Some theory for ordinal embedding

ERY ARIAS-CASTRO

Department of Mathematics, University of California, San Diego 9500 Gilman Dr., La Jolla, CA 92093-0112, USA. E-mail: eariasca@ucsd.edu

Motivated by recent work on ordinal embedding (In *Proceedings of the 27th Conference on Learning Theory* (2014) 40–67), we derive large sample consistency results and rates of convergence for the problem of embedding points based on triple or quadruple distance comparisons. We also consider a variant of this problem where only local comparisons are provided. Finally, inspired by (In *Communication, Control, and Computing* (*Allerton*), *2011 49th Annual Allerton Conference on* (2011) 1077–1084 IEEE), we bound the number of such comparisons needed to achieve consistency.

Keywords: dissimilarity comparisons; landmark multidimensional scaling; non-metric multidimensional scaling (MDS); ordinal embedding

1. Introduction

The problem of ordinal embedding, also called non-metric multidimensional scaling [4], consists of finding an embedding of a set of items based on pairwise distance comparisons. Specifically, suppose that $\delta_{ij} \geq 0$ is some dissimilarity measure between items $i, j \in [n] := \{1, \ldots, n\}$. We assume that $\delta_{ii} = 0$ and $\delta_{ij} = \delta_{ji}$ for all $i, j \in [n]$. These dissimilarities are either directly available but assumed to lack meaning except for their relative magnitudes, or only available via comparisons with some other dissimilarities, meaning that we are only provided with a subset $\mathcal{C} \subset [n]^4$ such that

$$\delta_{ij} < \delta_{k\ell}, \qquad \forall (i, j, k, \ell) \in \mathcal{C}. \tag{1}$$

Note that the latter setting encompasses the former. Given $\mathcal{C}$ and a dimension d, the goal is to embed the items as points $p_1, \ldots, p_n \in \mathbb{R}^d$ in a way that is compatible with the available information, specifically

$$\delta_{ij} < \delta_{k\ell} \quad \Rightarrow \quad \|p_i - p_j\| \leq \|p_k - p_\ell\|, \qquad \forall (i, j, k, \ell) \in \mathcal{C}, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm. The two most common situations are when all the quadruple comparisons are available, meaning $\mathcal{C} = [n]^4$, or all triple comparisons are available, meaning $\mathcal{C} = \{(i, j, i, k) : i, j, k \in [n]\}$, which can be identified with $[n]^3$. We emphasize that the dissimilarities δ_{ij} may not be available, but their ordering is, although only according to the set $\mathcal{C}$. In some situations, the dissimilarities may be available but cannot be considered a meaningful measure of distance, making the direct use of methods for metric embedding inappropriate. This problem has a long history surveyed in [27], with pioneering contributions from [18,19] and [15].

The main question we tackle here is that of *consistency*. Suppose that the items are in fact points $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\delta_{ij} = \|x_i - x_j\|$. Provided with a subset $\mathcal{C} = \mathcal{C}_n$ of dissimilarity comparisons as in (2), is it possible to reconstruct the original points in the large-sample limit

$n \to \infty$? Clearly, the reconstruction can only be up to a similarity transformation – that is, a transformation $f : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that, for some $\lambda > 0$, $\|f(x) - f(y)\| = \lambda \|x - y\|$ for all $x, y \in \mathbb{R}^d$, or equivalently, of the form $f(x) = \lambda R(x) + b$ where R is an orthogonal transformation and b is a constant vector – since such a transformation leaves the distance comparisons unchanged. This question is at the foundation of non-metric multidimensional scaling.

Early work only addressed the continuous case, where the x's span a whole convex subset $U \subset \mathbb{R}^d$. In that setting, the goal becomes to characterize *isotonic functions* on U, that is, functions $f : U \mapsto \mathbb{R}^d$ satisfying

$$\|x - y\| < \|x' - y'\| \quad \Rightarrow \quad \|f(x) - f(y)\| \le \|f(x') - f(y')\|, \qquad \forall x, y, x', y' \in U. \quad (3)$$

Shepard [20] argues that such functions must be similarities, and cites earlier work [3,22] dealing with the case $d = 1$.

Only recently has the finite sample case been formally considered. Indeed, [14] prove a consistency result, showing that if $x_1, \ldots, x_n \in U \subset \mathbb{R}^d$, where U is a bounded, connected, and open subset of $\mathbb{R}^d$ satisfying some additional conditions – for example, a finite union of open balls – and $\mathcal{C} = [n]^4$, then in the large sample limit, it is possible to recover the x's up to a similarity transformation. We note that [14] focus on the strictly isotonic case, where the second inequality in (3) is strict. Our first contribution is an extension of this consistency result for quadruple learning to triple learning where $\mathcal{C} = [n]^3$. In the process, we greatly simplify the arguments of [14] and weaken the conditions on the sampling domain U. We note that [23] have partially solved this problem by a reduction to the problem of embedding a nearest-neighbor graph. However, their arguments are based on an apparently incomplete proof in [25], which is itself based on a rather sophisticated approach. Our proofs are comparatively much simpler and direct.

Our second contribution is to provide rates of convergence, a problem left open by [14].

The last decade has seen a surge of interest in ordinal embedding, motivated by applications to recommender systems and large-scale psychometric studies made available via the internet, for example, databases for music artists similarity [9,16]. Sensor localization [17] is another possible application. Modern datasets being large, all quadruple or triple comparisons are rarely available, motivating the proposal of embedding methods based on a sparse set of comparisons [1,4,12,23]. Terada and Von Luxburg [23] study what they call *local ordinal embedding*, which they define as the problem of embedding an unweighted K-nearest neighbor (K-NN) graph. With our notation, this is the situation where $\mathcal{C} = \{(i, j, k) : \delta_{ij} \le \delta_{i(K)} < \delta_{ik}\}$, $\delta_{i(K)}$ being the dissimilarity between item i and its Kth nearest-neighbor. Terada and Von Luxburg [23] argue that, when the items are points $x_1, \ldots, x_n$ sampled from a smooth density on a bounded, connected, convex, and open subset $U \subset \mathbb{R}^d$ with smooth boundary, then $K = K_n \gg n^{2/(2+d)}(\log n)^{d/(2+d)}$ is enough for consistency. Our third contribution is to consider the related situation where $\mathcal{C} = \{(i, j, k, \ell) : \delta_{ij} < \delta_{k\ell} \text{ and } \max(\delta_{ij}, \delta_{ik}, \delta_{i\ell}) \le \delta_{i(K)}\}$, which provides us with the K-NN graph and also all the quadruple comparisons between the nearest neighbors. In this setting, we are only able to show that $K_n \gg \sqrt{n \log n}$ is enough.

Beyond local designs, which may not be feasible in some settings, [12] consider the problem of adaptively (i.e., sequentially) selecting triple comparisons in order to minimize the number of such comparisons and yet deduce all the other triple comparisons. They consider a few methods, among which a non-metric version of the landmark MDS method of [8]. Less ambitious is the

problem of selecting few comparisons in order to consistently embed the items when these are points in a Euclidean space. Our fourth contribution is to show that one can obtain a consistent embedding with a landmark design based on $a_n n$ queries, where a_n is any diverging sequence. Moreover, the embedding can be computed in (expected) time $\zeta(a_n)n$, for some function $\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$.

The rest of the paper is organized as follows. In Section 2, we state our theoretical results and prove the simpler ones. We then gather the remaining proofs in Section 3. Section 4 concludes the paper with a short discussion.

2. Theory

In this section, we present our theoretical findings. Most proofs are gathered in Section 3.

We already defined isotonic functions in (3). Following [14], we say that a function $f : U \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ is *weakly isotonic* if

$$\|x - y\| < \|x - z\| \quad \Rightarrow \quad \|f(x) - f(y)\| \le \|f(x) - f(z)\|, \qquad \forall x, y, z \in U. \quad (4)$$

Obviously, if a function is isotonic (3), then it is weakly isotonic (4). Weak isotonicity is in fact not much weaker than isotonicity. Indeed, let P be a property (e.g., 'isotonic'), and say that a function $f : U \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ has the property P locally if for each $x \in U$ there is $r > 0$ such that f has property P on $B(x, r) \cap U$, where $B(x, r)$ denotes the open ball with center x and radius r.

Lemma 1. *Any locally weakly isotonic function on an open U is also locally isotonic on U.*

Proof. This is an immediate consequence of [14], Lemma 6, which implies that a weakly isotonic function on $B(x, r)$ is isotonic on $B(x, r/4)$. □

Suppose we have data points $x_1, \ldots, x_n \in \mathbb{R}^d$. Define

$$\Omega_n = \{x_1, \ldots, x_n\}, \qquad \Omega = \bigcup_{n \ge 1} \Omega_n = \{x_n : n \ge 1\}. \quad (5)$$

Let $\delta_{ij} = \|x_i - x_j\|$ and suppose that we are only provided with a subset $\mathcal{C}_n \subset [n]^4$ of distance comparisons as in (1). To an (exact) ordinal embedding $p : [n] \mapsto \mathbb{R}^d$ – which by definition satisfies (2) – we associate the map $\phi_n : \Omega_n \mapsto \mathbb{R}^d$ defined by $\phi_n(x_i) = p_i$ for all $i \in [n]$. We crucially observe that, in the case of all quadruple comparisons ($\mathcal{C}_n = [n]^4$), the resulting map ϕ_n is isotonic on Ω_n; in the case of all triple comparisons ($\mathcal{C}_n = [n]^3$), ϕ_n is only weakly isotonic on Ω_n, instead. In light of this, and the fact that the location, orientation and scale are all lost when only ordinal information is available, the problem of proving consistency of (exact) ordinal embedding reduces to showing that any such embedding is close to a similarity transformation as the sample size increases $n \to \infty$. This is exactly what [14] do under some assumptions.

2.1. Ordinal embedding based on all triple comparisons

Our first contribution is to extend the consistency results of [14] on quadruple learning to triple learning. Following their presentation, we start with a result where the sample is infinite, which is only a mild generalization of [14], Theorem 3.

Theorem 1. *Let $U \subset \mathbb{R}^d$ be bounded, connected and open. Suppose Ω is dense in U and consider a locally weakly isotonic function $\phi : \Omega \mapsto \mathbb{R}^d$. Then there is a similarity transformation S that coincides with ϕ on Ω.*

The proof is largely based on that of [14], Theorem 3, but a bit simpler; see Section 3.1.

We remark that there can only be one similarity with the above property, since similarities are affine transformations, and two affine transformations of $\mathbb{R}^d$ that coincide on $d + 1$ affine independent points are necessarily identical.

In this theorem, the set Ω is dense in an open subset of $\mathbb{R}^d$, and therefore infinite. In fact, [14] use this theorem as an intermediary result for proving consistency as the sample size increases. In fact, most of their paper is dedicated to establishing this, as their arguments are quite elaborate. We found a more direct route by 'tending to the limit as soon as possible', based on Lemma 2 below, which is at the core of the Arzelà–Ascoli theorem.

For the remaining of this section, we consider the finite sample setting:

$$U \subset \mathbb{R}^d \text{ is bounded, connected and open,}$$

$$\Omega_n = \{x_1, \ldots, x_n\} \subset U \text{ is such that } \Omega := \{x_n : n \geq 1\} \text{ is dense in } U, \tag{6}$$

$$\text{and } \phi_n : \Omega_n \mapsto Q \subset \mathbb{R}^d \text{ is a function with values in a bounded set } Q.$$

In the context of (6), we implicitly extend ϕ_n to Ω, for example, by setting $\phi_n(x) = q$ for all $x \in \Omega \setminus \Omega_n$, where q is any fixed point in Q, although the following holds for any extension.

Lemma 2. *Consider $\Omega_n \subset \mathbb{R}^d$ finite and $\phi_n : \Omega_n \mapsto Q \subset \mathbb{R}^d$, where Q is bounded. Then there is $N \subset \mathbb{N}$ infinite such that $\phi(x) := \lim_{n \in N} \phi_n(x)$ exists for all $x \in \Omega := \bigcup_n \Omega_n$.*

This is called the diagonal process in [13], Problem D, Chapter 7. Although the result is classical, we provide a proof for completeness.

