}[g_j^{(2)}(\tau)] \right\|_{\ell_\infty} = \|g_j * p'_\sigma\|_{\ell_\infty} \lesssim \frac{\mathsf{d}}{\sigma} \lesssim \frac{\mathsf{d}^2\Lambda(\theta,\theta)}{\varepsilon}.$$

□

Claim 5.7. *For every $x \in \mathbb{C}^d$ with $\|x\|_{[W_0,W_1]_\theta} = 1$ and every $\varepsilon > 0$, there exists $f_x^{(3)} \in \mathscr{F}$ such that:*

- $\|f_x^{(3)}(\theta) - x\|_{\ell_\infty} \lesssim \varepsilon$,
- *for every $\tau \in \mathbb{R}$, $B_{\mathscr{F}}(f_x^{(3)}, \tau) \leqslant (1+\varepsilon) \cdot e^{\frac{\varepsilon(1-\tau^2)}{\mathsf{d}}}$ and $B_\infty(f_x^{(3)}, \tau) \lesssim \mathsf{d} \cdot e^{-\frac{\varepsilon\tau^2}{\mathsf{d}}}$, and*
- *for every $\tau \in \mathbb{R}$, one has: $D_\infty(f_x^{(3)}, \tau) \lesssim \frac{\mathsf{d}^2\cdot\Lambda(\theta,\theta)}{\varepsilon} \cdot e^{-\frac{\varepsilon\tau^2}{\mathsf{d}}} + \varepsilon|\tau| \cdot e^{-\frac{\varepsilon\tau^2}{\mathsf{d}}}$.*

Proof. We may consider $f_x^{(2)} \in \mathscr{F}$ from Claim 5.6 and set

$$f_x^{(3)}(z) = e^{\frac{\varepsilon z^2}{\mathsf{d}}} \cdot f_x^{(2)}(z).$$

All the desired properties are immediate to check. □

Claim 5.8. *For every $x \in \mathbb{C}^d$ with $\|x\|_{[W_0,W_1]_\theta} = 1$ and every $\varepsilon > 0$, there exists $f_x^{(4)} \in \mathscr{F}$ such that:*

- $\|f_x^{(4)}(\theta) - x\|_{\ell_\infty} \lesssim \varepsilon$,
- *for every $\tau \in \mathbb{R}$, $B_{\mathscr{F}}(f_x^{(4)}, \tau) \leqslant 1 + O(\varepsilon)$, $B_\infty(f_x^{(4)}, \tau) \lesssim \mathsf{d}$, $D_\infty(f_x^{(4)}, \tau) \lesssim \frac{\mathsf{d}^2\cdot\Lambda(\theta,\theta)}{\varepsilon}$, and*
- *$f_x^{(4)}$ is $2\pi iL$-periodic for*

$$L \lesssim \sqrt{\frac{\mathsf{d} \cdot \log\frac{\mathsf{d}}{\varepsilon}}{\varepsilon}}.$$

Proof. We take $f_x^{(3)}$ from Claim 5.7, and set:

$$f_x^{(4)}(z) = \sum_{k\in\mathbb{Z}} f_x^{(3)}(z + 2\pi iLk).$$

All the desired properties are immediate to check. □

Denote for $n \in \mathbb{Z}$ the n-th Fourier coefficient:

$$a_n = \frac{1}{2\pi L} \int\limits_{-\pi L}^{\pi L} f_x^{(4)}(i\tau) e^{-\frac{in\tau}{L}}\, d\tau \in \mathbb{C}^d.$$

Claim 5.9. *For every $0 \leqslant \widetilde{\theta} \leqslant 1$, one has:*

$$\frac{1}{2\pi L} \int\limits_{-\pi L}^{\pi L} f_x^{(4)}(\widetilde{\theta} + i\tau) e^{-\frac{n(\widetilde{\theta}+i\tau)}{L}}\, d\tau = a_n.$$

Proof. First, let us show that the left-hand side, which we denote by $A(\widetilde{\theta})$, does not change when $\widetilde{\theta}$ is varied in $(0;1)$. Consider the following contour on the complex plane for $k \geqslant 1$: $\theta_1 - i\pi Lk$ to $\theta_1 + i\pi Lk$ to $\theta_2 + i\pi Lk$ to $\theta_2 - i\pi Lk$ to $\theta_1 - i\pi Lk$. On the one hand, the integral of $f_x^{(4)}(z)e^{-\frac{nz}{L}}$ over it equals to zero. On the other hand, it is equal to $k \cdot (A(\theta_1) - A(\theta_2))$ plus a term that is independent of k. Since k is arbitrary, we get $A(\theta_1) = A(\theta_2)$. On the other hand, for $0 < \theta < 1$, we have $A(\theta) = A(0) = A(1) = a_n$, since the function $f_x^{(4)}(z)$ converges uniformly to the corresponding boundary value as the real part of z converges to zero or to one. $\square$

Claim 5.10. *One has:*

- $\|a_n\|_{\ell_\infty} \leqslant \min\{1, e^{-n/L}\} \cdot \|f_x^{(4)}\|_{\ell_\infty} \lesssim \min\{1, e^{-n/L}\} \cdot \mathsf{d}$.
- *If we denote $f_x^{(5)} \in \mathcal{F}$ by*

$$f_x^{(5)}(z) = \sum_{|n| \leqslant M} \frac{M+1-|n|}{M+1} \cdot a_n e^{\frac{nz}{L}},$$

then $\|f_x^{(5)}(z) - f_x^{(4)}(z)\|_{\ell_\infty} \leqslant \widetilde{\varepsilon}$ for

$$M \lesssim \frac{\mathsf{d}^5 \cdot L^2 \cdot \Lambda(\theta, \theta)^2}{\widetilde{\varepsilon}^3}.$$

Proof. The bound $\|a_n\|_{\ell_\infty} \leqslant \|f_x^{(4)}\|_{\ell_\infty} \lesssim \mathsf{d}$ is trivial. To show that $\|a_n\|_{\ell_\infty} \leqslant e^{-n/L} \cdot \|f_x^{(4)}\|_{\ell_\infty}$, we just use Claim 5.9 with $\widetilde{\theta} = 1$:

$$a_n = \frac{1}{2\pi L} \int\limits_{-\pi L}^{\pi L} f_x^{(4)}(1 + i\tau) e^{-\frac{n(1+i\tau)}{L}}\, d\tau = e^{-n/L} \cdot \frac{1}{2\pi L} \int\limits_{-\pi L}^{\pi L} f_x^{(4)}(1 + i\tau) e^{-\frac{in\tau}{L}}\, d\tau.$$

$f_x^{(5)}(z)$ converges to $f_x^{(4)}(z)$ uniformly in ℓ_∞ for $\operatorname{Re} z = 0$ by Lemma 5.4. But due to Claim 5.9, it is also the case for $\operatorname{Re} z = 1$. We get the required bound on M from the conclusions of Lemma 5.4 and Claim 5.8. For $0 < \operatorname{Re} z < 1$, we simply use the Hadamard three-line theorem. $\square$

Lemma 5.11. *For every $x \in \mathbb{C}^d$ with $\|x\|_{[W_0,W_1]_\theta} = 1$ and every $\varepsilon > 0$, there exists a function $\tilde{f}_x \in \mathcal{F}$ with*

$$\tilde{f}_x(z) = e^{\frac{z^2}{M}} \sum_{q \in \mathbb{Q}_M} v_q \cdot e^{qz},$$

where $\mathbb{Q}_M = \{\frac{s}{L} : |s| \leqslant ML\}$ and $v_q \in \mathbb{C}^d$ for all $q \in \mathbb{Q}_M$ satisfying:

- $L \lesssim M = \mathrm{poly}(d/\varepsilon)$*;*
- $\|v_q\|_H \lesssim \min\{1, e^{-q}\} \cdot M$*;*
- $\|\tilde{f}_x\|_{\mathcal{F}} \leqslant 1 + \frac{\varepsilon}{\mathsf{d}^2}$*;*
- $\|\tilde{f}_x(\theta) - x\|_{[W_0,W_1]_\theta} \leqslant \frac{\varepsilon}{\mathsf{d}^2}$*.*

Proof. We take $f_x^{(5)}(z)$ from Claim 5.10 with $\widetilde{\varepsilon} \lesssim \mathrm{poly}(\varepsilon, \frac{1}{\mathsf{d}})$. This implies that $\|f_x^{(5)} - f_x^{(4)}\|_{\mathcal{F}} \lesssim \dfrac{\varepsilon}{\mathsf{d}^2}$ and $\|f_x^{(5)}(\theta) - x\|_{[W_0,W_1]_\theta} \leqslant \frac{\varepsilon}{\mathsf{d}^2}$. Finally, setting $\tilde{f}_x(z) = e^{\frac{z^2}{M}} \cdot f_x^{(5)}(z)$, and a similar argument to that of Claim 5.7, the required properties hold. ☐

5.3 Convex program for $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$

In this subsection, we present a convex program for solving $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$ and argue that a good enough solution to this program can be a valid response for $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$.

Given any vector $x \in \mathbb{C}^d$, we may compute the vector $y \in \mathbb{C}^d$ with $y = \frac{x}{\|x\|_H}$. We note that since $\|x\|_H \leqslant \|x\|_\theta \leqslant \mathsf{d}\|x\|_H$ from (28), we have $1 \leqslant \|y\|_\theta \leqslant \mathsf{d}$, and recall that $\mathsf{d} = \mathrm{poly}(d)$. Thus, we may assume that calls to $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$ always have $x \in \mathbb{C}^d$ satisfying

$$\|x\|_H = 1 \qquad\qquad \text{and} \qquad\qquad 1 \leqslant \|x\|_\theta \leqslant \mathsf{d}. \tag{35}$$

Given $x \in \mathbb{C}^d$ satisfying (35), we will define a convex program $\texttt{Rep}(x, \theta, \varepsilon; W)$ which takes a parameter $\varepsilon > 0$ and has size $\mathrm{poly}(d/\varepsilon)$ whose optimum will give a valid response for $\texttt{ApproxRep}(x, \theta, 10\varepsilon; W)$. Recall the parameter $M = \mathrm{poly}(d, \frac{1}{\varepsilon}, \Lambda(\theta, \theta))$, as well as the definition of the set $\mathbb{Q}_M \subseteq \mathbb{R}$ from Lemma 5.11, and let $N \lesssim \frac{M^4 \mathsf{d}}{\varepsilon}$ for a large enough constant.[7] For $j \in \{0, 1\}$, consider the subset:

$$\mathbb{D}_N^{(j)} \stackrel{\text{def}}{=} \left\{ j + \frac{is}{N} \in \partial\mathfrak{S} : s \in \mathbb{Z}, |s| \leqslant MN \right\}.$$

Let the sequence of vectors $V = (v_q \in \mathbb{C}^d : q \in \mathbb{Q}_M)$ define a map $f_V \in \mathcal{F}$ given by:

$$f_V(z) \stackrel{\text{def}}{=} e^{\frac{z^2}{M}} \sum_{q \in \mathbb{Q}_M} v_q \cdot e^{qz}. \tag{36}$$

[7] see (6) for the definition of $\Lambda(\theta_1, \theta_2)$, and note that $\Lambda(\theta, \theta) = \mathrm{poly}(d)$ for $\theta \in (\frac{1}{\mathrm{poly}(d)}, 1 - \frac{1}{\mathrm{poly}(d)})$.

The convex program $\texttt{Rep}(x, \theta, \varepsilon; W)$ is given by:

$$\texttt{Rep}(x, \theta, \varepsilon; W) = \begin{cases} \min_{\substack{V \in (\mathbb{C}^d)^{|\mathbb{Q}_M|} \\ \alpha \in \mathbb{R}^{\geqslant 0}}} \alpha & \\ \text{s.t} & \text{i.} \quad \forall z \in \mathbb{D}_N^{(0)}, \|f_V(z)\|_W \leqslant \alpha + \varepsilon \\ & \text{ii.} \quad \forall z \in \mathbb{D}_N^{(1)}, \|f_V(z)\|_H \leqslant \alpha + \varepsilon \\ & \text{iii.} \quad \forall q \in \mathbb{Q}_M, \|v_q\|_H \max\{e^q, 1\} \leqslant 2M\mathsf{d} \\ & \text{iv.} \quad \|f_V(\theta) - x\|_H \leqslant \frac{2\varepsilon}{\mathsf{d}} \end{cases}. \tag{37}$$

In the language of Grötschel, Lovász and Schrijver [GLS12], we will argue that solving the weak optimization problem of $\texttt{Rep}(x, \theta, \varepsilon; W)$ satisfies the requirements of $\texttt{ApproxRep}(x, \theta, 8\varepsilon; W)$. After that, we address the problem of computing $\texttt{Rep}(x, \theta, \varepsilon; W)$.

Lemma 5.12. *For $x \in \mathbb{C}^d$ and $\varepsilon > 0$, and $\theta \in (0,1)$ with $\Lambda(\theta, \theta) \leqslant \mathrm{poly}(d)$, let $(V, \alpha) \in (\mathbb{C}^d)^{|\mathbb{Q}_M|} \times \mathbb{R}$ be a feasible solution to $\texttt{Rep}(x, \theta, \varepsilon; W)$, where the optimum of $\texttt{Rep}(x, \theta, \varepsilon; W)$ is at least $\alpha - 5\varepsilon$. Then $f_V \in \mathcal{F}$ is a valid output of $\texttt{ApproxRep}(x, \theta, 8\varepsilon; W)$.*

Before giving the proof of Lemma 5.12, we give some discussion as well as a sequence of claims from which Lemma 5.12 will easily follow.

Consider the unit vector $a = \frac{x}{\|x\|_\theta}$,[8] and let $f_a \in \mathcal{F}$ be an *optimal representative* for a at θ in $\mathcal{F}$. In other words, $f_a(\theta) = a$ and $\|f_a\|_{\mathcal{F}} = 1$. Applying Lemma 5.11, there exists some $\tilde{f}_a \in \mathcal{F}$ such that:

- $\tilde{f}_a(z) = e^{\frac{z^2}{M}} \sum_{q \in \mathbb{Q}_N} v_q e^{qz}$, and for all $q \in \mathbb{Q}_M$, $v_q \in \mathbb{C}^d$ with $\|v_q\|_H \cdot \max\{e^q, 1\} \leqslant M$.
- $\|\tilde{f}_a\|_{\mathcal{F}} \leqslant 1 + \frac{\varepsilon}{\mathsf{d}^2}$ and $\|\tilde{f}_a(\theta) - a\|_H \leqslant \frac{\varepsilon}{\mathsf{d}^2}$.

Therefore, let the function $\tilde{f}_x \in \mathcal{F}$ be

$$\tilde{f}_x(z) \stackrel{\text{def}}{=} \|x\|_\theta \tilde{f}_a(z) \tag{38}$$

which satisfies,

$$\|\tilde{f}_x(\theta) - x\|_H \leqslant \|x\|_\theta \cdot \|\tilde{f}_a(\theta) - a\|_H \leqslant \|x\|_\theta \cdot \frac{\varepsilon}{\mathsf{d}^2} \leqslant \frac{\varepsilon}{\mathsf{d}}. \tag{39}$$

In addition, we have:

$$\|\tilde{f}_x(\theta) - x\|_\theta \leqslant \mathsf{d}\|\tilde{f}_x(\theta) - x\|_H \leqslant \frac{\varepsilon}{\mathsf{d}} \cdot \|x\|_\theta \tag{40}$$

$$\|x\|_\theta \|v_q\| \cdot \max\{e^q, 1\} \leqslant M\mathsf{d} \qquad \text{for all } q \in \mathbb{Q}_M, \tag{41}$$

$$\|\tilde{f}_x\|_{\mathcal{F}} \leqslant \|x\|_\theta \cdot \|\tilde{f}_a\|_{\mathcal{F}} \leqslant \left(1 + \frac{\varepsilon}{\mathsf{d}^2}\right) \|x\|_\theta. \tag{42}$$

We note that the above facts imply that if $(V, \alpha) \in (\mathbb{C}^d)^{|\mathbb{Q}_M|} \times \mathbb{R}$, where $V = (v_q : q \in \mathbb{Q}_M)$ and the vectors v_q define $\tilde{f}_x \in \mathcal{F}$ according to (36), and $\alpha = \|\tilde{f}_x\|_{\mathcal{F}}$, then (V, α) is a feasible solution for $\texttt{Rep}(x, \theta, \varepsilon; W)$.