Proof of Lemma 2. Without loss of generality, suppose $\Omega_n = \{x_1, \ldots, x_n\}$. Let $N_0 = \mathbb{N}$. Since $(\phi_n(x_1) : n \in N_0) \in Q$ and Q is bounded, there is $N_1 \subset N_0$ infinite such that $\lim_{n \in N_1} \phi_n(x_1)$ exists. In turn, since $(\phi_n(x_2) : n \in N_1)$ is bounded, there is $N_2 \subset N_1$ infinite such that $\lim_{n \in N_2} \phi_n(x_2)$ exists. Continuing this process – which formally corresponds to a recursion – we obtain $\cdots \subset N_{k+1} \subset N_k \subset \cdots \subset N_1 \subset N_0 = \mathbb{N}$ such that, for all k, N_k is infinite and $\lim_{n \in N_k} \phi_n(x_k)$ exists. Let n_k denote the kth element (in increasing order) of N_k and note that $(n_k : k \geq 1)$ is strictly increasing. Define $N = \{n_k : k \geq 1\}$. Since $\{n_p, p \geq k\} \subset N_k$, we have $\lim_{n \in N} \phi_n(x_k) = \lim_{n \in N_k} \phi_n(x_k)$, and this is valid for all $k \geq 1$. $\square$

Corollary 1. *Consider the setting* (6) *and assume that ϕ_n is weakly isotonic. Then (ϕ_n) is sequentially compact for the pointwise convergence topology for functions on Ω and all the functions where it accumulates are similarity transformations restricted to Ω.*

The corresponding result ([14], Theorem 4) was obtained for isotonic (instead of weakly isotonic) functions and for domains U that are finite unions of balls, and the convergence was uniform instead of pointwise. For now, we provide a proof of Corollary 1, which we derive as a simple consequence of Theorem 1 and Lemma 2.

Proof of Corollary 1. Lemma 2 implies that (ϕ_n) is compact for the pointwise convergence topology. Let ϕ be an accumulation point of (ϕ_n), meaning that there is $N \subset \mathbb{N}$ infinite such that $\phi(x) = \lim_{n \in N} \phi_n(x)$ for all $x \in \Omega$. Take $x, y, z \in \Omega$ such that $\|x - y\| < \|x - z\|$. By definition, there is m such that $x, y, z \in \Omega_m$, and therefore $\|\phi_n(x) - \phi_n(y)\| \leq \|\phi_n(x) - \phi_n(z)\|$ for all $n \geq m$. Passing to the limit along $n \in N$, we obtain $\|\phi(x) - \phi(y)\| \leq \|\phi(x) - \phi(z)\|$. Hence, ϕ is weakly isotonic on Ω and, by Theorem 1, it is therefore the restriction of a similarity transformation to Ω. $\qquad\square$

It is true that ([14], Theorem 4) establishes a uniform convergence result. We do the same in Theorem 2 below, but with much simpler arguments. The key are the following two results bounding the modulus of continuity of a (resp. weakly) isotonic function. We note that the second result (for weakly isotonic functions) is very weak but sufficient for our purposes here. For $\Lambda \subset V \subset \mathbb{R}^d$, define $\delta_H(\Lambda, V) = \sup_{y \in V} \inf_{x \in \Lambda} \|y - x\|$, which is their Hausdorff distance. We say that $(y_i : i \in I) \subset \mathbb{R}^d$ is an η-packing if $\|y_i - y_j\| \geq \eta$ for all $i \neq j$. We recall that the size of the largest η-packing of a Euclidean open ball of radius r is of order exactly $(r/\eta)^{-d}$. For a set $V \subset \mathbb{R}^d$, let $\mathrm{diam}(V) = \sup_{x,y \in V} \|x - y\|$ be its diameter and let

$$\rho(V) = \arg\sup_{r > 0} \big\{ \exists v \in V : B(v, r/2) \subset V \big\}, \tag{7}$$

which is the diameter of a largest ball inscribed in V.

Everywhere in the paper, d is fixed, and in fact implicitly small as we assume repeatedly that the sample (of size n) is dense in a full-dimensional domain of $\mathbb{R}^d$. In particular, all the implicit constants of proportionality that follow depend solely on d.

Lemma 3. *Let $V \subset \mathbb{R}^d$ be open. Consider $\Lambda \subset V$ and set $\varepsilon = \delta_H(\Lambda, V)$. Let $\psi : \Lambda \mapsto Q$ be isotonic, where $Q \subset \mathbb{R}^d$ is bounded. There is $C \propto \mathrm{diam}(Q)/\rho(V)$, such that*

$$\big\|\psi(x) - \psi(x')\big\| \leq C\big(\big\|x - x'\big\| + \varepsilon\big), \qquad \forall x, x' \in \Lambda. \tag{8}$$

Proof. The proof is based on the fact that an isotonic function transforms a packing into a packing. Take $x, x' \in \Lambda$ such that $\xi := \|\psi(x) - \psi(x')\| > 0$, and let $\eta = \|x - x'\|$. Since V is open it contains an open ball of diameter $\rho(V)$. Let $y_1, \ldots, y_m$ be an $(\eta + 3\varepsilon)$-packing of such a ball with $m \geq C_1(\rho(V)/(\eta + \varepsilon))^d$ for some constant C_1 depending only on d. Then let $x_1, \ldots, x_m \in \Lambda$ such that $\max_i \|y_i - x_i\| \leq \varepsilon$. By the triangle inequality, for all $i \neq j$ we have $\|x_i - x_j\| \geq \|y_i - y_j\| - 2\varepsilon \geq \eta + \varepsilon > \|x - x'\|$. Because ψ is isotonic, we have $\|\psi(x_i) - \psi(x_j)\| \geq \xi$, so

that $\psi(x_1), \ldots, \psi(x_m)$ is a ξ-packing. Therefore, there is a constant C_2 depending only on d such that $m \le C_2(\text{diam}(Q)/\xi)^d$. We conclude that $\xi \le (C_2/C_1)^{1/d}(\text{diam}(Q)/\rho(V))(\eta + \varepsilon)$. $\square$

For $V \subset \mathbb{R}^d$ and $h > 0$, let $V^h = \{x \in V : \exists y \in V \text{ s.t. } x \in B(y, h) \subset V\}$. We note that V^h is the complement of the h-convex hull of $V^c := \mathbb{R}^d \setminus V$ – see [6] and references therein.

Lemma 4. *In the context of Lemma* 3, *if ψ is only weakly isotonic, then there is $C \propto \text{diam}(Q)$, such that for all $h > 0$,*

$$\|\psi(x) - \psi(x')\| \le C(\|x - x'\|/h + \sqrt{\varepsilon/h})^{1/d}, \qquad \forall x \in \Lambda \cap V^h, \forall x' \in \Lambda. \tag{9}$$

Proof. Assume that $V^h \ne \varnothing$, for otherwise there is nothing to prove. Take $x \in \Lambda \cap V^h$ and $x' \in \Lambda$ such that $\xi := \|\psi(x) - \psi(x')\| > 0$, and let $\eta = \|x - x'\|$. Because ψ is bounded, it is enough to prove the result when $\eta, \varepsilon < h/2$. Let $y \in V$ be such that $x \in B(y, h) \subset V$. There is $y' \in B(y, h)$ such that $y \in [xy']$ and $\|x - y'\| \ge 2h/3$. Define $u = (y' - x)/\|y' - x\|$. Let $z_0 = x$, and for $j \ge 1$, define $z_j = z_{j-1} + (\eta + 5j\varepsilon)u$. Let $k \ge 0$ be maximum such that $\sum_{j=1}^{k}(\eta + 5j\varepsilon) < h/2$. Since k satisfies $k\eta + 5k^2\varepsilon \ge h/2$, we have $k \ge \min(h/(4\eta), \sqrt{h/(10\varepsilon)})$. By construction, for all $j \in [k]$, $z_j \in [xy']$ and $B(z_j, 2\varepsilon) \subset B(y, h)$. Let $x_{-1} = x'$, $x_0 = x$ and take $x_1, \ldots, x_k \in \Lambda$ such that $\max_j \|x_j - z_j\| \le \varepsilon$. By the triangle inequality, for $j = 2, \ldots, k$,

$$\|x_j - x_{j-1}\| \ge \|z_j - z_{j-1}\| - 2\varepsilon \ge \|z_{j-1} - z_{j-2}\| + 3\varepsilon \ge \|x_{j-1} - x_{j-2}\| + \varepsilon,$$

which implies by induction that

$$\|x_j - x_{j-1}\| \ge \|x_1 - x_0\| + \varepsilon \ge \|z_1 - z_0\| = \eta + 5\varepsilon > \|x - x'\|.$$

By weak isotonicity, this implies that $\|\psi(x_j) - \psi(x_{j-1})\| \ge \|\psi(x) - \psi(x')\| = \xi$. We also have, for any $i, j \in [k]$ such that $1 \le i \le j - 2$,

$$\|x_j - x_i\| \ge \|z_j - z_i\| - 2\varepsilon \ge \|z_j - z_{j-1}\| + \eta + 5\varepsilon - 2\varepsilon \ge \|x_j - x_{j-1}\| + \eta + \varepsilon.$$

By weak isotonicity, this implies that $\|\psi(x_j) - \psi(x_i)\| \ge \|\psi(x_j) - \psi(x_{j-1})\|$ for all $0 \le i < j \le k$. Consequently, $(\psi(x_j) : j \in [k])$ forms a ξ-packing of Q. Hence, $k \le C'(\text{diam}(Q)/\xi)^d$, for some constant C'. We conclude with the lower bound on k. $\square$

From this control on the modulus of continuity, we obtain a stronger version of Corollary 1.

Theorem 2. *Under the same conditions as Corollary* 1, *we have the stronger conclusion that there is a sequence (S_n) of similarities such that, for all $h > 0$, $\max_{x \in \Omega_n \cap U^h} \|\phi_n(x) - S_n(x)\| \to 0$ as $n \to \infty$. If in fact each ϕ_n is isotonic, then this remains true when $h = 0$.*

We remark that when U is a connected union of a possibly uncountable number of open balls of radius at least $h > 0$, then $U = U^h$. This covers the case of a finite union of open balls considered in [14]. We also note that, if U is bounded and open, and ∂U has bounded curvature, then there is $h > 0$ such that $U = U^h$. This follows from the fact that, in this case, U^c has positive reach

[10], and is therefore h-convex when h is below the reach by[1] [6], Proposition 1. Moreover, our arguments can be modified to accommodate sets U with boundaries that are only Lipschitz, by reasoning with wedges in Lemma 4.

Theorem 2 now contains ([14], Theorem 4), and extends it to weakly isotonic functions and to more general domains U. Overall, our proof technique is much simpler, shorter, and elementary.

Define $\varepsilon_n = \delta_H(\Omega_n, U)$, which quantifies the density of Ω_n in U. Because $\Omega_{n+1} \subset \Omega_n$ and Ω is dense in U, we have $\varepsilon_n \searrow 0$ as $n \to \infty$.

Proof of Theorem 2. Let ϕ be an accumulation point of (ϕ_n) for the pointwise convergence topology, meaning there is $N \subset \mathbb{N}$ infinite such that $\phi(x) = \lim_{n \in N} \phi_n(x)$ for all $x \in \Omega$. We show that, in fact, the convergence is uniform.

First, suppose that each ϕ_n is isotonic. In that case, Lemma 3 implies the existence of a constant $C > 0$ such that $\|\phi_n(x) - \phi_n(x')\| \leq C(\|x - x'\| + \varepsilon_n)$ for all $x, x' \in \Omega_n$, and for all n. Passing to the limit along $n \in N$, we get $\|\phi(x) - \phi(x')\| \leq C\|x - x'\|$ for all $x, x' \in \Omega$. (In fact, we already knew this from Corollary 1, since we learned there that ϕ coincides with a similarity, and is therefore Lipschitz.) Fix $\varepsilon > 0$. There is m such that $\varepsilon_m \leq \varepsilon$. Then there is $m' \geq m$ such that $\max_{i \in [m]} \|\phi_n(x_i) - \phi(x_i)\| \leq \varepsilon$ for all $n \in N$ with $n \geq m'$. For such an n, and $x \in \Omega_n$, let $i \in [m]$ be such that $\|x - x_i\| \leq \varepsilon_m$. By the triangle inequality,

$$
\begin{aligned}
\|\phi_n(x) - \phi(x)\| &\leq \|\phi_n(x) - \phi_n(x_i)\| + \|\phi_n(x_i) - \phi(x_i)\| + \|\phi(x_i) - \phi(x)\| \\
&\leq C(\|x - x_i\| + \varepsilon_n) + \|\phi_n(x_i) - \phi(x_i)\| + C\|x_i - x\| \\
&\leq C(\varepsilon_m + \varepsilon_n) + \varepsilon + C\varepsilon_m \leq (3C + 1)\varepsilon.
\end{aligned}
$$

Since $x \in \Omega$ is arbitrary and ε can be taken as small as desired, this shows that the sequence $(\phi_n : n \in N)$ convergences uniformly to ϕ over $(\Omega_n : n \in N)$.

When the ϕ_n are only weakly isotonic, we use Lemma 4 to get a constant $C > 0$ depending on $\mathrm{diam}(Q)$ and $h > 0$ such that $\|\phi_n(x) - \phi_n(x')\| \leq C(\|x - x'\| + \sqrt{\varepsilon_n})^{1/d}$ for all $x \in \Omega_n \cap U^h$ and all $x' \in \Omega_n$, and for all n. Passing to the limit along $n \in N$, we get $\|\phi(x) - \phi(x')\| \leq C\|x - x'\|^{1/d}$ for all $x, x' \in \Omega$. (In fact, $\|\phi(x) - \phi(x')\| \leq C\|x - x'\|$ for all $x, x' \in \Omega$ from Corollary 1, as explained above.) The rest of the arguments are completely parallel. We conclude that $(\phi_n : n \in N)$ convergences uniformly to ϕ over $(\Omega_n \cap U^h : n \in N)$.

Let $\mathcal{S}$ denote the similarities of $\mathbb{R}^d$. For any functions $\phi, \psi : \Omega \mapsto \mathbb{R}^d$, define $\delta_n(\phi, \psi) = \max_{x \in \Omega_n \cap U^h} \|\phi(x) - \psi(x)\|$, and also $\delta_n(\phi, \mathcal{S}) = \inf_{S \in \mathcal{S}} \delta_n(\phi, S)$. Our end goal is to show that $\delta_n(\phi_n, \mathcal{S}) \to 0$ as $n \to \infty$. Suppose this is not the case, so that there is $\eta > 0$ and $N \subset \mathbb{N}$ infinite such that $\delta_n(\phi_n, \mathcal{S}) \geq \eta$ for all $n \in N$. By Corollary 1, there is $N_1 \subset N$ and $S \in \mathcal{S}$ such that $S(x) = \lim_{n \in N_1} \phi_n(x)$ for all $x \in \Omega$. As we showed above, the convergence is in fact uniform over $(\Omega_n \cap U^h : n \in N_1)$, meaning $\lim_{n \in N_1} \delta_n(\phi_n, S) = 0$. At the same time, we have $\delta_n(\phi_n, S) \geq \delta_n(\phi_n, \mathcal{S}) \geq \eta$. We therefore have a contradiction. $\square$

---

[1]This proposition is stated for compact sets (which is not the case of U^c) but easily extends to the case where set is closed with compact boundary.

2.2. Rates of convergence

Beyond consistency, we are able to derive convergence rates. We do so for the isotonic case, that is, the quadruple comparison setting. Recall that $\varepsilon_n = \delta_H(\Omega_n, U)$.

Theorem 3. *Consider the setting* (6) *with ϕ_n isotonic. There is C depending only on (d, U), and a sequence of similarities S_n such that $\max_{x \in \Omega_n} \|\phi_n(x) - S_n(x)\| \leq C \operatorname{diam}(Q)\varepsilon_n$. If $U = U^h$ for some $h > 0$, then $C = C'/\operatorname{diam}(U)$ where C' is a function of $(d, h/\operatorname{diam}(U), \rho(U)/\operatorname{diam}(U))$.*

The proof of Theorem 3 is substantially more technical than the previous results, and thus postponed to Section 3. Although [14] are not able to obtain rates of convergence, the proof of Theorem 3 bares resemblance to their proof technique, and in particular, is also based on a result of [2] on the approximation of ε-*isometries*; see Lemma 18. We will also make use of a related result of [24] on the approximation of approximately midlinear functions; see Lemma 17. We mention that we know of a more elementary proof that only makes use of [2], but yields a slightly slower rate of convergence.

We note that there is a constant c depending only on d such that $\varepsilon_n \geq cn^{-1/d}$. This is because U being open, it contains an open ball, and this lower bound trivially holds for an open ball. And such a lower bound is achieved when the x_i's are roughly regularly spread out over U. If instead the x_i's are i.i.d. uniform in U, and U is sufficiently regular – for example, $U = U^h$ for some $h > 0$ – then $\varepsilon_n = O(\log(n)/n)^{1/d}$, as is well-known. This would give the rate, and we do not know whether it is optimal, even in dimension $d = 1$.

Remark. We are only able to get a rate in $\sqrt{\varepsilon_n}$ for the weakly isotonic case. We can do so by adapting of the arguments underlying Theorem 3, but only after assuming that $U = U^h$ for some $h > 0$ and resolving a few additional complications.

2.3. Ordinal embedding with local comparisons

Terada and Von Luxburg [23] consider the problem of embedding an unweighted nearest-neighbor graph, which as we saw in the Introduction, is a special case of ordinal embedding. Their arguments – which, as we explained earlier, seem incomplete at the time of writing – indicate that $K = K_n \gg n^{2/(2+d)}(\log n)^{d/(2+d)}$ is enough for consistently embedding a K-NN graph.

We consider here a situation where we have more information, specifically, all the distance comparisons between K-nearest-neighbors. Formally, this is the situation where

$$\mathcal{C}_n = \big\{(i, j, k, \ell) : \delta_{ij} < \delta_{k\ell} \text{ and } \{j, k, \ell\} \subset N_{K_n}(i)\big\},$$

where $N_K(i)$ denotes the set of the K items nearest item i. If the items are points $\Omega_n = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, an exact ordinal embedding ϕ_n is only constrained to be locally weakly isotonic as we explain now. We start by stating a standard result which relates a K-NN graph to an r-ball graph.

Lemma 5. *Let $U \subset \mathbb{R}^d$ be bounded, connected and open, and such that $U = U^h$ for some $h > 0$. Sample $x_1, \ldots, x_n$ i.i.d. from a density f supported on U with (essential) range in $(0, \infty)$ strictly. There is a constant C such that, if $K := [nr^d] \geq C \log n$, then with probability tending to 1,*

$$\mathrm{Neigh}_{K/2}(x_i) \subset \left\{x_j : \|x_j - x_i\| \leq r\right\} \subset \mathrm{Neigh}_{2K}(x_i), \qquad \forall i \in [n],$$

where $\mathrm{Neigh}_K(x_i)$ denotes the set of the K points in $\{x_j : j \in [n]\}$ nearest x_i.

The proof is postponed to Section 3 and only provided for completeness. Therefore, assuming that $K \geq C \log n$, where C is the constant of Lemma 5, we may equivalently consider the case where

$$\mathcal{C}_n = \left\{(i, j, k, \ell) : \delta_{ij} < \delta_{k\ell} \text{ and } \max(\delta_{ij}, \delta_{ik}, \delta_{i\ell}) < r_n\right\},$$

for some given $r_n > 0$. An exact embedding $\phi_n : \Omega_n \mapsto \mathbb{R}^d$ in that case is isotonic on $\Omega_n \cap B(x, r_n)$ for any $x \in \Omega_n$. We require in addition that

$$\|x - x'\| < r_n \leq \|x^\dagger - x^\ddagger\| \implies$$
$$\|\phi_n(x) - \phi_n(x')\| \leq \|\phi_n(x^\dagger) - \phi_n(x^\ddagger)\|, \qquad \forall x, x', x^\dagger, x^\ddagger \in \Omega_n. \tag{10}$$

This is a reasonable requirement since it is possible to infer it from $\mathcal{C}_n$. Indeed, for $k, \ell \in [n]$, we have $\delta_{k\ell} < r_n$ if, and only if, $(k, k, k, \ell) \in \mathcal{C}_n$ or $(\ell, \ell, \ell, k) \in \mathcal{C}_n$. (Here we assume that $\delta_{ii} = 0$ for all i and $\delta_{ij} > 0$ if $i \neq j$, as is the case for Euclidean distances.) We can still infer this even if the quadruples in $\mathcal{C}_n$ must include at least three distinct items. Indeed, suppose $k, \ell \in [n]$ are such that there is no i such that $(i, k, i, \ell) \in \mathcal{C}_n$ or $(i, \ell, i, k) \in \mathcal{C}_n$, then (a) $\delta_{ik} = \delta_{i\ell}$ for all i such that $\max(\delta_{ik}, \delta_{i\ell}) < r_n$, or (b) $\delta_{k\ell} \geq r_n$. Assume that $r_n \geq C\varepsilon_n$ with $C > 0$ sufficiently large, so that situation (a) does not happen. Conversely, if (k, ℓ) is such that $\delta_{k\ell} < r_n$, then when (a) does not happen, there is i such that $(i, k, i, \ell) \in \mathcal{C}_n$ or $(i, \ell, i, k) \in \mathcal{C}_n$.

Theorem 4. *Consider the setting (6) and assume in addition that $U = U^h$ for some $h > 0$, and that ϕ_n is isotonic over balls of radius r_n and satisfies (10). There is a constant $C > 0$ depending only on $(d, h, \rho(U), \mathrm{diam}(U), \mathrm{diam}(Q))$ and similarities S_n such that $\max_{x \in \Omega_n} \|\phi_n(x) - S_n(x)\| \leq C\varepsilon_n/r_n^2$.*

Assume the data points are generated as in Lemma 5. In that case, we have $\varepsilon_n = O(\log(n)/n)^{1/d}$ and Theorem 4 implies consistency when $r_n \gg (\log(n)/n)^{1/d}$. By Lemma 5, this corresponds to the situation where we are provided with comparisons among K_n-nearest neighbors with $K_n \gg \sqrt{n \log n}$. If the result of [23] holds in all rigor, then this is a rather weak rate. Initially, we were hopeful to prove that $K_n \gg \log n$ is enough for consistency, but even that we were not able to show.

2.4. Landmark ordinal embedding

Inspired by [12], we consider the situation where there are landmark items indexed by $L_n \subset [n]$, and we are given all distance comparisons from any point to the landmarks. Formally, with triple

comparisons, this corresponds to the situation where

$$\mathcal{C}_n = \left\{ (i, j, k) \in [n] \times L_n^2 : \delta_{ij} < \delta_{ik} \right\}.$$

If the items are points $\Omega_n = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, an exact ordinal embedding ϕ_n is only constrained to be weakly isotonic on the set of landmarks and, in addition, is required to respect the ordering of the distances from any point to the landmarks. The following is an easy consequence of Theorem 2.

Corollary 2. *Theorem* 2 *remains valid in the landmark triple comparisons setting* (*meaning with* ϕ_n *as just described*) *as long as the landmarks become dense in* U.

Jamieson and Nowak [12] study the number of triple comparisons that are needed for exact ordinal embedding. With a counting argument, they show that at least $Cn \log n$ comparisons are needed, where C is a constant depending only on d. If we only insist that the embedding respects the comparisons that are provided, then Corollary 2 implies that a landmark design is able to be consistent as long as the landmarks become dense in U. This consistency implies that, as the sample size increases, an embedding that respects the landmark comparisons also respects all other comparisons approximately. This is achieved with $O(n\ell_n^2 + \ell_n^3)$ triple comparisons, where $\ell_n := |L_n|$ is the number of landmarks, and the conditions of Corollary 2 can be fulfilled with $\ell_n \to \infty$ at any speed, so that the number of comparisons is nearly linear in n.