[8] algorithmically, we do not have access to this vector

Claim 5.13. *Suppose $V = (v_q : q \in \mathbb{Q}_M), \alpha \in \mathbb{R}^d$ defines a feasible solution for* $\mathtt{Rep}(x, \theta, \varepsilon; W)$*, and suppose $z = j + i\tau \in \partial\mathfrak{S}$ with $|\tau| \geqslant M$, then $\|f_V(z)\|_H \leqslant \frac{1}{2\mathsf{d}}$.*

Proof. We simply follow the computations, using the third constraint of (37):

$$\begin{aligned}\|f_V(z)\|_H &\leqslant e^{\frac{j^2-\tau^2}{M}} \left(\sum_{q\in\mathbb{Q}_M} \|v_q\|_H \cdot e^{qj}\right)\\ &\leqslant e^{-\frac{\tau^2-1}{M}} \left(|\mathbb{Q}_M| \cdot 2M\mathsf{d}\right) \leqslant \frac{1}{2\mathsf{d}},\end{aligned}$$

when $\tau \geqslant M \gg \sqrt{M\log(d)\log(M)}$, since $\mathsf{d} = \mathrm{poly}(d)$. $\square$

Corollary 5.14. *Let $V = (v_q : q \in \mathbb{Q}_M), \alpha \in \mathbb{R}$ be a feasible solution for* $\mathtt{Rep}(x, \theta, \varepsilon; W)$*. We have that:*

$$\|f_V\|_{\mathcal{F}} = \max\left\{\sup_{|\tau|\leqslant M} \|f_V(i\tau)\|_W, \sup_{|\tau|\leqslant M} \|f_V(1+i\tau)\|_H\right\}.$$

Proof. From the fourth constraint of (37), $\|f_V(\theta) - x\|_\theta \leqslant \mathsf{d}\|f_V(\theta) - x\|_H \leqslant 2\varepsilon$. Therefore, we have $\|f_V\|_{\mathcal{F}} \geqslant \|f_V(\theta)\|_\theta \geqslant \|x\|_\theta - 2\varepsilon \geqslant 1 - 2\varepsilon > \frac{1}{2}$ for small enough $\varepsilon < \frac{1}{4}$. Since by Claim 5.13, any $|\tau| \geqslant M$ satisfies $\|f_V(i\tau)\|_W \leqslant \mathsf{d}\|f_V(i\tau)\|_H \leqslant \frac{1}{2}$ and $\|f_V(1+i\tau)\|_H \leqslant \frac{1}{2\mathsf{d}}$, we must have

$$\|f_V\|_{\mathcal{F}} = \max\left\{\sup_{|\tau|\leqslant M} \|f_V(i\tau)\|_W, \sup_{|\tau|\leqslant M} \|f_V(1+i\tau)\|_H\right\}.$$

$\square$

Let us write $f_V : \overline{\mathfrak{S}} \to \mathbb{C}^d$ as $f_V = (g_1(z), \ldots, g_d(z))$, where $g_k : \overline{\mathfrak{S}} \to \mathbb{C}$ is given by the k-th coordinate of f_V. Then, taking derivatives, all $k \in [d]$ satisfy

$$g_k'(z) = e^{z^2/M} \sum_{q\in\mathbb{Q}_M} (v_q)_k e^{qz} \left(q + \frac{2z}{M}\right).$$

Thus, when (V, α) is feasible in $\mathtt{Rep}(x, \theta, \varepsilon; W)$, the third constraint of (37) implies that for each $z = j + i\tau \in \partial\mathfrak{S}$ with $|\tau| \leqslant M$ and every $k \in [d]$,

$$\begin{aligned}|g_k'(z)| &\leqslant e \cdot e^{-\tau^2/M} \sum_{q\in\mathbb{Q}_M} |(v_q)_k \cdot e^q|(q+2)\\ &\leqslant e|\mathbb{Q}_M| \cdot 2M\mathsf{d} \cdot (M+2) \lesssim M^4\mathsf{d}. \qquad (43)\end{aligned}$$

Claim 5.15. *For large enough $N \lesssim \frac{M^4\mathsf{d}^3}{\varepsilon} = \mathrm{poly}(d/\varepsilon)$, we have that any (V, α) which is feasible for* $\mathtt{Rep}(x, \theta, \varepsilon; W)$ *satisfies*

$$\|f_V\|_{\mathcal{F}} \leqslant \max\left\{\sup_{z\in\mathbb{D}_N^{(0)}} \|f_V(z)\|_W, \sup_{z\in\mathbb{D}_N^{(1)}} \|f_V(z)\|_H\right\} + \varepsilon.$$

Proof. By definition of $\mathbb{D}_N^{(j)}$, any $z \in \partial\mathfrak{S}$ with$|\operatorname{Im}(z)| \leqslant M$ and $\operatorname{Re}(z) = j$, there exists some $z' \in \mathbb{D}_N^{(j)}$ with $|z - z'| \leqslant \frac{1}{N}$. This implies that every $k \in [d]$ satisfies $|g_k(z) - g_k(z')| \lesssim |z - z'| \cdot M^4\mathsf{d} = \frac{M^4\mathsf{d}}{N}$ by (43). Thus,

$$\|f_V(z) - f_V(z')\|_H^2 \leqslant \sum_{k=1}^{d} |g_k(z) - g_k(z')|^2 \leqslant \frac{\varepsilon^2}{\mathsf{d}^2},$$

for a large enough setting of N. Therefore, we have $\|f_V(z) - f_V(z')\|_H \leqslant \frac{\varepsilon}{\mathsf{d}}$, and that $\|f_V(z) - f_V(z')\|_W \leqslant \varepsilon$, which completes the proof. □

Claim 5.16. *We have that*

$$\|x\|_\theta \leqslant \mathtt{Rep}(x, \theta, \varepsilon; W) + 3\varepsilon.$$

Proof. Every vector (V, α) which is feasible in $\mathtt{Rep}(x, \theta, \varepsilon; W)$ defines a function $f_V \in \mathscr{F}$ by (36), which by Claim 5.15, satisfies

$$\|f_V\|_{\mathscr{F}} \leqslant \max\left\{ \sup_{z \in \mathbb{D}_N^{(0)}} \|f_V(z)\|_W, \sup_{z \in \mathbb{D}_N^{(1)}} \|f_V(z)\|_H \right\} + \varepsilon.$$

Therefore, for all (V, α) which are feasible for $\mathtt{Rep}(x, \theta, \varepsilon; W)$, we have:

$$\|x\|_\theta \leqslant \|f_V(\theta)\|_\theta + \|f_V(\theta) - x\|_\theta \leqslant \|f_V\|_{\mathscr{F}} + \varepsilon \leqslant \alpha + 3\varepsilon,$$

which implies $\|x\|_\theta \leqslant \mathtt{Rep}(x, \theta, \varepsilon; W) + 3\varepsilon$. □

Proof of Lemma 5.12. First, consider the solution (V, α) whose values of $(v_q : q \in \mathbb{Q}_M)$ define the function $\tilde{f}_x$ according to (36), and $\alpha = \|\tilde{f}_x\|_{\mathscr{F}}$. Then, (V, α) lies in the feasible set of $\mathtt{Rep}(x, \theta, \varepsilon; W)$. In particular, (39) shows that v_q satisfies the last constraint of (37). In addition, since $\|x\|_\theta \leqslant \mathsf{d}$, the third constraint of (37) is also satisfied. Therefore, we have that the optimal solution of $\mathtt{Rep}(x, \theta, \varepsilon; W)$ satisfies:

$$\mathtt{Rep}(x, \theta, \varepsilon; W) \leqslant \|\tilde{f}_x\|_{\mathscr{F}} \leqslant \left(1 + \frac{\varepsilon}{\mathsf{d}^2}\right) \|x\|_\theta.$$

If, in addition, (V, α) has $\alpha \leqslant \mathtt{Rep}(x, \theta, \varepsilon; W) + 5\varepsilon$, then by Claim 5.16 and $\|x\|_\theta \geqslant 1$,

$$(1 - 8\varepsilon)\|x\|_\theta \leqslant \alpha \leqslant \left(1 + \frac{\varepsilon}{\mathsf{d}^2} + \varepsilon\right) \|x\|_\theta,$$

which is a valid output of $\mathtt{ApproxRep}(x, \theta, 8\varepsilon; W)$. □

5.4 Computing $\mathtt{ApproxRep}(x, \theta, \varepsilon; W)$ with $\mathtt{MEM}(B_W)$

In this section, we show how to compute an approximate optimum of $\mathtt{Rep}(x, \theta, \varepsilon; X)$ described in (37). In particular, we will compute some value (V, α) which satisfies the conditions of Lemma 5.12. In other words, the solution (V, α) is feasible for $\mathtt{Rep}(x, \theta, \varepsilon; W)$, and the optimal value of $\mathtt{Rep}(x, \theta, \varepsilon; W)$ is at least $\alpha - 5\varepsilon$. If the algorithm runs in $\operatorname{poly}(d/\varepsilon)$ time, by Lemma 5.12, this would give us the required algorithm for $\mathtt{ApproxRep}(x, \theta, 8\varepsilon; X)$.

Let $P \subseteq (\mathbb{C}^d)^{|\mathbb{Q}_M|} \times \mathbb{R}^{\geqslant 0}$ be the set:

$$P = \left\{ ((v_q \max\{e^q, 1\} : q \in \mathbb{Q}_M), \alpha) \in (\mathbb{C}^d)^{|\mathbb{Q}_M|} \times \mathbb{R}^{\geqslant 0} : \begin{array}{ll} \text{i. } \forall z \in \mathbb{D}_N^{(0)} & \|f_V(z)\|_W \leqslant \alpha + \frac{\varepsilon}{2} \\ \text{ii. } \forall z \in \mathbb{D}_N^{(1)} & \|f_V(z)\|_H \leqslant \alpha + \frac{\varepsilon}{2} \\ \text{iii. } \forall q \in \mathbb{Q}_M & \|v_q\|_H \max\{e^q, 1\} \leqslant \frac{3M \cdot \mathsf{d}}{2} \\ \text{iv.} & \|f_V(\theta) - x\|_H \leqslant \frac{3\varepsilon}{2\mathsf{d}} \\ \text{v.} & \sum_{q \in \mathbb{Q}_M} \|v_q\|_H^2 \max\{e^{2q}, 1\} + \alpha^2 \leqslant R^2 \end{array} \right\}, \tag{44}$$

where $R = \text{poly}(d/\varepsilon)$ is a large enough parameter. In addition, we view the set $P \subseteq (\mathbb{C}^d)^{|\mathbb{Q}_M|} \times \mathbb{R}$ as a subset of $(\mathbb{R}^{2d})^{|\mathbb{Q}_M|} \times \mathbb{R}$ by separating the real and imaginary parts of the vectors v_q for $q \in \mathbb{Q}_M$.

5.4.1 Properties of the set P

The following are a couple of useful facts showing that the convex set P is nice. In particular, we will have that P is convex and inscribed within a Euclidean ball of polynomial radius. Any solution *close* to P (by an inverse polynomial amount), gives a feasible solution to $\texttt{Rep}(x, \theta, \varepsilon; W)$. The *approximate* representative to x, given by $\tilde{f}_x$ in (38) lies well within P (by an inverse polynomial amount). Finally, there exists an explicit solution well within P (by an inverse polynomial amount).

For the remainder of the section, we let $\delta = \text{poly}\left(\varepsilon, \frac{1}{d}\right)$ be the parameter:

$$\delta \stackrel{\text{def}}{=} \frac{\varepsilon}{2e \cdot |\mathbb{Q}_M| \cdot \mathsf{d}}.$$

Fact 5.17. *The set P is convex and is contained in a Euclidean ball of radius R.*

Lemma 5.18. *Let $(U, \alpha) \in B(P, \delta)$ where $U = (u_q : q \in \mathbb{Q}_M)$, and let $V = (v_q : q \in \mathbb{Q}_M)$ where $v_q = \frac{u_q}{\max\{e^q, 1\}}$. Then, (V, α) is a feasible solution to $\texttt{Rep}(x, \theta, \varepsilon; W)$.*

Proof. Consider $(U, \alpha) \in B(P, \delta)$ and let $(\widetilde{U}, \widetilde{\alpha}) \in P$ with $\|(U, \alpha) - (\widetilde{U}, \widetilde{\alpha})\|_2 \leqslant \delta$. We denote $\widetilde{U} = (\widetilde{u_q} : q \in \mathbb{Q}_M)$ and $\widetilde{V} = (\widetilde{v_q} : q \in \mathbb{Q}_M)$ where $\widetilde{v_q} = \frac{\widetilde{u_q}}{\max\{e^q, 1\}}$. Therefore, we have

$$\sum_{q \in \mathbb{Q}_M} \|u_q - \widetilde{u_q}\|_H^2 + |\alpha - \widetilde{\alpha}|^2 \leqslant \delta^2.$$

We will check that assuming constraints (i–iv) in (44), we can satisfy constraints (i–iv) in (37) for the point (V, α). All the subsequent checks proceed in the same fashion: we first use the triangle inequality to argue about $(\widetilde{U}, \widetilde{\alpha})$ and use (28) and the constraints (i–iv) in (44) of $(\widetilde{U}, \widetilde{\alpha})$ to deduce (V, α) is feasible for $\texttt{Rep}(x, \theta, \varepsilon; W)$.

iv. We simply follow the computations:

$$\begin{aligned} \|f_V(\theta) - x\|_H &\leqslant \|f_{\widetilde{V}}(\theta) - x\|_H + \|f_V(\theta) - f_{\widetilde{V}}(\theta)\|_H \leqslant \frac{3\varepsilon}{2\mathsf{d}} + e^{\theta^2/M} \sum_{q \in \mathbb{Q}_M} e^{q\theta} \|v_q - \widetilde{v_q}\|_H \\ &\leqslant \frac{3\varepsilon}{2\mathsf{d}} + e^{\theta^2/M} \sum_{q \in \mathbb{Q}_M} \frac{e^{q\theta}}{\max\{e^q, 1\}} \|u_q - \widetilde{u_q}\|_H \leqslant \frac{3\varepsilon}{2\mathsf{d}} + e \sum_{q \in \mathbb{Q}_M} \|u_q - \widetilde{u_q}\|_H \\ &\leqslant \frac{3\varepsilon}{2\mathsf{d}} + e|\mathbb{Q}_M|\delta \leqslant \frac{2\varepsilon}{\mathsf{d}}. \end{aligned}$$

iii. For $q \in \mathbb{Q}_M$, we have:

$$\|v_q\|_H \max\{e^q, 1\} \leqslant \|\widetilde{v_q}\|_H \max\{e^q, 1\} + \|u_q - \widetilde{u_q}\|_H \leqslant \frac{3M \cdot \mathsf{d}}{2} + \delta \leqslant 2M \cdot \mathsf{d}.$$

ii. For $z = 1 + i\tau \in \mathbb{D}_N^{(1)}$, we have:

$$\|f_V(z)\|_H \leqslant \|f_{\widetilde{V}}(z)\|_H + \|f_V(z) - f_{\widetilde{V}}(z)\|_H \leqslant \alpha + \frac{\varepsilon}{2} + e^{-\frac{\tau^2-1}{M}} \sum_{q \in \mathbb{Q}_M} e^q \|\widetilde{v_q} - v_q\|_H$$

$$\leqslant \alpha + \frac{\varepsilon}{2} + e|\mathbb{Q}_M|\delta \leqslant \alpha + \varepsilon.$$

i. In addition, for $z = i\tau \in \mathbb{D}_N^{(1)}$, we similarly have,

$$\|f_V(z)\|_W \leqslant \|f_{\widetilde{V}}(z)\|_W + \mathsf{d}\|f_V(z) - f_{\widetilde{V}}(z)\|_H \leqslant \alpha + \frac{\varepsilon}{2} + \mathsf{d} \cdot e^{-\frac{\tau^2}{M}} \sum_{q \in \mathbb{Q}_M} \|\widetilde{v_q} - v_q\|_H$$

$$\leqslant \alpha + \frac{\varepsilon}{2} + \mathsf{d}|\mathbb{Q}_M|\delta \leqslant \alpha + \varepsilon.$$

This completes the proof. □

Lemma 5.19. *We have $B((U, \alpha), \delta) \subseteq P$, where (U, α) is given by $U = (v_q \max\{e^q, 1\} : q \in \mathbb{Q}_M)$ where the vectors $v_q \in \mathbb{C}^d$ define $\tilde{f}_x$ in (38) according to (36), and $\alpha = \|\tilde{f}_x\|_{\mathcal{F}}$. In other words, $(U, \alpha) \in B(P, -\delta)$.*