Proof of Corollary 2. We focus on the weakly isotonic case, where we assume that $U = U^h$ for some $h > 0$. Let $\Lambda_n = \{x_l : l \in L_n\}$ denote the set of landmarks. Since Λ_n becomes dense in U, meaning $\eta_n := \delta_H(\Lambda_n, U) \to 0$, by Theorem 2, there is a sequence of similarities S_n such that $\zeta_n := \max_{x \in \Lambda_n} \|\phi_n(x) - S_n(x)\| \to 0$. Now, for $x \in \Omega_n$, let $\tilde{x} \in \Lambda_n$ such that $\|x - \tilde{x}\| \leq \eta_n$. We have

$$\|\phi_n(x) - S_n(x)\| \leq \|\phi_n(x) - \phi_n(\tilde{x})\| + \|\phi_n(\tilde{x}) - S_n(\tilde{x})\| + \|S_n(\tilde{x}) - S_n(x)\|. \quad (11)$$

The first term is bounded by $C\eta_n^{1/2d}$ by Lemma 4, for some constant C. The middle term is bounded by ζ_n. For the third term, express S_n in the form $S_n(x) = \beta_n R_n(x) + b_n$, where $\beta_n \in \mathbb{R}$, R_n is an orthogonal transformation, and $b_n \in \mathbb{R}^d$. Take two distinct landmarks $x^\dagger, x^\ddagger \in \Lambda_n$ such that $\|x^\dagger - x^\ddagger\| \geq \operatorname{diam}(U)/2$, which exist when n is sufficiently large. Since $\|S_n(x^\dagger) - S_n(x^\ddagger)\| = \beta_n \|x^\dagger - x^\ddagger\| \geq \beta_n \operatorname{diam}(U)/2$ and, at the same time,

$$\|S_n(x^\dagger) - S_n(x^\ddagger)\| \leq \|S_n(x^\dagger) - \phi_n(x^\dagger)\| + \|\phi_n(x^\dagger) - \phi_n(x^\ddagger)\| + \|\phi_n(x^\ddagger) - S_n(x^\ddagger)\|$$
$$\leq \zeta_n + \operatorname{diam}(Q) + \zeta_n \leq 2\operatorname{diam}(Q), \qquad \text{eventually,}$$

we have $\beta_n \leq \bar{\beta} := 4\operatorname{diam}(Q)/\operatorname{diam}(U)$. Hence, the third term on the RHS of (11) is bounded by $\bar{\beta}\eta_n$. Thus, the RHS of (11) is bounded by $C\eta_n^{1/2d} + \zeta_n + \bar{\beta}\eta_n$, which tends to 0 as $n \to \infty$. This being valid for any $x \in \Omega_n$, we conclude. $\square$

We remark that at the very end of the proof, we obtained a rate of convergence as a function of the density of the landmarks and the convergence rate implicit in Theorem 2. This leads to the

following rate for the quadruple comparisons setting, which corresponds to the situation where

$$\mathcal{C}_n = \left\{(i, j, k) \in [n] \times L_n^2 : \delta_{ij} < \delta_{ik}\right\} \cup \left\{(i, j, k, \ell) \in L_n^4 : \delta_{ij} < \delta_{ik}\right\}.$$

Here, ϕ_n is constrained to be isotonic on the set of landmarks and, as before, is required to respect the ordering of the distances from any data point to the landmarks.

Corollary 3. *Consider the setting* (6) *in the landmark quadruple comparisons setting (meaning with ϕ_n as just described). Let Λ_n denote the set of landmarks and set $\eta_n = \delta_H(\Lambda_n, U)$. There is a constant $C > 0$ and a sequence of similarities S_n such that $\max_{x \in \Omega_n} \|\phi_n(x) - S_n(x)\| \leq C\eta_n$.*

Proof. The proof is parallel to that of Corollary 2. Here, we apply Theorem 3 to get $\zeta_n \leq C_0\eta_n$. This bounds the second term on the RHS of (11). The first term is bounded by $C_1\eta_n$ by Lemma 3, while the third term is bounded by $\bar{\beta}\eta_n$ as before. (C_0, C_1 are constants.) $\square$

Computational complexity. We now discuss the computational complexity of ordinal embedding with a landmark design. The obvious approach has two stages. In the first stage, the landmarks are embedded. This is the goal of [1], for example. Here, we use brute force.

Proposition 1. *Suppose that m items are in fact points in Euclidean space and their dissimilarities are their pairwise Euclidean distances. Then whether in the triple or quadruple comparisons setting, an exact ordinal embedding of these m items can be obtained in finite expected time.*

Proof. The algorithm we discuss is very naive: we sample m points i.i.d. from the uniform distribution on the unit ball, and repeat until the ordinal constraints are satisfied. Since checking the latter can be done in finite time, it suffices to show that there is a strictly positive probability that one such sample satisfies the ordinal constraints. Let $\mathcal{X}_m$ denote the set of m-tuples $(x_1, \ldots, x_m) \in B(0, 1)$ that satisfy the ordinal constraints, meaning that $\|x_i - x_j\| < \|x_k - x_\ell\|$ when $(i, j, k, \ell) \in \mathcal{C}$. Seeing $\mathcal{X}_n$ as a subset of $B(0, 1)^m \subset \mathbb{R}^{dm}$, it is clearly open. And sampling $x_1, \ldots, x_m$ i.i.d. from the uniform distribution on $B(0, 1)$ results in sampling $(x_1, \ldots, x_m)$ from the uniform distribution on $B(0, 1)^m$, which assigns a positive mass to any open set. $\square$

In the second stage, each point that is not a landmark is embedded based on the order of its distances to the landmarks. We quickly mention the work [7], who develops a convex method for performing this task. Here, we are contented with knowing that this can be done, for each point, in finite time, function of the number of landmarks. For example, a brute force approach starts by computing the Voronoi diagram of the landmarks, and iteratively repeats within each cell, creating a tree structure. Each point that is not a landmark is placed by going from the root to a leaf, and choosing any point in that leaf cell, say its barycenter.

Thus, if there are ℓ landmarks, the first stage is performed in expected time $F(\ell)$, and the second stage is performed in time $(n - \ell)G(\ell)$. The overall procedure is thus computed in expected time $F(\ell) + (n - \ell)G(\ell)$.

Remark 1. The procedure described above is not suggested as a practical means to perform ordinal embedding with a landmark design. The first stage, described in Proposition 1, has finite expected time, but likely *not* polynomial in m. For a practical method, we can suggest the following:

1. Embed the landmarks using the semidefinite [1] (which solves a semidefinite program) or the method of [23] (which uses an iterative minimization-majorization strategy).
2. Embed the remaining points using the method of [7] (which solves a quadratic program).

We cannot provide any theoretical guarantees for this method.

3. More proofs

In this section, we gather the remaining proofs and some auxiliary results. We introduce some additional notation and basic concepts. For $z_1, \dots, z_m \in \mathbb{R}^d$, let $\mathrm{Aff}(z_1, \dots, z_m)$ denote their affine hull, meaning the affine subspace they generate in $\mathbb{R}^d$. For a vector x in a Euclidean space, let $\|x\|$ denote its Euclidean norm. For a matrix $M \in \mathbb{R}^{p \times q}$, let $\|M\|$ denote its usual operator norm, meaning, $\|M\| = \max\{\|Mx\| : \|x\| \le 1\}$ and $\|M\|_F = \sqrt{\mathrm{tr}(M^\top M)}$ its Frobenius norm.

Regular simplexes. These will play a central role in our proofs. We say that $z_1, \dots, z_m \in \mathbb{R}^d$, with $m \ge 2$, form a regular simplex if their pairwise distances are all equal. We note that, necessarily, $m \le d + 1$, and that regular simplexes in the same Euclidean space and with same number of (distinct) nodes m are similarity transformations of each other – for example, segments ($m = 2$), equilateral triangles ($m = 3$), tetrahedron ($m = 4$). By recursion on the number of vertices, m, it is easy to prove the following.

Lemma 6. *Let $z_1, \dots, z_m$ form a regular simplex with edge length 1 and let μ denote the barycenter of $z_1, \dots, z_m$. Then $\|\mu - z_i\| = \sqrt{(m-1)/2m}$, and if $z, z_1, \dots, z_m$ form a regular simplex, then $\|z - \mu\| = \sqrt{(m+1)/2m}$. (In dimension m, there are exactly two such points z.)*

3.1. Proof of Theorem 1

We assume $d \ge 2$. See [14] for the case $d = 1$. We divide the proof into several parts.

Continuous extension. Lemma 4 implies that ϕ is locally uniformly continuous. Indeed, take $x_0 \in \Omega$ and let $r > 0$ such that $B(x_0, r) \subset U$ and ϕ is weakly isotonic on $B(x_0, r) \cap \Omega$. Applying Lemma 4 with $V = B(x_0, r)$ and $\Lambda = \Omega \cap B(x_0, r)$ – so that $\delta_H(\Lambda, V) = 0$ because Λ is dense in V – and noting that $V^r = V$, yields a constant $C_r > 0$ such that $\|\phi(x) - \phi(x')\| \le C_r \|x - x'\|^{1/d}$, for all $x, x' \in \Omega \cap B(x_0, r)$. Being locally uniformly continuous, we can uniquely extend ϕ to a continuous function on U, also denoted by ϕ. By continuity, this extension is locally weakly isotonic on U.

Isosceles preservation. Sikorska and Szostok [21] say that a function $f : V \subset \mathbb{R}^d \to \mathbb{R}^d$ preserves isosceles triangles if

$$\|x - y\| = \|x - z\| \quad \Rightarrow \quad \|f(x) - f(y)\| = \|f(x) - f(z)\|, \qquad \forall x, y, z \in V.$$

In our case, by continuity, we also have that ϕ preserves isosceles triangles locally. Indeed, for the sake of pedagogy, let $u \in U$ and $r > 0$ such that $B(u, r) \subset U$ and ϕ is weakly isotonic on $B(u, r)$. Take $x, y, z \in B(u, r/2)$ be such that $\|x - y\| = \|x - z\|$. For $t \in \mathbb{R}$, define $z_t = (1 - t)x + tz$. Let $t > 1$ such that $z_t \in B(u, r)$. Because $\|x - y\| < t\|x - z\| = \|x - z_t\|$, we have $\|\phi(x) - \phi(y)\| \leq \|\phi(x) - \phi(z_t)\|$. Letting $t \searrow 1$, we get $\|\phi(x) - \phi(y)\| \leq \|\phi(x) - \phi(z)\|$ by continuity of ϕ. Since y and z play the same role, the converse inequality is also true, and combined, yield an equality.

Midpoint preservation. Let $V \subset \mathbb{R}^d$ be convex. We say that a function $f : V \mapsto \mathbb{R}^d$ preserves midpoints if

$$f\left(\frac{x + y}{2}\right) = \frac{f(x) + f(y)}{2}, \qquad \forall x, y \in V.$$

We now show that ϕ preserves midpoints, locally. Kleindessner and Von Luxburg [14] also do that, however, our arguments are closer to those of [21], who make use of regular simplexes. The important fact is that a function that preserves isosceles preserves regular simplexes. Let $u \in U$ and $r > 0$ such that $B(u, r) \subset U$ and ϕ preserves isosceles on $B(u, r)$. Take $x, y \in B(u, r/2)$, and let $\mu = (x + y)/2$. Let $z_1, \ldots, z_d$ form a regular simplex with barycenter μ and side length s, and such that $\|x - z_i\| = s$ for all i. In other words, $x, z_1, \ldots, z_d$ forms a regular simplex placed so that μ is the barycenter of $z_1, \ldots, z_d$. By symmetry, $y, z_1, \ldots, z_d$ forms a regular simplex also. By Lemma 6, we have $\|z_i - \mu\|/\|x - \mu\| = \sqrt{(d - 1)/(d + 1)}$, so that $z_1, \ldots, z_d \in B(\mu, r/2) \subset B(u, r)$, by the triangle inequality and the fact that $\|x - \mu\| < r/2$. Hence, $\phi(x), \phi(z_1), \ldots, \phi(z_d)$ and $\phi(y), \phi(z_1), \ldots, \phi(z_d)$ are regular simplexes. If one of them is singular, so is the other one, in which case $\phi(x) = \phi(y) = \phi(\mu)$. Otherwise, necessarily $\phi(x)$ is the symmetric of $\phi(y)$ with respect to $\text{Aff}(\phi(z_1), \ldots, \phi(z_d))$; the only other possibility would be that $\phi(x) = \phi(y)$, but in that case we would still have that $\phi(z_i) = \phi(x)$ for all $i \in [d]$, since $\|x - z_i\|/\|x - \mu\| = \sqrt{2d/(d + 1)}$ by Lemma 6 – implying that $\|x - z_i\| < \|x - y\|$ – and ϕ is weakly isotonic in that neighborhood. So assume that $\phi(x)$ is the symmetric of $\phi(y)$ with respect to $\text{Aff}(\phi(z_1), \ldots, \phi(z_d))$. For $a \in \{x, y, \mu\}$, $\|a - z_i\|$ is constant in i, and therefore so is $\|\phi(a) - \phi(z_i)\|$, so that $\phi(a)$ belongs to the line of points equidistant to $\phi(z_1), \ldots, \phi(z_d)$. This implies that x, y, μ are collinear. And because $\|\mu - x\| = \|\mu - y\|$, we also have $\|\phi(\mu) - \phi(x)\| = \|\phi(\mu) - \phi(y)\|$, so that $\phi(\mu)$ is necessarily the midpoint of $\phi(x)$ and $\phi(y)$.

Conclusion. We arrived at the conclusion that ϕ can be extended to a continuous function on U that preserves midpoints locally. We then use the following simple results in sequence: with Lemma 7, we conclude that ϕ is locally affine; with Lemma 8, we conclude that ϕ is in fact affine on U; and with Lemma 9, we conclude that ϕ is in fact a similarity on U.

Lemma 7. *Let V be a convex set of a Euclidean space and let f be a continuous function on V with values in a Euclidean space that preserves midpoints. Then f is an affine transformation.*

Proof. This result is in fact well known, and we only provide a proof for completeness. It suffices to prove that f is such that $f((1-t)x+ty) = (1-t)f(x)+tf(y)$ for all $x, y \in V$ and all $t \in [0,1]$. Starting with the fact that this is true when $t = 1/2$, by recursion we have that this is true when t is dyadic, meaning, of the form $t = k2^{-j}$, where $j \geq 1$ and $k \leq 2^j$ are both integers. Since dyadic numbers are dense in $[0,1]$, by continuity of f, we deduce the desired property. $\qquad\square$

Lemma 8. *A locally affine function over an open and connected subset of a Euclidean space is the restriction of an affine function over the whole space.*

Proof. Let U be the domain and f the function. Cover U with a countable number of open balls $B_i, i \in I$ such that f coincides with an affine function f_i on B_i. Take $i, j \in I$ distinct. Since U is connected, there must be a sequence $i = k_1, \ldots, k_m = j$, all in I, such that $B_{k_s} \cap B_{k_{s+1}} \neq \varnothing$ for $s \in [m-1]$. Since $B_{k_s} \cap B_{k_{s+1}}$ is an open set, we must have $f_{k_s} = f_{k_{s+1}}$, and this being true for all s, it implies that $f_i = f_j$. $\qquad\square$

Lemma 9. *An affine function that preserves isosceles locally is a similarity transformation.*

Proof. Let f be an affine function that preserves isosceles in an open ball. Without loss of generality, we may assume that the ball is $B(0,2)$ and that $f(0) = 0$ (so that f is linear). Fix $u_0 \in \partial B(0,1)$ and let $a = \|f(u_0)\|$. Take $x \in \mathbb{R}^d$ different from 0 and let $u = x/\|x\|$. We have $\|f(x)\|/\|x\| = \|f(u)\| = \|f(u) - f(0)\| = \|f(u_0) - f(0)\| = \|f(u_0)\| = a$. Hence, $\|f(x)\| = a\|x\|$, valid for all $x \in \mathbb{R}^d$, and f being linear, this implies that f is a similarity. $\qquad\square$

3.2. Auxiliary results

We list here a number of auxiliary results that will be used in the proof of Theorem 3.

The following result is a perturbation bound for trilateration, which is the process of locating a point based on its distance to landmark points. For a real matrix Z, let $\sigma_k(Z)$ denote its kth largest singular value.

Lemma 10. *Let $z_1, \ldots, z_{d+1} \in \mathbb{R}^d$ such that $\mathrm{Aff}(z_1, \ldots, z_{d+1}) = \mathbb{R}^d$ and let Z denote the matrix with columns $z_1, \ldots, z_{d+1}$. Consider $p, q \in \mathbb{R}^d$ and define $a_i = \|p - z_i\|$ and $b_i = \|q - z_i\|$ for $i \in [d+1]$. Then*

$$\|p - q\| \leq \frac{1}{2}\sqrt{d}\sigma_d(Z)^{-1} \max_i |a_{d+1}^2 - a_i^2 - b_i^2 + b_{d+1}^2| \leq \sqrt{d}\sigma_d(Z)^{-1} \max_i |a_i^2 - b_i^2|.$$

Proof. Assume without loss of generality that $z_{d+1} = 0$. In that case, note that $a_{d+1} = \|p\|$ and $b_{d+1} = \|q\|$. Also, redefine Z as the matrix with columns $z_1, \ldots, z_d$, and note that the first d singular values remain unchanged. Since $\mathrm{Aff}(z_1, \ldots, z_{d+1}) = \mathbb{R}^d$, there is $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$ and $\beta = (\beta_1, \ldots, \beta_d) \in \mathbb{R}^d$ such that $p = \sum_{i \in [d]} \alpha_i z_i = Z\alpha$ and $q = \sum_{i \in [d]} \beta_i z_i = Z\beta$. For p, we have $a_i^2 = \|p - z_i\|^2 = \|p\|^2 + \|z_i\|^2 - 2z_i^\top Z\alpha$ for all $i \in [d]$, or in matrix form, $Z^\top Z\alpha = \frac{1}{2}u$,

where $u = (u_1, \ldots, u_d)$ and $u_i = a_{d+1}^2 - a_i^2 + \|z_i\|^2$. Similarly, we find $Z^\top Z\beta = \frac{1}{2}v$, where $v = (v_1, \ldots, v_d)$ and $v_i = b_{d+1}^2 - b_i^2 + \|z_i\|^2$. Hence, we have

$$\|Z^\top(p-q)\| = \|Z^\top Z\alpha - Z^\top Z\beta\| = \frac{1}{2}\|u - v\| = \frac{1}{2}\sqrt{\sum_{i \in [d]} \left(a_{d+1}^2 - b_{d+1}^2 - a_i^2 + b_i^2\right)^2}$$

$$\leq \frac{1}{2}\sqrt{d}\max_i |a_{d+1}^2 - a_i^2 - b_i^2 + b_{d+1}^2| \leq \sqrt{d}\max_i |a_i^2 - b_i^2|.$$

Simultaneously, $\|Z^\top(p-q)\| \geq \sigma_d(Z)\|p - q\|$. Combining both inequalities, we conclude. $\quad\square$

For $\eta \in [0, 1)$, we say that $z_1, \ldots, z_m \in \mathbb{R}^d$ form an η-approximate regular simplex if

$$\min_{i \neq j} \|z_i - z_j\| \geq (1 - \eta)\max_{i \neq j} \|z_i - z_j\|.$$

Lemma 11. *Let $z_1, \ldots, z_m$ form an η-approximate regular simplex with maximum edge length λ achieved by $\|z_1 - z_2\|$. There is a constant C_m and $z_1', \ldots, z_m' \in \mathrm{Aff}(z_1, \ldots, z_m)$ with $z_1' = z_1$ and $z_2' = z_2$ and forming a regular simplex with edge length λ, such that $\max_i \|z_i' - z_i\| \leq \lambda C_m \eta$.*

Proof. By scale equivariance, we may assume that $\lambda = 1$. We use an induction on m. In what follows, C_m, C_m', C_m'', etc., are constants that depend only on m. For $m = 2$, the statement is trivially true. Suppose that it is true for $m \geq 2$ and consider an η-approximate regular simplex $z_1, \ldots, z_{m+1} \in \mathbb{R}^d$ with maximum edge length 1. By changing d to m if needed, without loss of generality, assume that $\mathrm{Aff}(z_1, \ldots, z_{m+1}) = \mathbb{R}^d$. In that case, $z_1, \ldots, z_m$ is an η-approximate regular simplex with maximum edge length achieved by $\|z_1 - z_2\| = 1$, and by the inductive hypothesis, this implies the existence of $z_1', \ldots, z_m' \in A := \mathrm{Aff}(z_1, \ldots, z_m)$ with $z_1' = z_1$ and $z_2' = z_2$ and forming a regular simplex of edge length 1, such that $\max_{i \in [m]} \|z_i' - z_i\| \leq C_m \eta$ for some constant C_m. Let p be the orthogonal projection of z_{m+1} onto A. Before continuing, let $\mathcal{P}$ be the set of such p obtained when fixing $z_1', \ldots, z_m'$ and then varying $z_i \in B(z_i', C_m \eta)$ for $i \in [m]$ and z_{m+1} among the points that make an η-approximate regular simplex with $z_1, \ldots, z_m$. Let μ' be the barycenter of $z_1', \ldots, z_m'$ and note that $\mu' \in \mathcal{P}$. Now, set $\delta = \|z_{m+1} - p\|$. By the Pythagoras theorem, we have $\|p - z_i\|^2 = \|z_{m+1} - z_i\|^2 - \delta^2$, with $1 - \eta \leq \|z_{m+1} - z_i\| \leq 1$, so that $0 \leq 1 - \delta^2 - \|p - z_i\|^2 \leq 2\eta$. By the triangle inequality, $|\|p - z_i'\| - \|p - z_i\|| \leq \|z_i - z_i'\| \leq C_m \eta$, so that

$$\left|\|p - z_i'\|^2 - \|p - z_i\|^2\right| = \left|\|p - z_i'\| - \|p - z_i\|\right|\left(\|p - z_i'\| - \|p - z_i\|\right) \leq \|z_i - z_i'\|\|z_i - z_i'\|$$
$$\leq C_m \eta(2 + C_m \eta) \leq C_m' \eta,$$

using the fact that $\|p - z_i\| \leq \|z_{m+1} - z_i\| \leq 1$. Hence,

$$\mathcal{P} \subset \left\{q : \|q - z_i'\|^2 = 1 - \delta^2 \pm C_m'' \eta, \forall i \in [m]\right\}.$$

Since $\mu' \in \mathcal{P}$, we must therefore have $\|p - z_i'\|^2 = \|\mu' - z_i'\|^2 \pm 2C_m'' \eta$. By Lemma 10, this implies that $\|p - \mu'\| \leq \sqrt{m-1}\sigma_{m-1}^{-1}([z_1' \cdots z_m'])2C_m'' \eta =: C_m''' \eta$. Let z_{m+1}' be on the same side of

A as z_{m+1} and such that $z'_1, \ldots, z'_m, z'_{m+1}$ form a regular simplex. Note that μ' is the orthogonal projection of z'_{m+1} onto A. By the Pythagoras theorem, applied multiple times, we obtain the following. First, we have

$$
\begin{aligned}
\left\| z'_{m+1} - z_{m+1} \right\|^2 &= \left\| z'_{m+1} - \mu' + \mu' - p + p - z_{m+1} \right\|^2 \\
&= \left\| z'_{m+1} - \mu' \right\|^2 - 2(z'_{m+1} - \mu')^\top (z_{m+1} - p) + \| p - z_{m+1} \|^2 + \left\| \mu' - p \right\|^2 \\
&= \left(\left\| z'_{m+1} - \mu' \right\| - \| p - z_{m+1} \| \right)^2 + \left\| \mu' - p \right\|^2,
\end{aligned}
$$

because $z'_{m+1} - \mu'$ and $z_{m+1} - p$ are orthogonal to A, and therefore parallel to each other and both orthogonal to $\mu' - p$. For the second term, we already know that $\| \mu' - p \| \le C'''_m \eta$, while the first term is bounded by $(2C''_m + 2)^2 \eta^2$ since, on the one hand,

$$
\left\| z'_{m+1} - \mu' \right\|^2 = \left\| z'_{m+1} - z'_1 \right\|^2 - \left\| \mu' - z'_1 \right\|^2 = 1 - \left\| \mu' - z'_1 \right\|^2
$$

while, on the other hand,

$$
\| p - z_{m+1} \|^2 = \| z_{m+1} - z_1 \|^2 - \| p - z_1 \|^2 = 1 \pm 2\eta - \| p - z_1 \|^2,
$$

and we know that $\| \mu' - z'_1 \|^2 = \| p - z_1 \|^2 \pm 2C''_m \eta$. Hence, we find that $\| z'_{m+1} - z_{m+1} \|^2 \le C^2_{m+1} \eta^2$ for some constant C_{m+1} function of m only. This shows that the induction hypothesis holds for $m + 1$. $\qquad \square$

Lemma 12. *There are constants $C_m, C'_m > 0$ such that, if $z_1, \ldots, z_m$ form an η-approximate regular simplex with maximum edge length λ, then $\sigma_{m-1}([z_1 \cdots z_m]) \ge \lambda C_m (1 - C'_m \eta)$.*

Proof. By scale equivariance, we may assume that $\lambda = 1$. By Lemma 11, there is a constant C''_m and $z'_1, \ldots, z'_m \in \mathrm{Aff}(z_1, \ldots, z_m)$ forming a regular simplex with edge length 1 such that $\max_i \| z'_i - z_i \| \le C''_m \eta$. By Weyl's inequality ([11], Corollary 7.3.8), $\sigma_{m-1}(Z) \ge \sigma_{m-1}(Z') - \| Z - Z' \|$. On the one hand, $\sigma_{m-1}(Z')$ is a positive constant depending only on m, while on the other hand, $\| Z - Z' \| \le \| Z - Z' \|_F = \sqrt{\sum_i \| z_i - z'_i \|^2} \le \sqrt{m} C''_m \eta$. $\qquad \square$