Proof. Consider the point (V, α) where the vectors v_q define the function $\tilde{f}_x \in \mathcal{F}$ from (38) and $\alpha = \|\tilde{f}_x\|_{\mathcal{F}}$. Let $U = (u_q : q \in \mathbb{Q}_M)$ where $u_q = v_q \max\{e^q, 1\}$. We claim that $B((U, \alpha), \delta) \subseteq P$. In particular, consider any (U', α') with $U' = (u'_q : q \in \mathbb{Q}_M)$ where $\|(U', \alpha') - (U, \alpha)\|_2 \leqslant \delta$, i.e.,

$$\sum_{q \in \mathbb{Q}_M} \|u'_q - u_q\|_H^2 + |\alpha' - \alpha|^2 \leqslant \delta^2.$$

With arguments very similar to those in the proof of Lemma 5.18 above, we may check that constraints (i–iv) of (44) have $(U', \alpha') \in P$ since v_q satisfy (40) and (41)[9]. It remains to check that constraint (v) is satisfied. Letting $v'_q = \frac{u'_q}{\max\{e^q, 1\}}$,

$$\begin{aligned}
\sum_{q \in \mathbb{Q}_M} \|v'_q\|_H^2 \max\{e^{2q}, 1\} + \alpha'^2 &\leqslant \sum_{q \in \mathbb{Q}_M} (\|u_q\|_H + \|u'_q - u_q\|_H)^2 + (\alpha + |\alpha' - \alpha|)^2 \\
&\leqslant 2\left(\sum_{q \in \mathbb{Q}_M} \|u_q\|_H^2 + \alpha^2\right) + 2\left(\sum_{q \in \mathbb{Q}_M} \|u'_q - u_q\|_H^2 + |\alpha' - \alpha|^2\right) \\
&\leqslant 2\left(\sum_{q \in \mathbb{Q}_M} \|v_q\|_H^2 \max\{e^{2q}, 1\} + \alpha^2\right) + 2\delta^2 \\
&\leqslant 2\left(|\mathbb{Q}_M| M^2 \mathsf{d}^2 + \mathsf{d}^2\right) + 2\delta^2 \leqslant R.
\end{aligned}$$

Therefore, we may conclude that $(U', \alpha') \in P$, which implies that $B((U, \alpha), \delta) \subseteq P$. □

[9] Note that v_q here corresponds to $v_q \|x\|_\theta$ in (40) and (41), since in (40) and (41), v_q are the vectors which define $\tilde{f}_a$, and $\tilde{f}_x = \|x\|_\theta \tilde{f}_a$.

Corollary 5.20. *We have that* $\min_{(V,\alpha)\in B(P,-\delta)} \alpha \leqslant \|\tilde{f}_x\|_{\mathcal{F}}$.

In addition, using the very similar arguments as in the proof of Lemma 5.18, one may deduce the following lemma.

Lemma 5.21. *Let* (U, α) *be the vector given by:*

$$u_q = \begin{cases} x \cdot e^{-\theta^2/M} & q = 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \alpha = \mathsf{d}.$$

Then, $B((U, \alpha), \delta) \subseteq P$.

5.4.2 Optimizing over P

In this subsection, we show how to optimize over the convex set P defined in (44) using the tools from [LSV17].

Definition 5.22 (Optimization Oracle ($\mathtt{OPT}(K)$) [LSV17])**.** *For a convex set* $K \subseteq \mathbb{R}^m$*, given a unit vector* $c \in \mathbb{R}^m$ *and a real number* $\delta > 0$*, with probability* $1 - \delta$*, the oracle either finds a vector* $y \in \mathbb{R}^m$ *such that* $y \in B(K, \delta)$ *and* $c^\intercal x \leqslant c^\intercal y + \delta$ *for all* $B(K, -\delta)$*, or asserts* $B(K, -\delta)$ *is empty.*

The next lemma, combined with Lemma 5.12 shows that in order to solve $\mathtt{ApproxRep}(x, \theta, \varepsilon; W)$, it suffices to give an optimization oracle for P, $\mathtt{OPT}(P)$.

Lemma 5.23. *Let* (U, α) *be the response to a call to* $\mathtt{OPT}(P)$ *with vector* $c = (0, \ldots, 0, -1)$ *and* $\delta = \frac{\varepsilon}{2e|\mathbb{Q}_M|\mathsf{d}}$*, and let* $V = (v_q : q \in \mathbb{Q}_M)$ *be given by* $v_q = \frac{u_q}{\max\{e^q, 1\}}$*. Then,* (V, α) *is a feasible solution for* $\mathtt{Rep}(x, \theta, \varepsilon; W)$ *and the optimum of* $\mathtt{Rep}(x, \theta, \varepsilon; W)$ *is at least* $\alpha - 5\varepsilon$*.*

Proof. Note from Lemma 5.19 that $B(P, -\delta)$ is non-empty, so that $\mathtt{OPT}(P)$ always returns a vector $(U, \alpha) \in (\mathbb{R}^{2d})^{|\mathbb{Q}_M|} \times \mathbb{R}$ such that $(U, \alpha) \in B(P, \delta)$ and $\alpha' \geqslant \alpha - \delta$ for all $(U', \alpha') \in B(P, -\delta)$. We note that by Lemma 5.18, (V, q) is feasible for $\mathtt{Rep}(x, \theta, \varepsilon; W)$.

Additionally, let (U', α') where $U' = (v'_q \max\{e^q, 1\} : q \in \mathbb{Q}_M)$ where the vectors v'_q define $\tilde{f}_x$ from (38) according to (36), and $\alpha' = \|\tilde{f}_x\|_{\mathcal{F}}$. By Lemma 5.19, $(U', \alpha') \in B(P, -\delta)$, which implies $\alpha - \delta \leqslant \alpha' = \|\tilde{f}_x\|_{\mathcal{F}}$, and by (42), $\|\tilde{f}_x\|_{\mathcal{F}} \leqslant (1 + \frac{\varepsilon}{\mathsf{d}^2})\|x\|_\theta$. Finally, using Claim 5.16, and the fact that $\mathtt{Rep}(x, \theta, \varepsilon; W) \leqslant 2\mathsf{d}$,

$$\alpha \leqslant \mathtt{Rep}(x, \theta, \varepsilon; W) + \frac{2\varepsilon}{\mathsf{d}} + 4\varepsilon + \delta \leqslant \mathtt{Rep}(x, \theta, \varepsilon; W) + 5\varepsilon,$$

which completes the proof. □

Given Lemma 5.23 and Lemma 5.12, in order to solve $\mathtt{ApproxRep}(x, \theta, 8\varepsilon; W)$ it suffices to give an algorithm which solves $\mathtt{OPT}(K)$ with a unit vector c and error δ in time $\mathrm{poly}(\frac{d}{\varepsilon}, \frac{1}{\delta})$. In order to do this, we will utilize the reduction from [LSV17] which reduces the optimization problem to the separation problem.

Definition 5.24 (Separation Oracle ($\texttt{SEP}(K)$) [LSV17]). *For a convex set $K \subseteq \mathbb{R}^m$, given a vector $y \in \mathbb{R}^m$ and a real number $\delta > 0$, with probability $1 - \delta$, the oracle either asserts $y \in B(K, \delta)$, or finds a unit vector $c \in \mathbb{R}^m$ such that $c^\intercal x \leqslant c^\intercal y + \delta$ for all $x \in B(K, -\delta)$.*

Theorem 9 (Theorem 15 from [LSV17]). *Let $K \subseteq \mathbb{R}^m$ be a convex set satisfying $B(0, r) \subseteq K \subseteq B(0, 1)$, and let $\kappa = \frac{1}{r}$. For any $\delta \in (0, 1)$, with probability $1 - \delta$, one can compute $\texttt{OPT}(K)$ with a unit vector c and parameter δ with $O(m \log(\frac{m\kappa}{\delta}))$ calls to $\texttt{SEP}(K)$ with error parameter $\delta' = \mathrm{poly}(m, \delta, \frac{1}{\kappa})$ and $\mathrm{poly}(m, \log(\frac{\kappa}{\delta}))$ additional time.*

Given Theorem 9, we give an algorithm which solves $\texttt{OPT}(P)$ using calls to $\texttt{SEP}(P)$.

Lemma 5.25. *There exists an algorithm for $\texttt{OPT}(P)$ with a unit vector c and error parameter δ making $\mathrm{poly}\left(\frac{d}{\varepsilon}, \log\left(\frac{1}{\delta}\right)\right)$ calls to $\texttt{SEP}(P)$ with error parameter $\delta' = \mathrm{poly}\left(\frac{d}{\varepsilon}, \delta\right)$ and $\mathrm{poly}\left(\frac{d}{\varepsilon}, \log\left(\frac{1}{\delta}\right)\right)$ additional time.*

Proof. Recall that by Lemma 5.21, there exists a vector $(U_0, \alpha_0) \in B(P, -\delta)$, which we may compute algorithmically. This implies that the set $P' = \{(U, \alpha) - (U_0, \alpha_0) : (U, \alpha) \in P\}$, has $B(0, \delta) \subseteq P' \subseteq B(0, 2R)$. This in turn, implies that the set $P_0 = \{\frac{1}{2R}(U, \alpha) : (U, \alpha) \in P'\}$ has $B(0, \frac{\delta}{2R}) \subseteq P_0 \subseteq B(0, 1)$. Suppose $(\widetilde{U}, \widetilde{\alpha})$ is the output of $\texttt{OPT}(P_0)$ with a unit vector c and error parameter $\frac{\delta}{2R}$. In addition, one may easily verify that $2R(\widetilde{U}, \widetilde{\alpha}) + (U_0, \alpha_0)$ is a valid output for $\texttt{OPT}(P)$ with unit vector c and error parameter δ.

Given Theorem 9, and the fact that $R = \mathrm{poly}(d/\varepsilon)$, it suffices to show that one may implement $\texttt{SEP}(P_0)$ with $\texttt{SEP}(P)$. Consider a call to $\texttt{SEP}(P_0)$ with some point (U, α) and error parameter $\delta' > 0$. Let $(\widetilde{U}, \widetilde{\alpha}) = 2R(U, \alpha) + (U_0, \alpha_0)$. Then if $(\widetilde{U}, \widetilde{\alpha}) \in B(P, 2R\delta')$, then $(U, \alpha) \in B(P_0, \delta')$. If c is a unit vector where $c^\intercal \cdot (U', \alpha') \leqslant c^\intercal \cdot (\widetilde{U}, \widetilde{\alpha}) + 2R\delta'$ for all $(U', \alpha') \in B(P, -2R\delta')$, then, the vector c is a valid output of $\texttt{SEP}(P_0)$. □

Thus, given Lemma 5.25, it suffices to design an algorithm for $\texttt{SEP}(P)$ which runs in $\mathrm{poly}\left(\frac{d}{\varepsilon}, \delta\right)$ time with error parameter δ.

Lemma 5.26. *There exists an oracle for $\texttt{SEP}(P)$ with error parameter δ which makes $\mathrm{poly}\left(\frac{d}{\varepsilon}\right)$ calls to an oracle $\texttt{SEP}(B_W)$ with error parameter $\delta' = \mathrm{poly}\left(\frac{d}{\varepsilon}, \frac{1}{\delta}\right)$ and takes $\mathrm{poly}\left(\frac{d}{\varepsilon}\right)$ additional time.*

Proof. Consider some $(U, \alpha) \in (\mathbb{R}^{2d})^{|\mathbb{Q}_M|} \times \mathbb{R}$ with $U = (u_q : q \in \mathbb{Q}_M)$ which is an input to $\texttt{SEP}(P)$ with error parameter δ. Let $V = (v_q : q \in \mathbb{Q}_M)$ have $v_q = \frac{u_q}{\max\{e^q, 1\}}$. The algorithm proceeds as follows:

1. First, check whether constraint (v) in (44) is violated by computing $\|(U, \alpha)\|_2^2$ in $\mathrm{poly}\left(\frac{d}{\varepsilon}\right)$ time since $H = \ell_2^{2d}$. If $\|(U, \alpha)\|_2^2 \leqslant R^2$, then continue. Otherwise, we output the vector $c = \frac{(U,\alpha)}{\|(U,\alpha)\|_2}$, which is a valid output of $\texttt{SEP}(P)$.

2. Second, we may think of U_M as the $2d \times |\mathbb{Q}_M|$ matrix, whose columns are the vectors $v_q \in \mathbb{R}^{2d}$. Thus, the constraints (ii–iv) may be written as $\|U_M \gamma\|_H \leqslant \lambda$ for some $\gamma \in \mathbb{R}^{|\mathbb{Q}_M|}$ which is a column vector. If none of the constraints (ii–iv) are violated, then we continue. If some

constraint is violated, i.e., $\|U_M\gamma\|_H > \lambda$, let $b = U_M\gamma$, so that $\frac{b^\intercal U_M\gamma}{\|b\|_H} > \lambda$, but any $(U', \alpha') \in P$ with $\|U'_M\gamma\|_H \leqslant \lambda$ has $\frac{b^\intercal U'_M\gamma}{\|b\|_H} \leqslant \lambda$. Thus, we consider the vector $c = (\frac{\gamma_q b}{\|b\|_H} \in \mathbb{R}^{2d} : q \in \mathbb{Q}_M) \times (0) \in (\mathbb{R}^{2d})^{|\mathbb{Q}_M|} \times \mathbb{R}$, so that $c^\intercal \cdot (U, \alpha) = \frac{b^\intercal U_M\gamma}{\|b\|_H}$. Thus, $\frac{c}{\|c\|_2}$ is a valid output for $\mathtt{SEP}(P)$.

3. Finally, we consider constraint (i), which may be written as $\|U_M\gamma\|_W \leqslant \lambda$ for some $\lambda \in \mathbb{R}$ with $\lambda \in (\frac{\varepsilon}{2}, 2R)$. For each constraint of type (i), we query the oracle $\mathtt{SEP}(B_W)$ on the vector $y = \frac{U_M\gamma(1+\delta'\mathsf{d})}{\lambda}$ with error parameter δ' (which we specify later). If $\mathtt{SEP}(B_W)$ asserts $y \in B(B_W, \delta')$, then $\|U_M\gamma\|_W \leqslant \lambda$. So if all oracle calls to $\mathtt{SEP}(B_W)$ assert $y \in B(B_W, \delta')$, then since constraints (ii–v) are satisfied as well, we have $(U, \alpha) \in P$, so we may assert $(U, \alpha) \in B(P, \delta)$.

   Otherwise, suppose $\|U_M\gamma\|_W > \lambda$. Then, we have

$$\frac{\lambda}{\mathsf{d}} \leqslant \|U_M\gamma\|_H \leqslant \sum_{q\in\mathbb{Q}} |\gamma_q| \|u_q\|_H \leqslant \|\gamma\|_1 \cdot M \cdot \mathsf{d}.$$

   In this case, $\mathtt{SEP}(B_W)$ outputs a unit vector $b \in \mathbb{R}^{2d}$ where $b^\intercal \widetilde{y} \leqslant b^\intercal y + \delta'$ for all $\widetilde{y} \in B(B_W, -\delta')$. So suppose $(\widetilde{U}, \widetilde{\alpha}) \in P$ and in particular, $\|\widetilde{U}_M\gamma\|_W \leqslant \lambda$. Then, letting $\widetilde{y} = \frac{\widetilde{U}_M\gamma(1-\delta'\mathsf{d})}{\lambda}$, we have $\widetilde{y} \in B(B_W, -\delta')$. Therefore,

$$b^\intercal \widetilde{U}_M\gamma \leqslant b^\intercal U_M\gamma + \delta'\lambda + \delta'\mathsf{d}\|b\|_2\|\widetilde{U}_M\gamma\|_H \leqslant b^\intercal U_M\gamma + 2\delta'\lambda\mathsf{d}.$$

   Similarly to step 2 above, we consider the vector $c = (\gamma_q b : q \in \mathbb{Q}_M) \times (0) \in (\mathbb{R}^{2d})^{|\mathbb{Q}_M|} \times \mathbb{R}$ which satisfies $b^\intercal U_M\gamma = c^\intercal(U, \alpha)$ for all (U, α). Recall that since b is a unit vector in $\mathbb{R}^{2d}$, we have $\|c\|_2 = \|\gamma\|_2 \geqslant \frac{\|y\|_1}{2d|\mathbb{Q}_M|} \geqslant \frac{\varepsilon}{4d\mathsf{d}^2 M|\mathbb{Q}_M|}$, which in turn, implies that:

$$\frac{c^\intercal}{\|c\|_2} \cdot (\widetilde{U}, \widetilde{\alpha}) \leqslant \frac{c^\intercal}{\|c\|_2} \cdot (U, \alpha) + \frac{8d\lambda\mathsf{d}^3 M|\mathbb{Q}_M|}{\varepsilon} \cdot \delta' \leqslant \frac{c^\intercal}{\|c\|_2} \cdot (U, \alpha) + \delta,$$

   when $\delta' = \dfrac{\delta\varepsilon}{8dR\mathsf{d}^3 M|\mathbb{Q}_M|}$.