Lemma 13. *Let $z_1, \ldots, z_m$ form an η-approximate regular simplex with maximum edge length λ and barycenter μ. Let $p \in \mathrm{Aff}(z_1, \ldots, z_m)$ and define $\gamma = \max_i \| p - z_i \|^2 - \min_i \| p - z_i \|^2$. There is a constant $C_m \ge 1$ depending only on m such that $\| p - \mu \| \le C_m \lambda \gamma$ when $\eta \le 1/C_m$.*

Proof. By scale equivariance, we may assume that $\lambda = 1$. By Lemma 10, we have

$$
\| p - \mu \| \le \frac{1}{2} \sqrt{m - 1} \, \sigma^{-1}_{m-1}([z_1 \cdots z_m]) \max_i \left| \| p - z_m \|^2 - \| p - z_i \|^2 \right|.
$$

By Lemma 12, there is a constant C'_m such that $\sigma^{-1}_{m-1}([z_1 \cdots z_m]) \le C'_m$ when $\eta \le 1/C'_m$. And we also have $\max_i \left| \| p - z_m \|^2 - \| p - z_i \|^2 \right| \le \gamma$. From this, we conclude. $\qquad \square$

Lemma 14. *Let* $\psi : \Lambda \mapsto Q$ *be isotonic, where* $\Lambda, Q \subset \mathbb{R}^d$. *Let* $v \in \mathbb{R}^d$ *and* $r > 0$, *and set* $\varepsilon = \delta_H(\Lambda, B(v, r))$. *There is* $C \propto \operatorname{diam}(Q)/r$ *such that, for all* $x, x', x^\dagger, x^\ddagger \in \Lambda$ *with* $x, x' \subset B(v, 3r/4)$ *and for all* $\eta \in (0, r/4 - 2\varepsilon)$,

$$\|x - x'\| = \|x^\dagger - x^\ddagger\| \pm \eta \quad \Rightarrow \quad \|\psi(x) - \psi(x')\| = \|\psi(x^\dagger) - \psi(x^\ddagger)\| \pm C(\eta + \varepsilon). \quad (12)$$

Proof. Let $\xi = \|x - x'\|$ and $\xi^\dagger = \|x^\dagger - x^\ddagger\|$. Suppose that $\xi < \eta + 2\varepsilon$, which implies that $\xi^\dagger < 2\eta + 2\varepsilon$. In that case, Lemma 3 – where the constant there is denoted here by $C_1 \propto \operatorname{diam}(Q)/r$ – yields $\|\psi(x) - \psi(x')\| \leq C_1(\xi + \varepsilon) \leq C_1(\eta + 3\varepsilon)$ and, similarly, $\|\psi(x^\dagger) - \psi(x^\ddagger)\| \leq C_1(2\eta + 3\varepsilon)$. This proves (12). Henceforth, we assume that $\xi \geq \eta + 2\varepsilon$.

First assume that $\xi > \xi^\dagger$. In that case, we immediately have $\|\psi(x) - \psi(x')\| \geq \|\psi(x^\dagger) - \psi(x^\ddagger)\|$. For the reverse, let $y_t = (1 - t)x + tx'$, and note that $\|y_t - x\| = t\xi$. Take $t = 1 - (\eta + 2\varepsilon)/\xi$ and note that $t \in [0, 1]$, so that $y_t \in [xx'] \subset B(v, r)$, and therefore there is $x^\star \in \Lambda$ such that $\|x^\star - y_t\| \leq \varepsilon$. We have $\|x^\star - x\| \leq \|y_t - x\| + \|x^\star - y_t\| \leq \xi - \eta - \varepsilon < \xi^\dagger$, so that $\|\psi(x) - \psi(x^\star)\| \leq \|\psi(x^\dagger) - \psi(x^\ddagger)\|$. Applying the triangle inequality and Lemma 3, we then have

$$\|\psi(x) - \psi(x^\star)\| \geq \|\psi(x) - \psi(x')\| - \|\psi(x') - \psi(x^\star)\|$$
$$\geq \|\psi(x) - \psi(x')\| - C_1(\|x' - x^\star\| + \varepsilon),$$

with $\|x' - x^\star\| \leq \|x' - y_t\| + \|y_t - x^\star\| \leq \eta + 3\varepsilon$.

When $\xi < \xi^\dagger$, we choose $t = 1 + (\eta + 2\varepsilon)/\xi$. Because $x, x' \subset B(v, 3r/4)$, we still have $y_t \in B(v, r)$ because of the constraint on η. The remaining arguments are analogous.

When $\xi = \xi^\dagger$, repeating what we just did both ways and with $\eta = 0$ yields the result. $\qquad\square$

Lemma 15. *Consider* $\psi : \Lambda \mapsto \mathbb{R}^d$ *isotonic, where* $\Lambda \subset \mathbb{R}^d$. *Let* V *denote the convex hull of* Λ. *Set* $\varepsilon = \delta_H(\Lambda, V)$ *and* $c = \operatorname{diam}(\psi(\Lambda))/5\operatorname{diam}(\Lambda)$. *Then* $\|\psi(x) - \psi(x')\| \geq c\|x - x'\|$ *for all* $x, x' \in \Lambda$ *such that* $\|x - x'\| \geq 4\varepsilon$.

Proof. We first prove that, if $c > 0$ and $\eta \geq 4\varepsilon$ are such that $\|\psi(x) - \psi(x')\| \leq c\eta$ for all $x, x' \in \Lambda$ with $\|x - x'\| < \eta$, then $\operatorname{diam}(\psi(\Lambda)) < c(4\operatorname{diam}(\Lambda) + \eta)$. Indeed, take $x, x' \in \Lambda$. Let $u = (x' - x)/\|x' - x\|$ and $L = \|x - x'\|$, and define $y_j = x + s_j u$ where $s_j = j(\eta - 3\varepsilon)$ for $j = 0, \ldots, J := \lfloor L/(\eta - 3\varepsilon) \rfloor$, and then let $s_{J+1} = L$. By construction, $y_j \in [xx'] \subset V$, with $y_0 = x$ and $y_{J+1} = x'$. Let $x_j \in \Lambda$ be such that $\|x_j - y_j\| \leq \varepsilon$, with $x_0 = x$ and $x_{J+1} = x'$. By the triangle inequality, $\|x_{j+1} - x_j\| \leq \|y_{j+1} - y_j\| + 2\varepsilon = s_{j+1} - s_j + 2\varepsilon < \eta$. Hence,

$$\|\psi(x) - \psi(x')\| \leq \sum_{j=0}^{J} \|\psi(x_j) - \psi(x_{j+1})\| \leq (J+1)c\eta \leq c\frac{L\eta}{\eta - 3\varepsilon} + c\eta < c(4\operatorname{diam}(\Lambda) + \eta),$$

since $\eta - 3\varepsilon \geq \eta - 3\eta/4 = \eta/4$ and $L \leq \operatorname{diam}(\Lambda)$.

Now assume that ψ is isotonic and suppose that $\|\psi(x) - \psi(x')\| < c\|x - x'\|$ for some $x, x' \in \Lambda$ such that $\eta := \|x - x'\| \geq 4\varepsilon$. Then we have $\|\psi(x^\dagger) - \psi(x^\ddagger)\| \leq c\eta$ when $x^\dagger, x^\ddagger \in \Lambda$ satisfy $\|x^\dagger - x^\ddagger\| < \eta$. We just showed that this implies that $\operatorname{diam}(\psi(\Lambda)) < c(4\operatorname{diam}(V) + \eta)$, and we conclude using the fact that $\eta \leq \operatorname{diam}(\Lambda)$. $\qquad\square$

The following result is on 1-nearest neighbor interpolation.

Lemma 16. *Let Λ be a subset of isolated points in $V \subset \mathbb{R}^d$ and set $\varepsilon = \delta_H(\Lambda, V)$. For any function $\psi : \Lambda \mapsto \mathbb{R}^d$, define its 1-nearest neighbor interpolation as $\hat{\psi} : V \mapsto \mathbb{R}^d$ as*

$$\hat{\psi}(y) = \frac{1}{|N_\Lambda(y)|} \sum_{x \in N_\Lambda(y)} \psi(x), \qquad N_\Lambda(y) := \arg\min_{x \in \Lambda} \|x - y\|. \tag{13}$$

Consider the modulus of continuity of ψ, which for $\eta > 0$ is defined as $\omega(\eta) = \sup\{\|\psi(x) - \psi(x')\| : x, x' \in \Lambda, \|x - x'\| \le \eta\}$. Then the modulus of continuity of $\hat{\psi}$, denoted $\hat{\omega}$, satisfies $\hat{\omega}(\eta) \le \omega(\eta + 2\varepsilon)$. Moreover, for any $y, y' \in V$ and any $x, x' \in \Lambda$ such that $\|x - y\| \le \varepsilon$ and $\|x' - y'\| \le \varepsilon$,

$$\|\hat{\psi}(y) - \hat{\psi}(y')\| = \|\psi(x) - \psi(x')\| \pm 2\omega(2\varepsilon).$$

Proof. Fix $\eta > 0$ and take $y, y' \in V$ such that $\|y - y'\| \le \eta$. We have $\|x - y\| \le \varepsilon$ for all $x \in N_\Lambda(y)$ and $\|x' - y'\| \le \varepsilon$ for all $x' \in N_\Lambda(y')$, so that $\|x - x'\| \le \|y - y'\| + 2\varepsilon$ for all such x and x', by the triangle inequality. Therefore,

$$\|\hat{\psi}(y) - \hat{\psi}(y')\| \le \sup\{\|\psi(x) - \psi(x')\| : x \in N_\Lambda(y), x' \in N_\Lambda(y')\}$$
$$\le \sup\{\|\psi(x) - \psi(x')\| : x, x' \in \Lambda, \|x - x'\| \le \eta + 2\varepsilon\} = \omega(\eta + 2\varepsilon).$$

Since this is true for all $y, y' \in V$ such that $\|y - y'\| \le \eta$, we conclude that $\hat{\omega}(\eta) \le \omega(\eta + 2\varepsilon)$.

For the second part of the lemma, we have

$$\|\hat{\psi}(y) - \hat{\psi}(y')\| = \|\psi(x) - \psi(x')\| \pm \|\hat{\psi}(y) - \psi(x)\| \pm \|\hat{\psi}(y') - \psi(x')\|,$$

where the second term is bounded by

$$\|\hat{\psi}(y) - \psi(x)\| \le \sup\{\|\psi(\tilde{x}) - \psi(x)\| : \tilde{x} \in N_\Lambda(y)\}$$
$$\le \sup\{\|\psi(\tilde{x}) - \psi(x)\| : \|\tilde{x} - x\| \le 2\varepsilon\} \le \omega(2\varepsilon),$$

using the fact that $\|\tilde{x} - x\| \le \|\tilde{x} - y\| + \|y - x\| \le 2\varepsilon$, and similarly for the third term. $\square$

Let $V \subset \mathbb{R}^d$ be convex. In our context, we say that $f : V \mapsto \mathbb{R}^d$ is *η-approximately midlinear* if

$$\left\| f\left(\frac{x + y}{2}\right) - \frac{1}{2}(f(x) + f(y)) \right\| \le \eta, \qquad \forall x, y \in V.$$

Lemma 17. *Let $V \subset \mathbb{R}^d$ be star-shaped with respect to some point in its interior. There is a constant C depending only on V such that, for any η-approximately midlinear function $f : V \mapsto \mathbb{R}^d$, there is a affine function $T : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that $\sup_{x \in V} \|f(x) - T(x)\| \le C\eta$.*

Note that, if V is a ball, then by invariance considerations, C only depends on d.

Proof of Lemma 17. This is a direct consequence of [24], Theorem 1.4. □

We say that $f : V \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ is an ε-*isometry* if

$$\|x - y\| - \varepsilon \leq \|f(x) - f(y)\| \leq \|x - y\| + \varepsilon, \qquad \forall x, y \in V.$$

For a set $V \subset \mathbb{R}^d$, define its thickness as

$$\theta(V) = \inf\{\operatorname{diam}(u^\top V) : u \in \mathbb{R}^d, \|u\| = 1\}.$$

Recalling the definition of ρ in (7), we note that $\theta(V) \geq \rho(V)$, but that the two are distinct in general.