□

Finally, we will use a reduction from [LSV17], which asserts that one may implement $\mathtt{SEP}(B_W)$ with $\mathtt{MEM}(B_W)$.

Theorem 10 (Theorem 14 from [LSV17])**.** *Let $K \subseteq \mathbb{R}^m$ be a convex body satisfying $B(0, r) \subseteq K \subseteq B(0, 1)$, and let $\kappa = \frac{1}{r}$. For any $\delta \in (0, 1)$, with probability $1 - \delta$, one can compute $\mathtt{SEP}(K)$ with error parameter δ with $O\left(m \log\left(\frac{m\kappa}{\delta}\right)\right)$ calls to $\mathtt{MEM}(K)$ with error parameter $\delta' = \mathrm{poly}(\delta, \frac{1}{\kappa}, \frac{1}{m})$ and $\mathrm{poly}(m, \log(\frac{\kappa}{\delta}))$ time.*

Lemma 5.27. *There exists an algorithm for $\mathtt{SEP}(B_W)$ with parameter δ which makes $\mathrm{poly}\left(\frac{d}{\varepsilon}, \log\left(\frac{1}{\delta}\right)\right)$ calls to $\mathtt{MEM}(B_W)$ with parameter $\delta' = \mathrm{poly}(\delta, \frac{\varepsilon}{d})$.*

Proof. We simply use Theorem 14 from [LSV17], where we note that $B_W \subseteq B_2 \subseteq \mathsf{d}B_W$, which implies that $B(0, \frac{1}{\mathsf{d}}) \subseteq B_W \subseteq B(0, 1)$. □

5.5 Summary and instantiation for applications

From the discussion above, we may conclude the following theorem, whose proof simply follows by combining the reductions given in Lemma 5.12, Lemma 5.23, Lemma 5.25, Lemma 5.26, and Lemma 5.27.

Theorem 11. *There exists an algorithm which solves* $\texttt{ApproxRep}(x, \theta, \varepsilon; W)$ *with probability at least* $1 - \mathrm{poly}(\varepsilon, \frac{1}{d})$ *making* $\mathrm{poly}(d, \frac{1}{\varepsilon})$ *calls to* $\texttt{MEM}(W)$ *and using* $\mathrm{poly}(d, \frac{1}{\varepsilon})$ *additional time.*

In order to see how the above algorithm will be applied, recall from our discussion in Subsection 2.1, that in designing algorithms for ANN over X, we will assume oracle access to the real normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$.

Lemma 5.28. *Let* $\varepsilon < \frac{1}{10}$ *and assume access to computing* $\|\cdot\|_X$*, where* $X = (\mathbb{R}^d, \|\cdot\|_X)$ *is a real normed space with*

$$B_X \subseteq B_2 \subseteq \mathsf{d} \cdot B_X,$$

for $\mathsf{d} = \mathrm{poly}(d)$*. Given a vector* $x \in \mathbb{C}^d$ *with* $\|x\|_{X^{\mathbb{C}}}$*, one can compute a multiplicative* $(1 \pm 4\varepsilon)$*-approximation to* $\|x\|_{X^{\mathbb{C}}}$ *in time* $\mathrm{poly}(d/\varepsilon)$*.*

Proof. Let $x = u + iv \in \mathbb{C}^d$. We note that we may compute $\|x\|_H$ for $H = (\ell_2^d)^{\mathbb{C}}$, so that the vector $y = \frac{x}{\|x\|_H}$ satisfies $1 \leqslant \|y\|_{X^{\mathbb{C}}} \leqslant \mathsf{d}$. Thus, we may assume without loss of generality that $1 \leqslant \|x\|_{X^{\mathbb{C}}} \leqslant \mathsf{d}$. For a parameter $P > 0$ (which we set briefly to $\mathrm{poly}(d/\varepsilon)$), consider the set:

$$\mathbb{D}_P^{(C)} = \left\{ \frac{k}{P} : 0 \leqslant k \leqslant 2\pi P \right\}.$$

Then, we note that by differentiation, we have that for all $\varphi \in [0, 2\pi]$, if we let $\widetilde{\varphi} \in \mathbb{D}_P^{(C)}$ be the smallest element greater than φ, when $P = \frac{2R}{\varepsilon}$,

$$\|u \cos \widetilde{\varphi} + v \sin \widetilde{\varphi}\|_X - \varepsilon \leqslant \|u \cos \varphi + v \sin \varphi\|_X \leqslant \|u \cos \widetilde{\varphi} + v \sin \widetilde{\varphi}\|_X + \varepsilon.$$

The value $\ell_{X^{\mathbb{C}}}(x) = \frac{1}{\pi \cdot |\mathbb{D}_P^{(C)}|} \sum_{\widetilde{\varphi} \in \mathbb{D}_P^{(C)}} \|u \cos \widetilde{\varphi} + v \sin \widetilde{\varphi}\|_X^2$ may be computed in $\mathrm{poly}(\frac{d}{\varepsilon}, R)$ time with access to $\|\cdot\|_X$, and we have:

$$(1 - 2\varepsilon)\ell_{X^{\mathbb{C}}}(x) \leqslant \frac{1}{\pi} \int_0^{2\pi} \|u \cos \varphi + v \sin \varphi\|_X^2 d\varphi \leqslant (1 + 2\varepsilon)\ell_{X^{\mathbb{C}}}(x) + \varepsilon^2.$$

Therefore, we have $(1 - 2\varepsilon)\ell_{X^{\mathbb{C}}}(x) \leqslant \|x\|_{X^{\mathbb{C}}}$ and $\|x\|_{X^{\mathbb{C}}} \leqslant (1 + 2\varepsilon)\ell_{X^{\mathbb{C}}}(x) + \varepsilon\|x\|_{X^{\mathbb{C}}}$, so $\|x\|_{X^{\mathbb{C}}} \leqslant (1 + 4\varepsilon)\ell_{X^{\mathbb{C}}}(x)$. □

Given Lemma 5.28, we may now design a membership oracle for $B_{X^{\mathbb{C}}}$.

Lemma 5.29. *Assume oracle access to a real normed space* $X = (\mathbb{R}^d, \|\cdot\|_X)$ *with*

$$B_X \subseteq B_2 \subseteq \mathsf{d} \cdot B_X.$$

Let $B_{X^{\mathbb{C}}} \subseteq \mathbb{R}^{2d}$ *be the convex set given by the unit ball* $B_{X^{\mathbb{C}}} \subseteq \mathbb{C}^d$*. There exists an membership oracle* $\texttt{MEM}(B_{X^{\mathbb{C}}})$ *running in time* $\mathrm{poly}(d, \frac{1}{\delta})$*.*

Proof. Given a vector $y \in \mathbb{R}^{2d}$, first compute $\|y\|_2$, and if $\|y\|_2 > 1$, assert that $y \notin B(B_{X^{\mathbb{C}}}, -\delta)$ since $y \notin B_2$ and $B_{X^{\mathbb{C}}} \subseteq B_2$. We interpret the vector $y \in \mathbb{C}^d$ and compute $\ell_{X^{\mathbb{C}}}(y) \in \mathbb{R}^{\geqslant 0}$ satisfying

$$\|y\|_{X^{\mathbb{C}}} \leqslant \ell_{X^{\mathbb{C}}}(y) \leqslant \left(1 + \frac{\delta}{2\|y\|_2}\right) \|y\|_{X^{\mathbb{C}}}. \tag{45}$$

Note that since $\|y\|_2 \leqslant 1$, the computing $\ell_{X^{\mathbb{C}}}(y)$ takes $\mathrm{poly}(d, \frac{1}{\delta})$ time by Lemma 5.28. If $\ell_{X^{\mathbb{C}}}(y) \leqslant 1$, then by (45), $\|y\|_{X^{\mathbb{C}}} \leqslant 1$, i.e, $y \in B_{X^{\mathbb{C}}}$. This means we may safely assert that $y \in B(X^{\mathbb{C}}, \delta)$. On the other hand, if $\ell_{X^{\mathbb{C}}}(y) > 1$, then by (45), $\|y\|_{X^{\mathbb{C}}} > \frac{1}{1+\frac{\delta}{2\|y\|_2}}$. Therefore, $\|y + \frac{\delta y}{\|y\|_2}\|_{X^{\mathbb{C}}} = (1+\frac{\delta}{\|y\|_2})\|y\|_{X^{\mathbb{C}}} > 1$, which implies that $B(y, \delta) \notin B_{X^{\mathbb{C}}}$, i.e., we may safely assert $y \notin B(B_{X^{\mathbb{C}}}, -\delta)$. □

Looking ahead to Section 7 and Section 8, as well as the discussion from Section 4, starting from oracle access to the normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$, we will consider the normed spaces

$$A = [X^{\mathbb{C}}, H]_\alpha \qquad \text{and} \qquad Y = [A, H]_\beta,$$

for some values of $\alpha, \beta \in (0, 1)$ with $\frac{1}{d} \leqslant \alpha, \beta \leqslant 1 - \frac{1}{d}$ (which implies $\Lambda(\alpha, \alpha)$ and $\Lambda(\beta, \beta)$ are at most $\mathrm{poly}(d)$). During the course of the algorithms in Section 7 and Section 8, we will need to compute $\texttt{ApproxRep}(x, \beta, \varepsilon; A)$, which is needed for computing norms $\|x\|_Y$ as well as approximate representatives in $\mathscr{F}$ defined by the couple $[A, H]_\theta$ in the definition of Section 4.

From Theorem 11, we consider the setting with $W = A$, so it suffices to construct an oracle $\texttt{MEM}(B_A)$ which runs in $\mathrm{poly}(d, \frac{1}{\delta})$ time. Note that we do not have oracle access to $\|\cdot\|_A$ (which would give $\texttt{MEM}(B_A)$). However, by Corollary 5.3, we may design $\texttt{MEM}(B_A)$ running in $\mathrm{poly}(d, \frac{1}{\delta})$ time by solving $\texttt{ApproxRep}(x, \alpha, \delta; X^{\mathbb{C}})$ again. Thus, we construct $\texttt{MEM}(B_A)$ by using Theorem 11, this time with $W = X^{\mathbb{C}}$, to solve $\texttt{ApproxRep}(x, \alpha, \delta; X^{\mathbb{C}})$ using $\mathrm{poly}(d, \frac{1}{\delta})$ time and oracle calls to $\texttt{MEM}(X^{\mathbb{C}})$. Finally, Lemma 5.29 shows how to solve $\texttt{MEM}(X^{\mathbb{C}})$ in $\mathrm{poly}(d, \frac{1}{\delta})$ time from oracle access to $\|\cdot\|_X$. We thus conclude the discussion below into the following corollary.

Corollary 5.30. *Assume oracle access to a real normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$ with*

$$B_X \subseteq B_2 \subseteq \mathsf{d} \cdot B_X,$$

where $\mathsf{d} = \mathrm{poly}(d)$. Then, letting $H = (\ell_2^d)^{\mathbb{C}}$,

$$A = [X^{\mathbb{C}}, H] \qquad \textit{and} \qquad Y = [A, H],$$

there exists a $\mathrm{poly}(d/\varepsilon)$ time algorithm which for each $x \in \mathbb{C}^d$, computes a $(1 \pm \varepsilon)$-approximation to $\|x\|_Y$ with probability at least $1 - \varepsilon$.

6 Nonlinear Rayleigh quotient inequalities

6.1 High-level overview

In this section, we prove various nonlinear Rayleigh quotient inequalities using the approximate Hölder homeomorphism from Section 4. We refer the reader to the summary of the discussion of Section 4 in Subsection 4.3, where we collect the necessary properties of the approximate homeomorphism in Corollary 4.10 and Lemma 4.12. We recall the definition of the nonlinear Rayleigh quotient.

Definition 6.1 (Nonlinear Rayleigh quotient)**.** *For any $G \in \Delta(m)$, any metric space (X, d_X), and any $\boldsymbol{x} = (x_1, \ldots, x_m) \in X^m$, we denote:*

$$\mathsf{R}(\boldsymbol{x}, G, d_X^q) = \frac{\sum_{i=1}^m \sum_{j=1}^m g_{ij} d_X(x_i, x_j)^q}{\sum_{i=1}^m \sum_{j=1}^m \rho(i)\rho(j) d_X(x_i, x_j)^q},$$

where $\rho(i) = \sum_{j=1}^m g_{ij}$ denotes the row sums.

At a high level, the goal is to use the approximate Hölder homeomorphism to relate a nonlinear Rayleigh quotient in one normed space, to a nonlinear Rayleigh quotient in another normed space. In particular, we seek to derive a relationship between the nonlinear Rayleigh quotient of a matrix $G \in \Delta(m)$ and sequence of points $\boldsymbol{x} \in (\mathbb{C}^d)^m$ lying in a Banach space $X = (\mathbb{C}^d, \|\cdot\|_X)$ and the nonlinear Rayleigh quotient of the same matrix G with a transformed set of points $\boldsymbol{y} \in (\mathbb{C}^d)^m$ in a Hilbert space $H = (\mathbb{C}^d, \|\cdot\|_H)$. The key properties that we need is that 1) the underlying matrix G remains the same, and 2) the transformation $\boldsymbol{x} \mapsto \boldsymbol{y}$ depends very little on G and on $\boldsymbol{x}$, i.e., the transformation can be encoded with succinctly.

Considering the *exact* homeomorphism $\Phi_0 \colon \mathbb{C}^d \to \mathbb{C}^d$ (from Corollary 4.10 with $\varepsilon = 0$), we note that the arguments from from Lemma 6.3 and Lemma 6.5 applied to Φ_0 give the following lemma.

Lemma 6.2. *Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a Banach space and $H = (\mathbb{C}^d, \|\cdot\|_H)$ be a Hilbert space satisfying:*

$$B_X \subseteq B_H \subseteq \mathsf{d} \cdot B_X.$$

For any $G \in \Delta(m)$ and any $\boldsymbol{x} = (x_1, \ldots, x_m) \in (\mathbb{C}^d)^m$, there exists a point $z \in \mathbb{C}^d$ such that letting $\boldsymbol{y} = (y_1, \ldots, y_m) \in (\mathbb{C}^d)^m$ be $y_i = \Phi_0(x_i - z)$ satisfies:

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_H^2) \lesssim \log^4 \mathsf{d} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2)^{\sqrt{\frac{\log\log \mathsf{d}}{\log \mathsf{d}}}}.$$

Note that the above lemma indeed satisfies the two key properties: both nonlinear Rayleigh quotients use the same matrix G, and the map $\boldsymbol{x} \mapsto \boldsymbol{y}$ is specified by z since Φ_0 is independent of G and $\boldsymbol{x}$. However, since we will use the nonlinear Rayleigh quotients algorithmically, the computational efficiency of Φ_0 and the representation of z becomes an issue. In particular, instead of the exact map Φ_0, we have to settle for an approximation $\Phi_\varepsilon \colon \mathbb{C}^d \to \mathbb{C}^d$ from Corollary 4.10 for some specified $\varepsilon > 0$, and instead of storing z (which may be an arbitrary complex vector), we need to round z to a bounded number of bits of precision.

Let us briefly describe how these issues manifest themselves in the more complicated statement of the nonlinear Rayleigh quotient inequality in Lemma 6.6. First, we will require that the points $\boldsymbol{x} \in (\mathbb{C}^d)^m$ and $z \in \mathbb{C}^d$ be such that the approximate map Φ_ε applied to points $x_i - z$ has the desired Hölder properties. We achieve this condition by requiring that $\|x_i - z\|_X$ is not too large, and that the points $\boldsymbol{x}$ are not too close to each other. Second, we will want our inequality to be robust to small changes in z, which will allow us to round z to polynomially many bits, as well as small changes in G. In order to do this, we will require that points in $\boldsymbol{x}$ are "spread out" with respect to the distribution induced on points by the row sums of G.