Lemma 18. *Let $V \subset \mathbb{R}^d$ be compact and such that $\theta(V) \geq \eta \operatorname{diam}(V)$ for some $\eta > 0$. There is a constant C depending only on d such that, if $f : V \mapsto \mathbb{R}^d$ is an ε-isometry, then there is an isometry $R : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that $\max_{x \in V} \|f(x) - R(x)\| \leq C\varepsilon/\eta$.*

Proof. This is a direct consequence of [2], Theorem 3.3. □

Lemma 19. *Let $T : \mathbb{R}^d \mapsto \mathbb{R}^d$ be an affine function that transforms a regular simplex of edge length 1 into an η-approximate regular simplex of maximum edge length $\lambda > 0$. There is a constant C, depending only on d, and an isometry R, such that $\|T(x) - \lambda R(x)\| \leq C\lambda\eta$ for all $x \in B(0, 1)$.*

Proof. By invariance, we may assume T is linear and that the regular simplex is formed by $0, z_1, \ldots, z_d$ and has edge length 1. Letting $w_i = T(z_i)$, we have that $0, w_1, \ldots, w_d$ form an η-approximate regular simplex of maximum edge length $\lambda := \max_i \|w_i\|$. Lemma 11 gives 0, $w_1', \ldots, w_d'$ forming a regular simplex of edge length λ such that $\max_i \|w_i - w_i'\| \leq C_1\lambda\eta$ for some constant C_1. Let R be the orthogonal transformation such that $R(z_i) = w_i'/\lambda$ for all $i \in [d]$. We have $\|T(z_i) - \lambda R(z_i)\| = \|w_i - w_i'\| \leq C_1\lambda\eta$ for all i. In matrix notation, letting $Z := [z_1 \cdots z_d]$, we have

$$\|TZ - \lambda RZ\| \leq \|TZ - \lambda RZ\|_F = \sqrt{\sum_{i=1}^d \|Tz_i - \lambda Rz_i\|^2}$$

$$\leq \sqrt{d} \max_{i \in [d]} \|Tz_i - \lambda Rz_i\| \leq \sqrt{d} C_1\lambda\eta.$$

At the same time, $\|TZ - \lambda RZ\| \geq \|T - \lambda R\|/\|Z^{-1}\|$ with $\|Z^{-1}\| = 1/\sigma_d(Z) = 1/\sigma_d([0z_1 \cdots z_d])$ being a positive constant depending only on d. Hence, $\|T - \lambda R\| \leq (\sqrt{d}/\sigma_d(Z))C_1\lambda\eta =: C_2\lambda\eta$. □

Lemma 20. *Suppose that $S_1, S_2 : \mathbb{R}^d \mapsto \mathbb{R}^d$ are two affinities such that $\max_{x \in B(y,r)} \|S_1(x) - S_2(x)\| \leq \eta$ for some $y \in \mathbb{R}^d$ and $r > 0$. Then $\|S_1(x) - S_2(x)\| \leq 2\eta \|x - y\|/r + \eta$ for all $x \in \mathbb{R}^d$.*

Proof. By translation and scale invariance, assume that $y = 0$ and $r = 1$. Let $L_i = S_i - S_i(0)$. For $x \in B(0,1)$, we $\|L_1(x) - L_2(x)\| \leq \|S_1(x) - S_2(x)\| + \|S_1(0) - S_2(0)\| \leq 2\eta$. Hence, for $x \in \mathbb{R}^d$, $\|L_1(x) - L_2(x)\| \leq 2\eta \|x\|$, which in turn implies that $\|S_1(x) - S_2(x)\| \leq \|L_1(x) - L_2(x)\| + \|S_1(0) - S_2(0)\| \leq 2\eta \|x\| + \eta$. $\qquad \square$

3.3. Proof of Theorem 3

Without loss of generality, we may assume that $D_n := \operatorname{diam}(\phi_n(\Omega_n)) \geq 1$. Indeed, suppose that $D_n < 1$, but different from 0, for otherwise ϕ_n is a degenerate similarity and the result follows. Let $\tilde{\phi}_n = D_n^{-1}\phi_n$, which is isotonic on Ω_n and satisfies $\operatorname{diam}(\tilde{\phi}_n(\Omega_n)) = 1$. If the result is true for $\tilde{\phi}_n$, there is a similarity $\tilde{S}_n$ such that $\max_{x \in \Omega_n} |\tilde{\phi}_n(x) - \tilde{S}_n(x)| \leq C\varepsilon_n$ for some constant C. (We implicitly assume that the set $\phi_n(\Omega_n)$ contains the origin, so that $\tilde{\phi}_n(\Omega_n)$ remains bounded.) We then have $\max_{x \in \Omega_n} |\phi_n(x) - S_n(x)| \leq CD_n\varepsilon_n \leq C\varepsilon_n$, where $S_n := D_n\tilde{S}_n$ is also a similarity.

Let $r = \rho(U)$, so that there is some $u_\star$ such that $B(u_\star, r) \subset U$. Let $\Lambda_n = \Omega_n \cap B(u_\star, r/2)$ and $\delta_n = \operatorname{diam}(\phi_n(\Lambda_n))$. Let w be any unit-norm vector and define $y_\pm = u_\star \pm (r/2 - \varepsilon_n)w$. Let $x_\pm \in \Omega_n$ be such that $\|x_\pm - y_\pm\| \leq \varepsilon_n$. Necessarily, $x_\pm \in \Lambda_n$ because the distance from $y_\pm$ to $\partial B(u_\star, r/2)$ exceeds ε_n. Note that $\|x_- - x_+\| \geq r_1 := r - 4\varepsilon_n$. By isotonicity,

$$\left\|\phi_n(x) - \phi_n(x')\right\| \leq \left\|\phi_n(x_-) - \phi_n(x_+)\right\| \leq \delta_n, \qquad \text{whenever } \left\|x - x'\right\| < r_1. \quad (14)$$

Let $y_1, \ldots, y_K$ be a $(r_1/3)$-packing of U, so that $K \leq C(\operatorname{diam}(U)/r)^d$ for some constant $C > 0$. Let $x_{i_k} \in \Omega_n$ be such that $\|x_{i_k} - y_k\| \leq \varepsilon_n$, so that $U \subset \bigcup_{k \in [K]} B(y_k, r_1/3) \subset \bigcup_{k \in [K]} B(x_{i_k}, r_2)$, where $r_2 := r_1/3 + \varepsilon_n$. Let $z_k = x_{i_k}$ for clarity. Take $x, x' \in \Omega_n$. Because U is open, it is path-connected, so there is a continuous curve $\gamma : [0, 1] \mapsto U$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. Let $k_0 \in [K]$ be such that $x \in B(z_{k_0}, r_2)$ and $s_0 = 0$. Then for $j \geq 0$, let $s_{j+1} = \inf\{s > s_j : \gamma(s) \notin \bigcup_{l \in [j]} B(z_{k_l}, r_2)\}$, and let $k_{j+1} \in [K]$ be such that $\|z_{k_{j+1}} - \gamma(s_{j+1})\| \leq \varepsilon_n$. Let $J = \min\{j : s_{j+1} = \infty\}$, which is indeed finite. By construction, $\|z_{k_j} - z_{k_{j+1}}\| \leq 2r_2 < r_1$ when $\varepsilon_n < r/10$. By (14), we have $\|\phi_n(z_{k_j}) - \phi_n(z_{k_{j+1}})\| \leq \delta_n$. Thus, by the triangle inequality, $\|\phi_n(x) - \phi_n(x')\| \leq J\delta_n \leq K\delta_n$. This being true for all $x, x' \in \Omega_n$, this prove that $\delta_n \geq D_n/K \propto D_n(\operatorname{diam}(U)/r)^{-d}$.

1-NN interpolation. Let $\hat{\phi}_n$ denote the 1-NN interpolation of ϕ_n as in (13). We claim that there is a $C_0^\star \propto D_n/r$ and $c_0^\star \propto (\operatorname{diam}(U)/r)^{-d} D_n/r$ such that $\hat{\phi}_n$ satisfies the following properties: for all $y, y', y^\dagger, y^\ddagger \in U$,

$$\left\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\right\| \leq C_0^\star \left(\left\|y - y'\right\| + \varepsilon_n\right), \quad (15)$$

$$\left\|y - y'\right\| < \left\|y^\dagger - y^\ddagger\right\| - 4\varepsilon_n \quad \Rightarrow$$

$$\left\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\right\| \leq \left\|\hat{\phi}_n(y^\dagger) - \hat{\phi}_n(y^\ddagger)\right\| + C_0^\star \varepsilon_n, \quad (16)$$

and also

$$\big\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\big\| \geq c_0^\star \|y - y'\| - C_0^\star \varepsilon_n,$$
$$\text{if } y, y' \in B(u_\star, r/2) \text{ satisfy } \|y - y'\| \geq 10\varepsilon_n, \tag{17}$$

and

$$\|y - y'\| = \|y^\dagger - y^\ddagger\| \pm \eta \quad \Rightarrow \quad \big\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\big\| = \big\|\hat{\phi}_n(y^\dagger) - \hat{\phi}_n(y^\ddagger)\big\| \pm C_0^\star(\eta + \varepsilon_n),$$
$$\text{if } y, y' \in B(u_\star, r/2), \; \varepsilon_n < r/120 \text{ and } 0 \leq \eta \leq r/5. \tag{18}$$

Indeed, let $x, x', x^\dagger, x^\ddagger \in \Omega_n$ such that $\|x - y\|, \|x' - y'\|, \|x^\dagger - y^\dagger\|, \|x^\ddagger - y^\ddagger\| \leq \varepsilon_n$.

For (15), we start by applying Lemma 16 to get

$$\big\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\big\| = \big\|\phi_n(x) - \phi_n(x')\big\| \pm 2\omega_n(2\varepsilon_n)$$
$$\leq \omega_n\big(\|x - x'\|\big) + 2\omega_n(2\varepsilon_n) \leq \omega_n\big(\|y - y'\| + 2\varepsilon_n\big) + 2\omega_n(2\varepsilon_n),$$

where ω_n is the modulus of continuity of ϕ_n. We then use Lemma 3, which gives that $\omega_n(\eta) \leq C\eta$ for all η and some $C \propto D_n/r$, to get $\omega_n(\|y - y'\| + 2\varepsilon_n) \pm 2\omega_n(2\varepsilon_n) \leq C(\|y - y'\| + 6\varepsilon_n)$.

For (16), we first note that $\|x - x'\| < \|x^\dagger - x^\ddagger\|$ by the triangle inequality, which in turn implies that $\|\phi_n(x) - \phi_n(x')\| \leq \|\phi_n(x^\dagger) - \phi_n(x^\ddagger)\|$ since ϕ_n is isotonic. We then apply Lemma 16 to get that $\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\| \leq \|\hat{\phi}_n(y^\dagger) - \hat{\phi}_n(y^\ddagger)\| + 4\omega_n(2\varepsilon_n)$, and conclude with Lemma 3 as for (15).

For (17), we may apply Lemma 15 with Λ_n. Let V be the convex hull of Λ_n, so that $V \subset B(u_\star, r/2)$. Let z be a point in that ball. If $z \neq u_\star$, let $w = (u_\star - z)/\|u_\star - z\|$, and if $z = u_\star$, let w be any unit-norm vector. Define $z' = z + \varepsilon_n w$ and notice that the distance from z' to $\partial B(u_\star, r/2)$ exceeds ε_n. Therefore, if $x \in \Omega_n$ is such that $\|z' - x\| \leq \varepsilon_n$, then necessarily, $x \in \Lambda_n$. We then note that $\|z - x\| \leq 2\varepsilon_n$. We conclude that $\delta_H(\Lambda_n, V) \leq 2\varepsilon_n$. Since $\|x - x'\| \geq \|y - y'\| - 2\varepsilon_n \geq 4(2\varepsilon_n)$, we get that $\|\phi_n(x) - \phi_n(x')\| \geq c\|x - x'\|$, with $c := \mathrm{diam}(\phi_n(\Lambda_n))/5\,\mathrm{diam}(\Lambda_n) \geq \delta_n/5r$. We then apply Lemma 16 to obtain $\|\hat{\phi}_n(y) - \hat{\phi}_n(y')\| \geq c\|x - x'\| - 2\omega_n(2\varepsilon_n) \geq c\|y - y'\| - 2(c + C)\varepsilon_n$, using Lemma 3 as for (15).

For (18), note that $x, x' \in B(u_\star, r/2 + \varepsilon_n) \subset B(u_\star, 3r/4)$, and $\|x - x'\| = \|x^\dagger - x^\ddagger\| \pm (\eta + 4\varepsilon_n)$ by the triangle inequality. By Lemma 14 – where the constant there is denoted here by $C' \propto D_n/r$ – this implies that

$$\big\|\phi_n(x) - \phi_n(x')\big\| = \big\|\phi_n(x^\dagger) - \phi_n(x^\ddagger)\big\| \pm C'(\eta + \varepsilon_n)$$

when $\eta + 4\varepsilon_n < r/4 - 2\varepsilon_n$, which is true when $\varepsilon_n < r/120$ and $\eta \leq r/5$. We then apply Lemma 16 together with Lemma 3, as for (15).