Even though our applications require setting $\varepsilon > 0$, we encourage the reader to first consider the case when $\varepsilon = 0$. This case highlights the conceptual ideas in this section, and the particular parameter settings for $\varepsilon > 0$ are very loose, since our algorithm will allow us to set $\varepsilon = \frac{1}{d^{O(1)}}$.

In Section 7, we show how these inequalities are used in order to construct efficient partitions of points in an arbitrary normed space. We first show how to derive nonlinear Rayleigh quotient relationships given an approximate Hölder homeomorphism with a mild condition on the points. This is done in Lemma 6.3 in Subsection 6.2, and the mild condition on the points requires the image of points after the map be in a roughly centered position. We then show in Lemma 6.5 that a rough centering may be achieved by translating the original points by some vector whose norm is not too large. Finally, Lemma 6.6 in Subsection 6.3 shows how to combine the results from Subsection 6.2 and Section 4 in order to relate the nonlinear Rayleigh quotient of any norm to that of a Hilbert space.

6.2 Relating nonlinear Rayleigh quotients with Hölder homeomorphisms

Consider two Banach spaces $Y = (\mathbb{C}^d, \|\cdot\|_Y)$ and $Z = (\mathbb{C}^d, \|\cdot\|_Z)$ and let $R > 1$ be a large parameter and $\varepsilon > 0$ be a small parameter, $p \geqslant 1$, $r \geqslant \frac{1}{p}$ and $C > 1$. Suppose there exists a map $\Phi\colon Y \to Z$ satisfying the following conditions of Corollary 4.10 and Lemma 4.12[10]:

- For every $x \in \mathbb{C}^d$,

$$\|x\|_Y^r \leqslant 2\|\Phi(x)\|_Z \leqslant 4\|x\|_Y^r, \tag{46}$$

- For every $x, y \in \mathbb{C}^d$ with $\|x\|_Y, \|y\|_Y \leqslant R$ and $\|x-y\|_Y \geqslant \varepsilon$,

$$\|\Phi(x) - \Phi(y)\|_Z \leqslant C\|x-y\|_Y^{1/p}\left(\|x\|_Y^{r-1/p} + \|y\|_Y^{r-1/p}\right). \tag{47}$$

- Every $x \in \mathbb{C}^d$ satisfies

$$\|\Phi(x) - \Phi_0(x)\|_Z \lesssim \varepsilon^r \|x\|_Y^{r-\frac{1}{p}}, \tag{48}$$

  where $\Phi_0\colon Y \to Z$ is a homeomorphism such that for every $x \in \mathbb{C}^d$, $\|\Phi_0(x)\|_Z = \|x\|_Y^r$ and for every $x, y \in \mathbb{C}^d$,

$$\|\Phi_0(x) - \Phi_0(y)\|_Z \leqslant C\|x-y\|_Y^{1/p}\left(\|x\|_Y^{r-1/p} + \|y\|_Y^{r-1/p}\right). \tag{49}$$

For the remainder of this section, we fix $\varepsilon > 0$ to be a small enough parameter and refer to $\Phi\colon Y \to Z$ as the map satisfying the properties from above.

Lemma 6.3. *Let $G \in \Delta(m)$ with row sums $\rho(i) = \sum_{j=1}^m g_{ij}$. Suppose $\boldsymbol{x} = (x_1, \ldots, x_m) \in (\mathbb{C}^d)^m$, $\boldsymbol{y} = (y_1, \ldots, y_m) \in (\mathbb{C}^d)^m$, and $\delta \in \mathbb{C}^d$ satisfy:*

- *$\|x_i\|_Y \leqslant R$ for all $i \in [m]$, and $\|x_i - x_j\|_Y \geqslant \varepsilon$ for $i \neq j \in [m]$,*
- *$y_i = \Phi(x_i)$ for all $i \in [m]$, and $\delta = \sum_{i=1}^m \rho(i) y_i$,*

[10]Note that these are only the first two-conditions in Corollary 4.10, where we note that we will have $\varepsilon \ll \frac{1}{r}$, so that $(1-\varepsilon)^r\|x\|_Y^r \leqslant \|\Phi_\varepsilon(x)\|_Z$ implies $\|x\|_Y^r \leqslant 2\|\Phi_\varepsilon(x)\|_Z$.

- $\|\delta\|_Z^2 \leqslant \frac{1}{8}\left(\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r}\right)$.

Then,

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2) \leqslant 256C^2 \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_Y^{2r})^{\frac{1}{rp}}.$$

Proof. We use the first property of Φ, along with the definition of δ to say:

$$\begin{aligned}\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r} \overset{(46)}{\leqslant} 2\sum_{i=1}^m \rho(i)\|y_i\|_Z^2 = 2\sum_{i=1}^m \rho(i)\left\|y_i - \sum_{j=1}^m \rho(j)y_j + \delta\right\|_Z^2 \\ \leqslant 4\sum_{i=1}^m\sum_{j=1}^m \rho(i)\rho(j)\|y_i - y_j\|_Z^2 + 4\|\delta\|_Z^2.\end{aligned}$$

This, along with the upper bound on $\|\delta\|_Z^2$ and the definition of $\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2)$, implies:

$$\begin{aligned}\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r} &\leqslant 8\sum_{i=1}^m\sum_{j=1}^m \rho(i)\rho(j)\|y_i - y_j\|_Z^2 \\ &\leqslant \frac{8}{\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2)}\sum_{i=1}^m\sum_{j=1}^m g_{ij}\|y_i - y_j\|_Z^2.\end{aligned}$$

We now use the second condition of the map Φ to conclude:

$$\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r} \overset{(47)}{\leqslant} \frac{8C^2}{\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2)}\sum_{i=1}^m\sum_{j=1}^m g_{ij}\|x_i - x_j\|_Y^{2/p}\left(\|x_i\|_Y^{r-1/p} + \|x_j\|_Y^{r-1/p}\right)^2. \tag{50}$$

Applying Hölder's inequality to the right-hand side of (50) with exponents rp and $\frac{rp}{rp-1}$,

$$\begin{aligned}&\sum_{i=1}^m\sum_{j=1}^m g_{ij}\|x_i - x_j\|_Y^{2/p}\left(\|x_i\|_Y^{r-1/p} + \|x_j\|_Y^{r-1/p}\right)^2 \\ &\qquad\leqslant \left(\sum_{i=1}^m\sum_{j=1}^m g_{ij}\|x_i - x_j\|_Y^{2r}\right)^{\frac{1}{rp}}\left(\sum_{i=1}^m\sum_{j=1}^m g_{ij}\left(\|x_i\|_Y^{r-1/p} + \|x_j\|_Y^{r-1/p}\right)^{\frac{2rp}{rp-1}}\right)^{\frac{rp-1}{rp}} \\ &\qquad\leqslant 4\left(\sum_{i=1}^m\sum_{j=1}^m g_{ij}\|x_i - x_j\|_Y^{2r}\right)^{\frac{1}{rp}}\left(\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r}\right)^{\frac{rp-1}{rp}},\end{aligned} \tag{51}$$

where in the last inequality, we used the fact that

$$\begin{aligned}\sum_{i=1}^m\sum_{j=1}^m g_{ij}\left(\|x_i\|_Y^{r-1/p} + \|x_j\|_Y^{r-1/p}\right)^{\frac{2rp}{rp-1}} &\leqslant \sum_{i=1}^m\sum_{j=1}^n g_{ij}\left(2\|x_i\|_Y^{(r-\frac{1}{p})\cdot\frac{2rp}{rp-1}} + 2\|x_j\|_Y^{(r-\frac{1}{p})\cdot\frac{2rp}{rp-1}}\right) \\ &= 2\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r} + 2\sum_{j=1}^m \rho(j)\|x_j\|_Y^{2r} = 4\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r},\end{aligned}$$

as well as the fact that $4^{\frac{rp-1}{rp}} \leqslant 4$ since $rp \geqslant 1$. Thus, combining (50) and (51), we have:

$$\left(\sum_{i=1}^{m} \rho(i)\|x_i\|_Y^{2r}\right)^{\frac{1}{rp}} \leqslant \frac{32C^2}{\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2)} \left(\sum_{i=1}^{m}\sum_{j=1}^{m} g_{ij}\|x_i - x_j\|_Y^{2r}\right)^{\frac{1}{rp}}. \tag{52}$$

Finally, note that:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} \rho(i)\rho(j)\|x_i - x_j\|_Y^{2r} \leqslant \sum_{i=1}^{m}\sum_{j=1}^{m} \rho(i)\rho(j)\left(\|x_i\|_Y + \|x_j\|_Y\right)^{2r} \leqslant 2\cdot 4^r \sum_{i=1}^{m} \rho(i)\|x_i\|_Y^{2r}. \tag{53}$$

Thus combining (52) and (53), we may write:

$$\left(\frac{1}{2\cdot 4^r \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_Y^2)}\right)^{\frac{1}{rp}} \leqslant \frac{32C^2}{\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2)},$$

and the fact that $2^{\frac{1}{rp}}4^{\frac{1}{p}} \leqslant 8$ gives the final result. □

Note that Lemma 6.3 assumed a mild condition on the points. In particular, we assumed that $\boldsymbol{y} \in (\mathbb{C}^d)^m$ satisfied that the average (according to ρ) given by $\delta = \sum_{i=1}^m \rho(i)y_i$ is somewhat centered, i.e., $\|\delta\|_Z^2 \leqslant \frac{1}{8}\left(\sum_{i=1}^m \rho(i)\|x_i\|_Y^{2r}\right)$. Lemma 6.5 asserts that we may translate all original points $\boldsymbol{x} \in (\mathbb{C}^d)^m$ by a point $z \in \mathbb{C}^d$ in order to control $\|\delta\|_Z$ without $\|z\|_Y$ being too large. In order to prove it, we will need the following claim from [ANN+17].

Claim 6.4 (Claim 8.8 from [ANN+17])**.** *Let r be a positive integer and $h\colon \mathbb{R}^r \to \mathbb{R}^r$ be a continuous map such that for some norm $\|\cdot\|$ on $\mathbb{R}^r$ one has:*

$$\|h(w) - w\| = o(\|w\|)$$

as $w \to \infty$. Then, h is surjective.

Lemma 6.5. *Let $G \in \Delta(m)$ with row sums $\rho(i) = \sum_{j=1}^n g_{ij}$, and let $R_0 \leqslant R$ be a positive parameter. Suppose that $\boldsymbol{x} = (x_1, \ldots, x_m) \in (\mathbb{C}^d)^m$ satisfies:*

- *$\|x_i\|_Y \leqslant R_0$ for all $i \in [m]$, and*
- *$\|x_i - x_j\|_Y \geqslant \varepsilon$ for $i \neq j \in [m]$.*

Then, there exists a point $z \in \mathbb{C}^d$ with $\|z\|_Y \leqslant (6C)^p R_0$ written with $\mathrm{poly}(d, \log(\frac{1}{\varepsilon}), \log(R_0), \log(C), p, r)$ *bits such that:*

$$\delta = \sum_{i=1}^{m} \rho(i)\Phi(x_i - z) \qquad \textit{satisfies} \qquad \|\delta\|_Z \leqslant 2\varepsilon^r \sum_{i=1}^{m} \rho(i)\|x_i - z\|_Y^{r-\frac{1}{p}} + \varepsilon^{rp}.$$

Proof. We will prove that there exists a point $z \in \mathbb{C}^d$ with $\|z\|_Y \leqslant (6C)^p R_0$ such that

$$\delta_0 = \sum_{i=1}^{m} \rho(i)\Phi_0(x_i - z) = 0.$$

For any $z' \in \mathbb{C}^d$ with $\|z - z'\|_Y \leqslant (\frac{\varepsilon^r}{2C})^p$, which is obtained if we round coordinates of z to $\mathrm{poly}(\log d, \log(\frac{1}{\varepsilon}), \log(R_0), \log(C), p, r)$ many bits, we have that $\delta_0' = \sum_{i=1}^m \rho(i)\Phi_0(x_i - z')$ satisfies:

$$\begin{aligned}
\|\delta_0'\|_Z &\leqslant \sum_{i=1}^m \rho(i)\|\Phi_0(x_i - z') - \Phi_0(x_i - z)\|_Z \\
&\leqslant C \cdot \|z - z'\|_Y^{1/p} \sum_{i=1}^m \rho(i) \left(\|x_i - z'\|_Y^{r-1/p} + \|x_i - z\|_Y^{r-1/p}\right) \\
&\leqslant \varepsilon^r \sum_{i=1}^m \rho(i)\|x_i - z\|_Y^{r-1/p} + \varepsilon^{rp}.
\end{aligned}$$

Since $\Phi(x_i - z')$ is close to $\Phi_0(x_i - z')$ for every x_i, we obtain:

$$\|\delta\|_Z - \|\delta_0'\|_Z \leqslant \|\delta - \delta_0'\|_Z \leqslant \sum_{i=1}^m \rho(i)\|\Phi(x_i - z) - \Phi_0(x_i - z)\| \overset{(48)}{\leqslant} \varepsilon^r \sum_{i=1}^m \rho(i)\|x_i - z\|_Y^{r-\frac{1}{p}},$$

which would give the desired upper bound on $\|\delta\|_Z$. Towards this goal, consider the continuous map $h\colon Z \to Z$ given by:

$$h(u) = \sum_{i=1}^m \rho(i)\Phi_0\left(x_i - \Phi_0^{-1}(u)\right),$$

where we have:

$$\begin{aligned}
\|h(u) - u\|_Z &\leqslant \sum_{i=1}^m \rho(i) \left\|\Phi_0\left(x_i - \Phi_0^{-1}(u)\right) - \Phi_0\left(\Phi_0^{-1}(u)\right)\right\|_Z \\
&\overset{(49)}{\leqslant} C\sum_{i=1}^m \rho(i)\|x_i\|_Y^{1/p}\left(\|x_i - \Phi_0^{-1}(u)\|_Y^{r-1/p} + \|\Phi_0^{-1}(u)\|_Y^{r-1/p}\right) \\
&\leqslant \left(2C\sum_{i=1}^m \rho(i)\|x_i\|_Y^r\right) + \left(3C\sum_{i=1}^m \rho(i)\|x\|_Y^{1/p}\right)\|u\|_Z^{1-\frac{1}{rp}},
\end{aligned} \tag{54}$$

where we used the triangle inequality and the fact $\|\Phi_0^{-1}(u)\|_Y^r = \|u\|_Z$. Therefore, from (54) we conclude that $h\colon \mathbb{C}^d \to \mathbb{C}^d$ is a continuous map with $\|h(u) - u\|_Z = o(\|u\|_Z)$ as $\|u\|_Z \to \infty$, thus, we may view $h\colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ as a continuous which by Claim 6.4 implies h is surjective. Thus, consider the value $u \in \mathbb{C}^d$ where $h(u) = 0$, and let $z = \Phi_0^{-1}(u)$.

Suppose that $\|u\|_Z \geqslant 4C\sum_{i=1}^n \rho(i)\|x_i\|_Y^r$, then since $h(u) = 0$, (54) implies:

$$\|u\|_Z \leqslant \left(6C\sum_{i=1}^n \rho(i)\|x_i\|_Y^{1/p}\right)\|u\|_Z^{1-\frac{1}{rp}},$$

which implies

$$\|z\|_Y = \|u\|_Z^{1/r} \leqslant \left(6C\sum_{i=1}^n \rho(i)\|x_i\|_Y^{1/p}\right)^p \leqslant (6C)^p R_0.$$

In the other case, we have:

$$\|z\|_Y = \|u\|_Z^{1/r} \leqslant \left(4C\sum_{i=1}^n \rho(i)\|x_i\|_Y^r\right)^{1/r} \leqslant (4C)^{1/r}R \leqslant (6C)^p R_0,$$

since $r \geqslant \frac{1}{p}$. $\square$

6.3 A nonlinear Rayleigh quotient inequality for general norms

The goal of this subsection is to combine the results from Section 4 with Lemma 6.3 and Lemma 6.5 to relate the nonlinear Rayleigh quotient on an arbitrary norm to a nonlinear Rayleigh quotient in a Hilbert space. The main lemma of this section is the following:

Lemma 6.6. *Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a Banach space and $H = (\mathbb{C}^d, \|\cdot\|_H)$ be a Hilbert space satisfying:*

$$B_X \subseteq B_H \subseteq \mathsf{d} \cdot B_X,$$

where $\mathsf{d} = \mathrm{poly}(d)$, and let $0 < \varepsilon_0 < \frac{1}{10\mathsf{d}}$ and $R_0 > 1$ be two parameters. There exists a map $\Phi \colon \mathbb{C}^d \to \mathbb{C}^d$, such that for any $G \in \Delta(m)$ with row sums $\rho(i) = \sum_{j=1}^m g_{ij}$, and any $\boldsymbol{x} = (x_1, \ldots, x_m) \in (\mathbb{C}^d)^m$ satisfying:

- *$\|x_i\|_X \leqslant R_0$ for all $i \in [m]$, and*
- *$\|x_i - x_j\|_X \geqslant \varepsilon_0$ for all $i \neq j \in [m]$, and $\sum_{i=1}^m \sum_{j=1}^m \rho(i)\rho(j)\|x_i - x_j\|_X \geqslant \varepsilon_0$.*

There exists a point $z \in \mathbb{C}^d$ with $\|z\|_X \leqslant R_0 \cdot \mathsf{d}^2$ written with $\mathrm{poly}(d, \log(R_0), \log(\frac{1}{\varepsilon_0}))$ bits such that letting $\boldsymbol{y} = (y_1, \ldots, y_m) \in (\mathbb{C}^d)^m$ be $y_i = \Phi(x_i - z)$ satisfies:

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_H^2) \lesssim \log^4 \mathsf{d} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2)^{\sqrt{\frac{\log\log \mathsf{d}}{\log \mathsf{d}}}}.$$

In addition, the map $\Phi \colon \mathbb{C}^d \to \mathbb{C}^d$ is computable in $\mathrm{poly}(d, \frac{1}{\varepsilon_0}, R_0)$ time.

Proof. Consider the complex Banach space $A = [X, H]_\alpha$, where $\alpha = \sqrt{\frac{\log\log \mathsf{d}}{\log \mathsf{d}}}$. Since H is a Hilbert space, we have $K_2(H) = 1$, and we have $K_\infty(X) = 1$. Thus, by Lemma 2.14 we have:

$$K_p(A) \leqslant K_\infty(X)^{1-\alpha} K_2(H)^\alpha \leqslant 1,$$

when $p = \frac{1}{\alpha}$. In addition, we have that for every $x \in \mathbb{C}^d$, $\|x\|_A \leqslant \|x\|_X \leqslant \mathsf{d}^\alpha \|x\|_A$, which implies:

$$\mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_A^2) \leqslant e^{2\alpha \log \mathsf{d}} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2). \tag{55}$$

Consider the complex Banach spaces $Y = [A, H]_\beta$ and $Z = [A, H]_{1-\beta}$ where $\beta = \frac{1}{\log \mathsf{d}}$ which for every $x \in \mathbb{C}^d$ satisfy:

$$\|x\|_Y \leqslant \|x\|_A \leqslant \mathsf{d}^\beta \|x\|_Y \qquad \text{and} \qquad \|x\|_H \leqslant \|x\|_Z \leqslant \mathsf{d}^\beta \|x\|_H,$$

which allows us to conclude:

$$\mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_Y^2) \lesssim \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_A^2) \qquad \text{and} \qquad \mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_H^2) \lesssim \mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2). \tag{56}$$

We apply Corollary 4.10 with parameters

$$R = R_0 \cdot \mathsf{d}^2 \qquad \varepsilon = \frac{\beta}{5} \cdot \left(\frac{\varepsilon_0}{R \cdot \mathsf{d}^{\alpha+\beta}}\right)^{100} \qquad r = 1, \tag{57}$$

to obtain the map $\Phi = \Phi_\varepsilon \colon \mathbb{C}^d \to \mathbb{C}^d$ which by Section 5 is computable in $\mathrm{poly}(d, R_0, \varepsilon_0)$ time. We also note that we have: $\|x_i\|_Y \leqslant \mathsf{d}^{\alpha+\beta} R_0 \leqslant \frac{R}{2}$ for all $i \in [m]$ and that $\|x_i - x_j\|_Y \geqslant \varepsilon$ for $i \neq j \in [m]$. Letting $\rho(i) = \sum_{j=1}^m g_{ij}$ denote the row sums of G, we may apply Lemma 6.5 to obtain a point $z \in \mathbb{C}^d$ with $\|z\|_Y \leqslant \frac{R}{2}$ written with $\mathrm{poly}(d, \log(R_0), \log(\frac{1}{\varepsilon_0}))$ bits where $\delta = \sum_{i=1}^m \rho(i) y_i$ satisfies

$$\|\delta\|_Z - \varepsilon^p \leqslant \varepsilon \sum_{i=1}^m \rho(i) \|x_i - z\|_Y^{1-\frac{1}{p}} \leqslant \varepsilon \left(\sum_{i=1}^m \rho(i) \|x_i - z\|_Y \right)^{1-\frac{1}{p}},$$

where we used Jensen's inequality. Note that

$$\varepsilon_0 \leqslant \sum_{i=1}^m \sum_{j=1}^m \rho(i)\rho(j) \|x_i - x_j\|_X \leqslant 2 \sum_{i=1}^m \rho(i) \|x_i - z\|_X \leqslant 2\mathsf{d}^{\alpha+\beta} \sum_{i=1}^m \rho(i) \|x_i - z\|_Y,$$

and that $\varepsilon^p \leqslant (\frac{\varepsilon_0}{2\mathsf{d}^{\alpha+\beta}})^2 \leqslant \sum_{i=1}^m \rho(i) \|x_i - z\|_Y^2$. Thus we may conclude using this lower bound and Jensen's inequality

$$\|\delta\|_Z^2 \leqslant 8\varepsilon^2 \left(\frac{2\mathsf{d}^{\alpha+\beta}}{\varepsilon_0} \right)^{2/p} \sum_{i=1}^m \rho(i) \|x_i - z\|_Y^2 + 2\varepsilon^{2p} \leqslant \frac{1}{100} \sum_{i=1}^m \rho(i) \|x_i - z\|_Y^2. \tag{58}$$

Let $\boldsymbol{x}' = (x_1', \ldots, x_m') \in (\mathbb{C}^d)^m$ be $x_i' = x_i - z$. These points satisfy $\|x_i'\|_Y \leqslant R$ by the triangle inequality, and $\|x_i' - x_j'\|_Y \geqslant \varepsilon$ for when $i \neq j \in [m]$. These two conditions, along with (58) allows us to use Lemma 6.3 to obtain

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2) \lesssim \frac{1}{\beta^2} \cdot \mathsf{R}(\boldsymbol{x}', G, \|\cdot\|_Y^2)^\alpha = \frac{1}{\beta^2} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_Y^2)^\alpha. \tag{59}$$

We now combine the nonlinear Rayleigh quotient inequalities obtained to deduce:

$$\begin{aligned} \mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_H^2) \overset{(56)}{\lesssim} \mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_Z^2) \overset{(59)}{\lesssim} \frac{1}{\beta^2} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_Y^2)^\alpha &\overset{(56)}{\lesssim} \frac{1}{\beta^2} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_A^2)^\alpha \\ &\overset{(55)}{\lesssim} \frac{1}{\beta^2} \left(e^{2\alpha^2 \log \mathsf{d}} \right) \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2)^\alpha. \end{aligned}$$

□

The last lemma addressed the final point that the inequality from Lemma 6.6 is robust to small changes in the matrix G. Specifically, we show below that if we remove a small fraction of the rows and columns of G (with respect to the measure given by ρ), then we may "reuse" the same point z to derive the same nonlinear Rayleigh quotient inequality with respect to the modified matrix.

Lemma 6.7. *Suppose $G = (g_{ij})_{i,j \in [m]} \in \Delta(m)$. Suppose $\boldsymbol{x}, \boldsymbol{y} \in (\mathbb{C}^d)^m$ and $z, \delta \in \mathbb{C}^d$ satisfy the properties of Lemma 6.6. Let $G' = (g'_{ij})_{i,j \in [m]} \in \Delta(m)$ be obtained by considering any set $S \subseteq [m]$ with $\sum_{i \in S} \rho(i) \leqslant \frac{1}{\mathrm{poly}(\frac{1}{\varepsilon_0}, R_0, d)}$, and letting*

$$g'_{ij} = \begin{cases} 0 & i \in S \text{ or } j \in S \\ \frac{g_{ij}}{Z} & \text{otherwise} \end{cases},$$

where $Z \in \mathbb{R}^{\geqslant 0}$ is an appropriate coordinate. Then, if $\sum_{i=1}^m \sum_{j=1}^m \rho'(i)\rho'(j) \|x_i - x_j\| \geqslant \varepsilon_0$, we also have:

$$\mathsf{R}(\boldsymbol{y}, G', \|\cdot\|_H^2) \lesssim \log^4 \mathsf{d} \cdot \mathsf{R}(\boldsymbol{x}, G', \|\cdot\|_X^2)^{\sqrt{\frac{\log\log \mathsf{d}}{\log \mathsf{d}}}}.$$

Proof. Let $\rho'(i) = \sum_{j=1}^m g'_{ij}$ be the row sums of G'. Letting

$$\delta' = \sum_{i=1}^m \rho'(i)\Phi(x_i - z),$$

it remains to upper bound $\|\delta'\|_Z^2 \leqslant \frac{1}{8}\left(\sum_{i=1}^m \rho'(i)\|x_i - z\|_X^2\right)$, since then, the same argument as Lemma 6.6 will give the desired nonlinear Rayleigh quotient inequality. Recall by (58) in Lemma 6.6 that

$$\|\delta\|_Z^2 \leqslant \frac{1}{100}\sum_{i=1}^m \rho(i)\|x_i - z\|_Y^2. \tag{60}$$

Writing $\rho(S) = \sum_{i\in S}\rho(i)$, and recalling the settings of ε and R in (57), we have that:

$$\left\|\delta' - \frac{\delta}{Z}\right\|_Z \leqslant \frac{1}{Z}\sum_{i\in S}\rho(i)\|\Phi(x_i - z)\|_Z \overset{(46)}{\leqslant} \frac{2}{Z}\sum_{i\in S}\rho(i)\|x_i - z\|_Y \leqslant \frac{2R}{Z}\cdot\rho(S) \leqslant \frac{\varepsilon^2}{R}, \tag{61}$$

where we used the fact that $\|x_i - z\|_Y \leqslant R$, and that $\rho(S) \leqslant \frac{1}{\mathrm{poly}(\frac{1}{\varepsilon_0}, R_0, d)}$. Therefore, we have:

$$\begin{aligned}\|\delta'\|_H^2 &\leqslant \left(\left\|\delta' - \frac{\delta}{Z}\right\|_Z + \frac{1}{Z}\|\delta\|_Z\right)^2 \overset{(61)}{\leqslant} \frac{2\varepsilon^4}{R^2} + \frac{2}{Z^2}\|\delta\|_Z^2 \overset{(60)}{\leqslant} \frac{2\varepsilon^4}{R^2} + \frac{1}{50\cdot Z^2}\sum_{i=1}^m \rho(i)\|x_i - z\|_Y^2 \\ &\leqslant \frac{2\varepsilon^4}{R^2} + \frac{1}{25}\sum_{i\notin S}\rho'(i)\|x_i - z\|_Y^2 + \varepsilon^2 \leqslant \frac{1}{8}\sum_{i=1}^m \rho'(i)\|x_i - z\|_Y^2,\end{aligned}$$

where we used the fact that $2\mathsf{d}^{\alpha+\beta}\sum_{i=1}^m \rho'(i)\|x_i - z\|_Y \geqslant \sum_{i=1}^m\sum_{j=1}^m \rho'(i)\rho'(j)\|x_i - x_j\|_X \geqslant \varepsilon_0$ to conclude the last line. □

7 ANN via nonlinear Rayleigh quotient inequalities

We now describe a data structure for ANN over an arbitrary normed space $X = (\mathbb{C}^d, \|\cdot\|_X)$. The main ingredient is a distribution over subsets of $\mathbb{C}^d$ for partitioning X.

Before proceeding with the algorithm, it will be convenient to assume X has some nice properties (without loss of generality). First, we assume that the unit ball of X is close to $(\ell_2^2)^{\mathbb{C}}$, i.e.,

$$B_2 \subseteq B_X \subseteq \mathsf{d}\cdot B_2,$$

for $\mathsf{d} = \mathrm{poly}(d)$ where B_2 represents the unit ball of $(\ell_2^d)^{\mathbb{C}} = (\mathbb{C}^d, \|\cdot\|_{\ell_2^{\mathbb{C}}})$, which is isomorphic to $\ell_2^{2d} = (\mathbb{R}^{2d}, \|\cdot\|_2)$ considering the imaginary parts as distinct coordinates (see Section 2). We will consider the finite metric space given by rounding coordinates of points to $O(\log d)$ many bits within $R_0 B_X$ and measuring distance with respect to the norm $\|\cdot\|_X$. We use the following setting

$$\varepsilon_0 = \frac{1}{\mathsf{d}^{100}} \qquad \text{and} \qquad R_0 = O(\mathsf{d}^2).$$

We slightly abuse notation by denoting the set

$$X = \{x \in \mathbb{C}^d : \|x\|_X \leqslant R_0, \text{ and coordinates of } x \text{ are rounded to } c_0 \log d \text{ bits}\}$$

for some constant $c_0 > 0$, in order to consider the metric space (X, d_X), where $d_X(x, y) = \|x - y\|_X$. We may pick the constant c_0 so that X is ε_0-separated and $\frac{1}{\mathsf{d}}$-covering with respect to distances in $\|\cdot\|_X$. Following the reduction from Section 5 of [ANN+18], it suffices to design an ANN algorithm for (X, d_X). While it was not crucial in [ANN+18], we will rely on the property that for $x, y \in X$, $\|x - y\|_X \geqslant \varepsilon_0$ as it allows us to use the Rayleigh quotient inequality from Lemma 6.6.

In what follows, we assume that real numbers are written with $\mathrm{poly}(\log d)$ bits of precision.

7.1 Efficient partitions of normed spaces

In this subsection, we state the main partitioning lemma for normed spaces. We let $\mathscr{H}$ be the set of *coordinate cuts* in $\mathbb{R}^{2d}$.[11] In particular, each $H \in \mathscr{H}$ is specified by a index $i \in [2d]$, a real number $t \in \mathbb{R}$, and a direction {"+", "−"}. We interpret $H = (i, t, s)$ as a set specified by an axis-aligned hyperplane separator with respect to coordinate i with threshold at t and direction s, i.e., we write:

$$H = \begin{cases} \{x \in \mathbb{R}^{2d} : x_i \geqslant t\} & s = \text{“}+\text{”} \\ \{x \in \mathbb{R}^{2d} : x_i \leqslant t\} & s = \text{“}-\text{”} \end{cases}.$$

A *box* in $\mathbb{R}^{2d}$ is the intersection of coordinate cuts, and we note that a box may be encoded with at most $4d$ coordinate cuts. We let $\mathscr{B}$ be the set of all boxes in $\mathbb{R}^{2d}$. We note two simple facts about boxes in $\mathbb{R}^{2d}$. The first is that given the description of a box B in $\mathbb{R}^{2d}$ (by at most $4d$ coordinate cuts), for any point $x \in \mathbb{R}^{2d}$, one can determine whether $x \in B$ or not in $O(d)$ time. The second fact is that the intersection of two boxes in $\mathbb{R}^{2d}$ is also a box in $\mathbb{R}^{2d}$.

Given a map $f \colon X \to \mathbb{R}^{2d}$, we consider boxes in X after being transformed by the map. More specifically, for a map $f \colon X \to \mathbb{R}^{2d}$ and a box $B \in \mathscr{B}$, the *box transformed by f* is the set

$$B \circ f = \{x \in X : f(x) \in B\}.$$

The theorem below gives the partitioning result of a set of points in the normed space X. It shows that, for every n-point subset of X either there exists a dense ball in X, or there exists a distribution over *efficient* subsets of X which partition the n points into two nearly-balanced parts without separating "close" points too often. We will require two properties from "efficient subsets" of X. First, the encoding of a set should require at most $\mathrm{poly}(d)$ space, and second, there should be a $\mathrm{poly}(d)$ time algorithm which determines whether a point lies in the set.

The distributions will be supported on the following family of sets:

$$\mathcal{S} = \left\{ \bigcap_{i=1}^{k} B_i \circ \Phi_{z_i} : z_1, \dots, z_k \in \mathbb{C}^d, B_1, \dots, B_i \in \mathscr{B} \text{ and } 1 \leqslant k \leqslant \mathrm{poly}(d) \right\},$$

where the map $\Phi_z \colon \mathbb{C}^d \to \mathbb{R}^{2d}$ is the result of "decomplexification" of the map $x \mapsto \Phi(x - z)$, where Φ is specified by Lemma 6.6. More specifically, for a subset $S \in \mathcal{S}$ with

$$S = \bigcap_{i=1}^{k} B_i \circ \Phi_{z_i},$$

[11] We will work in $\mathbb{R}^{2d}$ since ℓ_2^{2d} is isomorphic to $(\ell_2^d)^{\mathbb{C}}$ (see Complexification in Section 2). We will use the nonlinear Rayleigh quotient inequalities relating $X = (\mathbb{C}^d, \|\cdot\|_X)$ to $(\ell_2^d)^{\mathbb{C}} = (\mathbb{C}^d, \|\cdot\|_{\ell_2^{\mathbb{C}}})$ and immediately interpret points in $(\ell_2^d)^{\mathbb{C}}$ as being in ℓ_2^{2d} by splitting all d coordinates of $\mathbb{C}^d$ from $\mathbb{C}$ to $\mathbb{R} \times \mathbb{R}$ encoding the real and imaginary parts.

we may determine if $x \in S$ by the following procedure: for all $i \in [k]$, 1) compute $y_i = \Phi(x - z_i) \in \mathbb{C}^d$, 2) interpret $y_i \in \mathbb{R}^{2d}$ by expanding each coordinate of y_i into two coordinates encoding the real and imaginary parts, and 3) $x \in S$ if all $i \in [k]$ satisfy $y_i \in \mathcal{B}_i$, otherwise, $x \notin S$.

Theorem 12. *Fix $\varepsilon \in (0, \frac{1}{2})$, $n \in \mathbb{N}$ and let P be any n-point dataset in X, either there exists a ball of radius $(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon})\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}$ containing $\frac{n}{50}$ points from P for some universal constant $c_1 > 0$, or there exists a distribution $\mathcal{D}$ supported on $\mathcal{S}$ such that:*

- *For every two $x, y \in X$ with $\|x - y\|_X \leqslant 1$,*

$$\Pr_{\mathbf{S} \sim \mathcal{D}}[\mathbf{S} \text{ separates } (x, y)] \leqslant \varepsilon,$$

- *For every set $\mathbf{S} \in \mathcal{D}$, we have*

$$\frac{1}{100} \leqslant \frac{|P \cap \mathbf{S}|}{n} \leqslant \frac{99}{100}.$$

We prove the above theorem by proving one partitioning lemma, and then building the distribution according to [ANN+18]. Following the notation from [ANN+18], we consider the space of all non-negative symmetric $m \times m$ matrices whose entries sum to 1 and denote this space $\Delta(m)$. For $G = (g_{ij})_{ij} \in \Delta(m)$, we denote $\rho_G \colon [m] \to \mathbb{R}$ and the row sums, i.e., $\rho_G(i) = \sum_{j=1}^m g_{ij}$, and for a subset $S \subseteq [m]$, we let $\rho_G(S) = \sum_{i \in S} \rho_G(i)$.

We associate $G \in \Delta(m)$ with a sequence of points $\boldsymbol{x} = (x_1, \ldots, x_m) \in X^m$, where x_i corresponds to row/column i. Therefore, we may view ρ_G as giving a probability distribution supported on the points $\boldsymbol{x}$. We will frequently interpret $\boldsymbol{x}$ as a set of m points in X. For instance, given a subset $S \subseteq X$ and a real number $\gamma \in [0, 1]$, we will say that S is γ-dense with respect to G and $\boldsymbol{x}$ if $\rho_G(S \cap \boldsymbol{x}) = \gamma$, where $S \cap \boldsymbol{x}$ denotes the set of points in $\boldsymbol{x}$ lying in S.

The main step of the proof of Theorem 12 is the following lemma (Lemma 7.1). Having established this upcoming lemma, the proof of Theorem 12 follows in exactly the same way as Section 3 of [ANN+18].

Lemma 7.1. *Let $\boldsymbol{x} \in X^N$ be the sequence of all points of X, where $|X| = N$, and let $G \in \Delta(N)$, where $g_{ij} > 0$ implies $\|x_i - x_j\|_X \leqslant 1$. Then, one of the following must hold:*

- *there exists a ball of radius $(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon})\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}$ which is at least $\frac{1}{4}$-dense with respect to G and $\boldsymbol{x}$;*
- *there exists a subset $S \in \mathcal{S}$ with:*

$$\frac{1}{3} \leqslant \rho_G(S \cap \boldsymbol{x}) \leqslant \frac{3}{4} \qquad \text{and} \qquad \sum_{i \in S, j \notin S} g_{ij} \leqslant 2\varepsilon.$$

Lemma 7.1 above is similar to Lemmas 3.7 and 8.3 of [ANN+18]. We use the Rayleigh quotient inequality from Lemma 6.6 and Cheeger's inequality to partition the points $\boldsymbol{x}$. We note that the cuts obtained by Cheeger's inequality form a collection of poly(d) boxes after applying the transformation Φ. Below is a formal proof.

Proof. Consider any $G \in \Delta(N)$ such that $g_{ij} > 0$ implies $\|x_i - x_j\|_X \leqslant 1$. If there is no ball of radius $(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon})^{\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}}$ which is $\frac{1}{4}$-dense with respect to G and $\boldsymbol{x}$, then,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_G(i)\rho_G(j)\|x_i - x_j\|_X^2 \gtrsim \left(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon}\right)^{2\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}}.$$

By Lemma 6.6, there exists a point $z \in \mathbb{C}^d$ with $\|z\|_X \leqslant R_0 \cdot \mathsf{d}$ such that if we define $\boldsymbol{y} = (y_1, \dots, y_N) \in (\mathbb{C}^d)^N$ by setting $y_i = \Phi(x_i - z)$, the following holds:

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_{\ell_2^{\mathbb{C}}}^2) \lesssim \log^4 \mathsf{d} \cdot \mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2)^{\sqrt{\frac{\log\log \mathsf{d}}{\log \mathsf{d}}}}. \tag{62}$$

Now we use the definition of $\mathsf{R}(\boldsymbol{x}, G, \|\cdot\|_X^2)$, using the fact that $\|x_i - x_j\|_X \leqslant 1$ whenever $g_{ij} > 0$, and (62), we obtain that

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_{\ell_2^{\mathbb{C}}}^2) \leqslant \varepsilon^2/2,$$

as long as we set c_1 to be a large enough constant. Decomplexifying $\ell_2^{\mathbb{C}}$ to consider $\ell_2^{2d} = (\mathbb{R}^{2d}, \|\cdot\|_2)$ by interpreting the points $\boldsymbol{y} \in (\mathbb{R}^{2d})^N$, we have:

$$\mathsf{R}(\boldsymbol{y}, G, \|\cdot\|_2) \leqslant \frac{\varepsilon^2}{2}.$$

This implies that for some coordinate $k \in [2d]$, we have $\mathsf{R}((\boldsymbol{y})_k, G, |\cdot|^2) \leqslant \varepsilon^2/2$, where $(\boldsymbol{y})_k = ((y_1)_k, (y_2)_k, \dots, (y_N)_k) \in \mathbb{R}^N$ is the projection of the points in $\boldsymbol{y}$ onto the k-th coordinate. Therefore, by Cheeger's inequality, there exists a threshold $t \in \mathbb{R}$ and a sign $s \in \{\text{“}+\text{”}, \text{“}-\text{”}\}$ such that the set $H \in \mathscr{H}$ specified by direction k with threshold t and sign s has $\rho_G(H \cap \boldsymbol{y}) \geqslant \frac{1}{2}$ and $\sum_{i: y_i \in H} \sum_{j: y_j \notin H} g_{ij} \leqslant \varepsilon$.

Note that we obtain a desired partition, modulo the balance condition. Hence, we repeat the above for a few stages iteratively on the larger side of the cut. In particular, we maintain a box $B \in \mathscr{B}$ which is the intersection of the coordinate cuts found by Cheeger's inequality. While $\rho_G(B \cap \boldsymbol{y}) \geqslant 1 - \frac{1}{\mathrm{poly}(d)}$, by Lemma 6.7, we may utilize the same point z to obtain another coordinate cut, which decreases the size of the box. As long as $\rho_G(B \cap \boldsymbol{y}) \leqslant 1 - \frac{1}{\mathrm{poly}(d)}$, we no longer modify the box, and re-compute a new center z to start another box B'. Once the intersection of all boxes is less than $\frac{3}{4}$-dense with respect to G and $\boldsymbol{y}$, we stop. Repeating this procedure, we obtain at most $\mathrm{poly}(d)$ boxes, while keeping the sparsity condition that $\sum_{i: y_i \in S} \sum_{j: y_j \notin S} g_{ij} \leqslant \varepsilon$. □

7.2 From Theorem 12 to ANN data structures

From Theorem 12, we conclude that the cutting modulus of any norm is bounded by:

$$\Xi(X, \varepsilon) \leqslant \left(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon}\right)^{\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}},$$

for some constant c_1. Furthermore, the cuts obtained by Theorem 12 can be encoded in $\mathrm{poly}(d)$ many bits (by specifying the $\mathrm{poly}(d)$ center points and the $\mathrm{poly}(d)$ boxes in $\mathbb{R}^d$). Finally, there exists an algorithm which, given the encoding of a set $S \in \mathcal{S}$, as well as a point $q \in X$, can decide in poly(d)-time whether $q \in S$ (by checking whether q lies in all poly(d) transformed boxes).

We now use the algorithm from Section 4 from [ANN+18], to obtain a data structure for ANN over any norm.

Theorem 13. *For any normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$ and any $\varepsilon > 0$, there exists a data structure for ANN over X achieving:*

- *approximation* $c \lesssim \left(\frac{c_1 \log^2 \mathsf{d}}{\varepsilon}\right)^{\sqrt{\frac{\log \mathsf{d}}{\log\log \mathsf{d}}}}$ *for some constant $c_1 > 0$, using*
- *space* $\mathrm{poly}(d) \cdot n^{1+\varepsilon}$*, and*
- *query time* $\mathrm{poly}(d) \cdot n^{\varepsilon}$*.*

Recalling the fact that $\mathsf{d} = \mathrm{poly}(d)$ gives us Theorem 4.

8 ANN via the embedding approach

8.1 High-level overview

In this section, we give an application of the approximate Hölder homeomorphism from Section 4. The goal will be utilize the tools from [BG18] to design an ANN data structure over any norm with subpolynomial approximation using polynomial space and sublinear query time. The main result is summarized in the next theorem.

Theorem 14. *For any normed space $X = (\mathbb{R}^d, \|\cdot\|_X)$, there exists a data structure for c-ANN over X achieving*

- *approximation* $c = 2^{O((\log d)^{2/3}(\log\log d)^{1/3})}$*, with*
- *space* $\mathrm{poly}(n, d)$*, and*
- *query time* $\mathrm{poly}(d)\log n$*.*

Even though the approximation guarantee from Theorem 14 is weaker than the approximation guarantee of the algorithm from Section 7, there are two main advantages. The first is that this data structure achieves query time which is logarithmic in n, as opposed to n^{ε}.[12] The second advantage is that the algorithm of Theorem 14 is conceptually very simple.

Before presenting a high level overview, as well as the proof of Theorem 14, we record the following lemma, which is a simple consequence of the definition of c-ANN.

[12] For a comparable approximation to that of Theorem 13, we may consider $\varepsilon = \frac{1}{\log^{1/4} n}$. Then, Theorem 13 has approximation $2^{O(\log^{3/4} d)}$ with query time $\mathrm{poly}(d) \cdot 2^{\log^{3/4} n} \gg \mathrm{poly}(d)\log n$

Lemma 8.1. *Suppose $W_0 = (\mathbb{C}^d, \|\cdot\|_{W_0})$ and $W_1 = (\mathbb{C}^d, \|\cdot\|_{W_1})$ are two Banach spaces, and $\gamma > 1$ is some parameter such that every $x \in \mathbb{C}^d$ satisfies*

$$\|x\|_{W_0} \leqslant \|y\|_{W_1} \leqslant \gamma \|x\|_{W_0}. \tag{63}$$

If there exists a data structure for c-ANN over W_0 using space $S(n)$ and query time $Q(n)$, then there exists a data structure for γc-ANN over W_1 using space $S(n)$ and query time $Q(n)$.

Proof. Let D be a data structure for c-ANN over W_0, the data structure D' will simulate D while computing distances in W_1. In particular, suppose D' is given a dataset $P \subseteq \mathbb{C}^d$ of n points to preprocess, it will simply interpret these points as belonging to the Banach space W_0. Upon receiving a query $x \in \mathbb{C}^d$ where $\|x - p\|_{W_1} \leqslant 1$ for some $p \in P$, by (63), $\|x - p\|_{W_0} \leqslant 1$, so the data structure D returns some point $p' \in P$ with $\|x - p'\|_{W_0} \leqslant c$. Again, by (63), $\|x - p'\|_{W_1} \leqslant \gamma c$, which completes the proof. $\square$

One of the conceptual contributions in [BG18] is a generic reduction from c-ANN to a c-bounded near neighbor (c-BNN), which we formally define next.

Definition 8.2 (c-bounded near neighbor [BG18])**.** *Consider a fixed Banach space $X = (\mathbb{C}^d, \|\cdot\|_X)$. The c-bounded near neighbor problem (c-BNN) over X asks to design a data structure which preprocesses a dataset $P \subseteq \mathbb{C}^d$ of n points where every point $p \in P$ satisfies $\|p\|_X \leqslant c$. Given a query point $q \in \mathbb{C}^d$ where some $p \in P$ satisfies $\|q - p\|_X \leqslant 1$, the data structure should return any point $p' \in P$ with $\|q - p'\|_X \leqslant \frac{c}{9}$.*

We now state Lemma 5.4 from [BG18] catered to general d-dimensional Banach spaces. While the norms considered in [BG18] are easy to compute, we state Lemma 5.4 assuming access to an approximate oracle for a normed space of interest.

Lemma 8.3 (Lemma 5.4 from [BG18])**.** *Let $X = (\mathbb{C}^d, \|\cdot\|_X)$ be a Banach space, and assume an algorithm which computes a function $\ell_X \colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$ with*

$$\|x\|_X \leqslant \ell_X(x) \leqslant 2\|x\|_X,$$

in time $T(n)$. Suppose D is a data structure for c-BNN over X using space $S(n)$ and query time $Q(n)$. Then, there exists a data structure $8c$-ANN over X using

- *space* $\mathrm{poly}(d) \cdot n \cdot S(n)$*, and*
- *query time* $\log d \cdot Q(n) \cdot T(n)$*.*

Applying Lemma 8.3, [BG18] gave algorithms for ANN under ℓ_p distances for $p > 2$ by solving the c-BNN problem and using the navigating net structure of [KL04]. In particular, they argued that the Mazur map $\ell_p \to \ell_2$ is a good enough embedding when points are within a ball of radius c. They solve the c-BNN problem in ℓ_p by embedding ℓ_p into ℓ_2, and use black-box ANN algorithms for ℓ_2. The proof of Theorem 14 proceeds in a similar high level fashion:

1. We aim to solve the c-BNN problem for a Banach space X. So first, we interpolate between X and a Hilbert space H, to obtain the Banach spaces Y and Z, as well as the approximate Hölder homeomorphism $\Phi_\varepsilon \colon Y \to Z$ (for some small $\varepsilon > 0$) from Corollary 4.10 from Section 4.

2. Since X will be relatively close to Y, using Lemma 8.1, we will solve the c-BNN over Y (up to an approximation loss), and similarly to [BG18], we view Φ_ε as an embedding into Z for vectors within radius c.

3. Finally, Z will be relatively close to H, so we may use any black-box ANN algorithms for ℓ_2 (up to an approximation loss) to solve the problem in Z.

We execute the above plan next which completes the proof of Theorem 14.

8.2 Proof of Theorem 14

From Theorem 14, we will aim to design an ANN data structure for the Banach space $X = (\mathbb{C}^d, \|\cdot\|_X)$. Assume that X satisfies

$$B_X \subseteq B_2 \subseteq \mathsf{d} \cdot B_X,$$

for $\mathsf{d} = \mathrm{poly}(d)$. Consider the parameters $\alpha \in (0,1)$ and $\beta \in (0,1)$, with

$$\alpha = \left(\frac{\log\log \mathsf{d}}{\log \mathsf{d}}\right)^{1/3} \qquad \text{and} \qquad \beta = \frac{1}{\log \mathsf{d}}, \tag{64}$$

and let A, Y and H be the complex normed spaces

$$A = [X, H]_\alpha, \qquad Y = [A, H]_\beta, \qquad \text{and} \qquad Z = [A, H]_{1-\beta}.$$

We note that for every $x \in \mathbb{C}^d$, we have the following inequalities:

$$\|x\|_Y \leqslant \|x\|_X \leqslant \mathsf{d}^{\alpha+\beta}\|x\|_Y \qquad \text{and} \qquad \|x\|_H \leqslant \|x\|_Z \leqslant \mathsf{d}^{\beta}\|x\|_H. \tag{65}$$

In addition, since $K_\infty(X) = 1$ and $K_2(H) = 1$, by Lemma 2.14, we have that $K_p(A) \leqslant 1$ when $p = \frac{1}{\alpha}$.

Corollary 8.4. *There exists a data structure for c-ANN over Z with approximation $c \lesssim \mathsf{d}^\beta$, using space* $\mathrm{poly}(n, d)$ *and query time $O(d \log n)$.*

Proof. We first note that by Lemma 8.1 and (65), it suffices to give a data structure for c-ANN over H. In addition, the complex Banach space $H = (\mathbb{C}^d, \|\cdot\|_H)$ is isomorphic to a Hilbert space over $\mathbb{R}^{2d}$, so it suffices consider data structures for 2-ANN over a real Hilbert space. For this task, we may use a data structure for 2-ANN over ℓ_2^{2d} with $\mathrm{poly}(n, d)$ space, and $O(d \log n)$ time [HIM12, KOR00]. $\square$

The following corollary is immediate from Lemma 8.1 and (65).

Corollary 8.5. *If there exists a data structure for c-ANN over Y with using space $S(n)$ and query time $Q(n)$, there exists a data structure for $c \cdot \mathsf{d}^{\alpha+\beta}$-ANN over X using space $S(n)$ and query time $Q(n)$.*

Given Corollary 8.5, we state Lemma 8.6 which builds an approximate norm oracle and Lemma 8.7 which solves c-BNN over Y. Combining Lemma 8.6 and Lemma 8.7 with Lemma 8.3 gives Theorem 14. The following lemma simply follows from Corollary 5.30.

Lemma 8.6. *There exists an algorithm that computes a function $\ell_Y \colon \mathbb{C}^d \to \mathbb{R}^{\geqslant 0}$ such that*

$$\|x\|_Y \leqslant \ell_Y(x) \leqslant 2\|x\|_Y$$

in $\mathrm{poly}(d)$ *time with probability* $1 - \frac{1}{\mathrm{poly}(d)}$.

The probability of error in Lemma 8.6 is smaller than the query time, so via a union bound, we may assume that all distance computations in the algorithm are correct up to a factor of 2 with high probability.

Lemma 8.7. *There exists a data structure for c-BNN over Y using space* $\mathrm{poly}(d, n)$ *and time* $O(d \log n) + \mathrm{poly}(d)$*, whenever* $c \gtrsim \left(\frac{C\mathsf{d}^{\beta}}{\beta}\right)^{\frac{\alpha+1}{\alpha^2}}$ *for some constant $C > 0$.*

Proof. From Corollary 8.4, let D be an $O(\mathsf{d}^{\beta})$-ANN data structure over Z using space $\mathrm{poly}(n, d)$ and query time $O(d \log n)$. A data structure D' for c-BNN over Y will obtain a dataset P of n points and proceed as follows:

1. We first discretize every coordinate to $O(\log d)$ bits of precision in order to consider the finite metric space $(\mathcal{M}, d_{\mathcal{M}})$. The discretization produces a set $\mathcal{M} \subseteq \mathbb{C}^d$ where every $x \in \mathbb{C}^d$ with $\|x\|_Y \leqslant c$ has $x' \in \mathcal{M}$ with $\|x - x'\|_Y \leqslant \frac{1}{\mathrm{poly}(d)}$, and for every two points $x, y \in \mathcal{M}$, $\|x - y\|_Y \geqslant \frac{1}{\mathrm{poly}(d)}$. The distance $d_{\mathcal{M}}(x, y) = \|x - y\|_Y$.

2. For every point $x \in \mathcal{M}$, we consider the points $\Phi_\varepsilon(x)$, where $\Phi_\varepsilon \colon Y \to Z$ is the approximate Hölder homeomorphism from Corollary 4.10 from Section 4, defined according to the interpolation between $Y = [A, H]_\beta$ and $Z = [A, H]_{1-\beta}$, where A is a uniformly convex space with p-convexity constant $K_p(A) \leqslant 1$ and $p = \frac{1}{\alpha}$. We initialize the parameters from Corollary 4.10 as
$$R = c, \qquad r = 1, \qquad \text{and} \qquad \varepsilon = \frac{1}{\mathrm{poly}(d)},$$
where ε is small enough so that $x, y \in \mathcal{M}$ satisfy $\|x - y\|_Y \geqslant 5R(\frac{\varepsilon}{\beta})^{1/100}$. Specifically, the map $\Phi_\varepsilon \colon Y \to Z$ satisfies that every $x, y \in \mathcal{M}$,
$$\|\Phi_\varepsilon(x) - \Phi_\varepsilon(y)\|_Z \leqslant \frac{c_0}{\beta} \cdot \|x - y\|_Y^{\alpha} \cdot c^{1-\alpha}, \tag{66}$$
for some constant $c_0 > 0$. In addition, by Lemma 4.12, the Hölder homeomorphism $\Phi_0 \colon Y \to Z$ is invertible and every $x, y \in \mathcal{M}$ satisfy
$$\|x - y\|_Y \leqslant \frac{c_1}{\beta} \cdot \|\Phi_0(x) - \Phi_0(y)\|_Z^{\alpha} c^{1-\alpha}, \tag{67}$$
$$\|\Phi_\varepsilon(x) - \Phi_0(x)\|_Z \leqslant \frac{c_2}{\beta} \cdot c^{1-\alpha}, \tag{68}$$
for constants $c_1, c_2 > 0$.

3. Finally, we use the data structure D to solve the $O(\mathsf{d}^{\beta})$-ANN problem over Z. In particular, we preprocess the dataset $P' = \{\Phi_\varepsilon(p) : p \in P\}$ in D. Upon receiving a query $q \in \mathbb{C}^d$, we query $\Phi_\varepsilon(q)$ to D, and if D returns $\Phi_\varepsilon(p) \in P'$, we return point p.

Note that the required space for the data structure D' is $\mathrm{poly}(n,d)$, and the query time for D' is $O(d\log n)+\mathrm{poly}(d)$, since for each query $q \in \mathbb{C}^d$, computing $\Phi_\varepsilon(q)$ takes $\mathrm{poly}(d)$ time. Suppose $p \in P$ satisfies $\|q-p\|_Y \leqslant 1$. By Corollary 4.10, we have $\|\Phi_\varepsilon(q)-\Phi_\varepsilon(p)\|_Z \leqslant \frac{c_0}{\beta}\cdot c^{1-\alpha}$. Thus, the point $\Phi_\varepsilon(p')$ returned by D satisfies $\|\Phi_\varepsilon(q)-\Phi_\varepsilon(p')\|_Z \lesssim \frac{\mathsf{d}^\beta}{\beta}\cdot c^{1-\alpha}$. Thus, we have that:

$$\begin{aligned}\|\Phi_0(q)-\Phi_0(p')\|_Z &\leqslant \|\Phi_0(q)-\Phi_\varepsilon(q)\|_Z + \|\Phi_\varepsilon(q)-\Phi_\varepsilon(p')\|_Z + \|\Phi_\varepsilon(p')-\Phi_0(p')\|_Z \\ &\overset{(68)}{\leqslant} \frac{2c_2}{\beta}\cdot c^{1-\alpha} + \|\Phi_\varepsilon(q)-\Phi_\varepsilon(p')\|_Z \\ &\lesssim \frac{\mathsf{d}^\beta c^{1-\alpha}}{\beta},\end{aligned}$$

and this implies that for some constant $C>0$,

$$\|q-p'\|_Y \overset{(67)}{\lesssim} \frac{1}{\beta}\left(\frac{C\cdot \mathsf{d}^\beta \cdot c^{1-\alpha}}{\beta}\right)^\alpha c^{1-\alpha} \lesssim \frac{C^\alpha \mathsf{d}^{\beta\alpha}}{\beta^{1+\alpha}}\cdot c^{1-\alpha^2} \ll c,$$

when $c^{\alpha^2} \gg (\frac{C\mathsf{d}^\beta}{\beta})^{\alpha+1}$. □

Proof of Theorem 14. Combining Lemma 8.3 and Lemma 8.7 with Corollary 8.5, we conclude that there exists a data structure for c-ANN with $\mathrm{poly}(n,d)$ space, using query time $O(d\log n)+\mathrm{poly}(d)$ where:

$$c = O\left(\mathsf{d}^{\alpha+\beta}\cdot\left(\frac{C\mathsf{d}^\beta}{\beta}\right)^{\frac{\alpha+1}{\alpha^2}}\right),$$

and recalling the parameters settings of α and β in (64), as well as the fact that $\mathsf{d} = \mathrm{poly}(d)$ gives the desired approximation. □

References

[AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The Multiplicative Weights Update Method: a Meta-Algorithm and Applications. *Theory of Computing*, 8(1):121–164, 2012.

[AI06] 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.

[AI17] Alexandr Andoni and Piotr Indyk. Nearest Neighbors in High-Dimensional Spaces. In Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Tóth, editors, *Handbook of Discrete and Computational Geometry*, pages 1133–1153. CRC Press LLC, 2017.

[AIK09] Alexandr Andoni, Piotr Indyk, and Robert 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)*, pages 865–874, 2009.

[AINR14] 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.

[AIR18] Alexandr Andoni, Piotr Indyk, and Ilya Razenshteyn. Approximate nearest neighbor search in high dimensions. In *Proceedings of ICM 2018 (to appear)*, 2018.

[ALRW17] 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.

[And09] Alexandr Andoni. *Nearest Neighbor Search: the Old, the New, and the Impossible*. PhD thesis, MIT, 2009.

[ANN+17] Alexandr Andoni, Huy L. Nguyen, Aleksandar Nikolov, Ilya Razenshteyn, and Erik Waingarten. Approximate Near Neighbors for General Symmetric Norms. In *Proceedings of the 49th ACM Symposium on the Theory of Computing (STOC '2017)*, pages 902–913, 2017. Available as arXiv:1611.06222.

[ANN+18] Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya Razenshteyn, and Erik 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.

[AR15] 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.

[BCL94] Keith Ball, Eric A. Carlen, and Elliott H. Lieb. Sharp Uniform Convexity and Smoothness Inequalities for Trace Norms. *Inventiones Mathematicae*, 115(1):463–482, 1994.

[BG18] Yair Bartal and Lee-Ad 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.

[BKL06] Alina Beygelzimer, Sham Kakade, and John Langford. Cover Trees for Nearest Neighbor. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '2006)*, pages 97–104, 2006.

[BL76] Jöran Bergh and Jörgen Löfström. *Interpolation spaces: an introduction*. Springer-Verlag, 1976.

[BL00] 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.

[Cal64] Alberto Calderón. Intermediate Spaces and Interpolation, the Complex Method. *Studia Mathematica*, 24(2):113–190, 1964.

[Cha02] Moses Charikar. Similarity Estimation Techniques from Rounding Algorithms. In *Proceedings of the 34th ACM Symposium on the Theory of Computing (STOC '2002)*, pages 380–388, 2002.

[Che69] Jeff Cheeger. A Lower Bound for the Smallest Eigenvalue of the Laplacian. In *Proceedings of the Princeton conference in honor of Professor S. Bochner*, pages 195–199, 1969.

[Cla99] Kenneth L. Clarkson. Nearest neighbor queries in metric spaces. *Discrete and Computational Geometry*, 22(1):63–93, 1999.

[Dah93] Mohamed Daher. Homéomorphismes uniformes entre les sphères unité des espaces d'interpolation. *Comptes Rendus Mathematique*, 316(10):1051–1054, 1993.

[GLS12] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Springer, 2012.

[HIM12] Sariel Har-Peled, Piotr Indyk, and Rajeev Motwani. Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality. *Theory of Computing*, 8(1):321–350, 2012.

[IM98] 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.

[Ind01] Piotr Indyk. On Approximate Nearest Neighbors under ℓ_∞ Norm. *Journal of Computer and System Sciences*, 63(4):627–638, 2001.

[Ind02] Piotr Indyk. Approximate Nearest Neighbor Algorithms for Fréchet Distance via Product Metrics. In *Proceedings of the 18th ACM Symposium on Computational Geometry (SoCG '2002)*, pages 102–106, 2002.

[Ind04] Piotr Indyk. Approximate Nearest Neighbor under Edit Distance via Product Metrics. In *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms (SODA '2004)*, pages 646–650, 2004.

[IT03] Piotr Indyk and Nitin Thaper. Fast Color Image Retrieval via Embeddings. Workshop on Statistical and Computational Theories of Vision (at ICCV), 2003.

[Joh48] Fritz John. Extremum Problems with Inequalities as Subsidiary Conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Kat04] Yitzhak Katznelson. *An introduction to harmonic analysis.* Cambridge University Press, 2004.

[KL04] Robert Krauthgamer and James R. Lee. Navigating Nets: Simple Algorithms for Proximity Search. In *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms (SODA '2004)*, pages 798–807, 2004.

[KLS97] Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $o*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 1997.

[KOR00] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces. *SIAM Journal on Computing*, 30(2):457–474, 2000.

[KR02] David R. Karger and Matthias Ruhl. Finding Nearest Neighbors in Growth-Restricted Metrics. In *Proceedings of the 34th ACM Symposium on the Theory of Computing (STOC '2002)*, pages 741–750, 2002.

[LSV17] Yin Tat Lee, Aaron Sidford, and Santosh S. Vempala. Efficient convex optimization with membership oracles. *arXiv:1706.07357*, 2017.

[Mat96] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.

[Maz29] S. Mazur. Une remarque sur l'homéomorphie des champs fonctionels. *Studia Math.*, 1:83–85, 1929.

[Mil99] Peter Bro Miltersen. Cell Probe Complexity – a Survey. In *Advances in Data Structures*, 1999.

[MN13] Manor Mendel and Assaf Naor. Spectral Calculus and Lipschitz Extension for Barycentric Metric Spaces. *Analysis and Geometry in Metric Spaces*, 1:163–199, 2013.

[Nao12] Assaf Naor. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon–Roichman graphs. *Combinatorics, Probability and Computing*, 21(4):623–634, 2012.

[Nao17] Assaf Naor. A Spectral Gap Precludes Low-Dimensional Embeddings. In *Proceedings of the 33rd International Symposium on Computational Geometry (SoCG '2017)*, pages 50:1–50:16, 2017. Available as arXiv:1611.08861.

[NR06] Assaf Naor and Yuval Rabani. On Approximate Nearest Neighbor Search in ℓ_p, $p > 2$. Manuscript, 2006.

[OR07] Rafail Ostrovsky and Yuval Rabani. Low Distortion Embedding for Edit Distance. *Journal of the ACM*, 54(5):23:1–23:16, 2007.

[OS94] Edward Odell and Thomas Schlumprecht. The distortion problem. *Acta Math.*, 173(2):259–281, 1994.

[Raz17] Ilya Razenshteyn. *High-Dimensional Similarity Search and Sketching: Algorithms and Hardness.* PhD thesis, Massachusetts Institute of Technology, 2017.

[Ric15] Éric Ricard. Hölder Estimates for the Noncommutative Mazur Map. *Archiv der Mathematik*, 104(1):37–45, 2015.

[SS03] Elias M. Stein and Rami Shakarchi. *Complex analysis. Princeton Lectures in Analysis, II.* Princeton University Press, 2003.

[Wid61] David V. Widder. Functions Harmonic in a Strip. *Proceedings of the American Mathematical Society*, 12(1):67–72, 1961.

[Yao81] Andrew Chi-Chih Yao. Should tables be sorted? *Journal of the ACM*, 28(3):615–628, 1981.