CASE $d = 1$. This case is particularly simple. Note that U is a bounded open interval of $\mathbb{R}$. We show that the function $\hat{\phi}_n$ is approximately midlinear on U. Take $x, y \in U$ and define $\mu = (x + y)/2$. By the fact that $\hat{\phi}_n$ takes its values in $\mathbb{R}$, and (18), we have

$$\left| \frac{1}{2}\big(\hat{\phi}_n(x) + \hat{\phi}_n(y)\big) - \hat{\phi}_n(\mu) \right| = \frac{1}{2}\big| \big|\hat{\phi}_n(x) - \hat{\phi}_n(\mu)\big| - \big|\hat{\phi}_n(y) - \hat{\phi}_n(\mu)\big| \big| \leq C_0^\star \varepsilon_n/2,$$

when ε_n/r is small enough. Hence, $\hat{\phi}_n$ is $(C_0^\star \varepsilon_n)$-approximate midlinear on U. By the result of [24], namely Lemma 17, there is $C \propto 1$ – since U is a ball – and an affine function T_n such that $\max_{y \in U} |\hat{\phi}_n(y) - T_n(y)| \le C C_0^\star \varepsilon_n$. Since all affine transformations from $\mathbb{R}$ to $\mathbb{R}$ are (possibly degenerate) similarities, we conclude.

CASE $d \ge 2$. For the remaining of this subsection, we assume that $d \ge 2$.

Approximate midlinearity. We show that there is a constant C such that $\hat{\phi}_n$ is locally $C\varepsilon_n$-approximately midlinear. Take $x, y \in B(u_\star, r/4)$, and let $\mu = (x + y)/2$. Let $t > 0$ be a constant to be set large enough later.

If $\|x - y\| \le t\varepsilon_n$, then by (15), $\hat{\phi}_n(x), \hat{\phi}_n(y) \in B(\hat{\phi}_n(\mu), C_0^\star(t/2 + 1)\varepsilon_n)$, so that

$$\left\| \hat{\phi}_n(\mu) - \frac{1}{2}\big(\hat{\phi}_n(x) + \hat{\phi}_n(y)\big) \right\| \le C_0^\star(t/2 + 1)\varepsilon_n.$$

Therefore, assume that $\|x - y\| \ge t\varepsilon_n$. Let $z_1, \dots, z_d$ be constructed as in the proof of Theorem 1. By construction, both $x, z_1, \dots, z_d$ and $y, z_1, \dots, z_d$ form regular simplexes, and μ is the barycenter of $z_1, \dots, z_d$. By Lemma 6, for any $i \ne j$,

$$\|z_i - \mu\| = \sqrt{(d-1)/2d}\|z_i - z_j\| = \sqrt{(d-1)/2d}\sqrt{2d/(d+1)}\|x - \mu\| \le \|x - y\|/2,$$

which coupled with the fact that $x, y \in B(u_\star, r/4)$ yields that $z_i \in B(u_\star, r/2)$ for all i. Now, let $z_0 = x$. By (18), we have $\min_{i \ne j} \|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| \ge \max_{i,j} \|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| - C_0^\star \varepsilon_n$. Let $c_d = \sqrt{d/(2d+2)}$. By (17) and Lemma 6,

$$\big\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\big\| \ge c_0^\star \|z_i - z_j\| - C_0^\star \varepsilon_n = c_0^\star c_d \|x - y\| - C_0^\star \varepsilon_n \ge \big(c_0^\star c_d t - C_0^\star\big)\varepsilon_n.$$

Hence, assuming $t \ge 2C_0^\star/c_0^\star c_d$, we have $\min_{i \ne j} \|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| \ge (1 - \eta) \max_{i,j} \|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\|$, where $\eta := 2C_0^\star/(c_0^\star c_d t)$. In that case, $\hat{\phi}_n(x), \hat{\phi}_n(z_1), \dots, \hat{\phi}_n(z_d)$ form a η-approximate regular simplex. By symmetry, the same is true of $\hat{\phi}_n(y), \hat{\phi}_n(z_1), \dots, \hat{\phi}_n(z_d)$.

Define $\lambda = \|\hat{\phi}_n(x) - \hat{\phi}_n(y)\|$. By Lemma 6, $\|z_i - z_j\| = c_d\|x - y\| < \|x - y\| - 4\varepsilon_n$ when $t > 4/(1 - c_d)$, since $\|x - y\| \ge t\varepsilon_n$ and $c_d < 1$. By (16), this implies that $\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| \le \lambda + C_0^\star \varepsilon_n$. By (17), $\lambda \ge (c_0^\star t - C_0^\star)\varepsilon_n$, so that $\lambda + C_0^\star \varepsilon_n \le 2\lambda$ since we already assumed that $t \ge 2C_0^\star/c_0^\star c_d > 2C_0^\star/c_0^\star$.

For $a \in \{x, y, \mu\}$, $\|a - z_i\|$ is constant in $i \in [d]$. Therefore, by (18), $\min_i \|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\| \ge \max_i \|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\| - C_0^\star \varepsilon_n$. Define ξ_a as the orthogonal projection of $\hat{\phi}_n(a)$ onto the affine space $A := \text{Aff}(\hat{\phi}_n(z_1), \dots, \hat{\phi}_n(z_d))$ and let $\delta_a = \|\hat{\phi}_n(a) - \xi_a\|$. By the Pythagoras theorem, we have $\|\xi_a - \hat{\phi}_n(z_i)\|^2 = \|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\|^2 - \delta_a^2$. In particular,

$$\max_i \big\|\xi_a - \hat{\phi}_n(z_i)\big\|^2 - \min_i \big\|\xi_a - \hat{\phi}_n(z_i)\big\|^2 = \max_i \big\|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\big\|^2 - \min_i \big\|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\big\|^2$$

$$\le 2C_0^\star \varepsilon_n \min_i \big\|\hat{\phi}_n(a) - \hat{\phi}_n(z_i)\big\| + \big(C_0^\star \varepsilon_n\big)^2 \le C_1 \varepsilon_n,$$

where $C_1 := 2C_0^\star D_n + C_0^\star r$, once $\varepsilon_n \le r$. Let ζ denotes the barycenter of $\hat{\phi}_n(z_1), \dots, \hat{\phi}_n(z_d)$. Assume that t is sufficiently large that $\eta \le 1/C_2$, where $C_2 \propto 1$ is the constant of Lemma 13.

By that lemma, and the fact that $\hat{\phi}_n(z_1), \ldots, \hat{\phi}_n(z_d)$ form a η-approximate regular simplex of maximum edge length bounded by λ, we have $\|\xi_a - \zeta\| \leq C_2\lambda C_1\varepsilon_n$. Let L be the line passing through ζ and perpendicular to A. We just proved that $\hat{\phi}_n(x), \hat{\phi}_n(y), \hat{\phi}_n(\mu)$ are within distance $C_3\lambda\varepsilon_n$ from L, where $C_3 := C_1 C_2$.

Let ξ denote the orthogonal projection of $\hat{\phi}_n(\mu)$ onto $(\hat{\phi}_n(x)\hat{\phi}_n(y))$. Since $\|x - \mu\| = \|y - \mu\|$, we can apply (18) to get

$$\left|\left\|\xi - \hat{\phi}_n(x)\right\|^2 - \left\|\xi - \hat{\phi}_n(y)\right\|^2\right|$$
$$= \left|\left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(x)\right\|^2 - \left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(y)\right\|^2\right|$$
$$= \left|\left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(x)\right\| + \left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(y)\right\|\right| \times \left|\left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(x)\right\| - \left\|\hat{\phi}_n(\mu) - \hat{\phi}_n(y)\right\|\right|$$
$$\leq 4C_0^\star\lambda\varepsilon_n,$$

using the fact that $\max(\|\hat{\phi}_n(\mu) - \hat{\phi}_n(x)\|, \|\hat{\phi}_n(\mu) - \hat{\phi}_n(y)\|) \leq \lambda + C_0^\star\varepsilon_n \leq 2\lambda$, due to (16) and $\|x - \mu\| = \|y - \mu\| = \frac{1}{2}\|x - y\| < \|x - y\| - 4\varepsilon_n$ when t is large enough. By Lemma 13, we then obtain $\|\xi - \frac{1}{2}(\hat{\phi}_n(x) + \hat{\phi}_n(y))\| \leq C_4\lambda\varepsilon_n$ for some constant $C_4 \propto C_0^\star$. In particular, recalling that $\lambda = \|\hat{\phi}_n(x) - \hat{\phi}_n(y)\|$, this implies that $\xi \in [\hat{\phi}_n(x)\hat{\phi}_n(y)]$ when $\varepsilon_n \leq 1/2C_4$.

It remains to argue that $\hat{\phi}_n(\mu)$ is close to ξ. We already know that $\hat{\phi}_n(x), \hat{\phi}_n(y), \hat{\phi}_n(\mu)$ are within distance $C_3\lambda\varepsilon_n$ from L, and by convexity, the same must be true of ξ. Let $M = (\hat{\phi}_n(x)\hat{\phi}_n(y))$ and $\theta = \angle(L, M)$. Let P_M denote the orthogonal projection onto M, when M is a linear subspace. By Pythagoras theorem,

$$\lambda^2 = \left\|\hat{\phi}_n(x) - \hat{\phi}_n(y)\right\|^2 = \left\|P_L\left(\hat{\phi}_n(x) - \hat{\phi}_n(y)\right)\right\|^2 + \left\|P_{L^\perp}\left(\hat{\phi}_n(x) - \hat{\phi}_n(y)\right)\right\|^2$$
$$\leq (\cos\theta)^2\lambda^2 + (2C_3\lambda\varepsilon_n)^2,$$

implying that $\sin\theta \leq 2C_2\varepsilon_n$. Since $\|P_L - P_M\| = \sin\theta$ and $\hat{\phi}_n(\mu) - \xi$ is parallel to M, we also have

$$\left\|\hat{\phi}_n(\mu) - \xi\right\|^2 = \left\|P_L\left(\hat{\phi}_n(\mu) - \xi\right)\right\|^2 + \left\|P_{L^\perp}\left(\hat{\phi}_n(\mu) - \xi\right)\right\|^2$$
$$\leq (\sin\theta)^2\left\|\hat{\phi}_n(\mu) - \xi\right\|^2 + (2C_3\lambda\varepsilon_n)^2,$$

so that $\|\hat{\phi}_n(\mu) - \xi\| \leq 2C_3\lambda\varepsilon_n/\cos\theta \leq 2C_3\lambda\varepsilon_n/\sqrt{1 - (2C_3\varepsilon_n)^2} \leq C_5\lambda\varepsilon_n$, for some constant $C_5 \propto C_3$, once $C_3\varepsilon_n$ is small enough.

We conclude that $\|\hat{\phi}_n(\mu) - \frac{1}{2}(\hat{\phi}_n(x) + \hat{\phi}_n(y))\| \leq (C_4 + C_5)\lambda\varepsilon_n$, by the triangle inequality.

Approximate affinity. We now know that $\hat{\phi}_n$ is $C\varepsilon_n$-approximate midlinear on $B(u_\star, r/4)$ for some constant $C \propto C_0^\star(D_n + r) \propto C_0^\star(\text{diam}(Q) + r)$. This implies, by the result of [24], that is Lemma 17, that there is an affine function T_n such $\|\hat{\phi}_n(x) - T_n(x)\| \leq C_1^\star\varepsilon_n$ for all $x \in W$, for some constant $C_1^\star \propto rC \propto rC_0^\star(\text{diam}(Q) + r)$.

Approximate similarity. (Reinitialize the constants $C_k, k \geq 1$.) We saw above that $\hat{\phi}_n$ transforms the regular simplex $z_0(= x), z_1, \ldots, z_d$ with height denoted h satisfying $h \geq t\varepsilon_n/2$ into a η-

approximate one, where $\eta = 2C_0^\star/(c_0^\star c_d t)$. In what follows, choose these points so that they are all in $B(u_\star, r/2)$ and the simplex has height $h \geq r/8$. (From here on, reinitialize the variables x, y, λ, etc.) We can then take $t = r/4\varepsilon_n$, yielding $\eta = C_1\varepsilon_n$ for a constant $C_1 \propto C_0^\star/(c_0^\star r)$. By the triangle inequality, we have

$$\min_{i \neq j}\|T_n(z_i) - T_n(z_j)\| \geq \min_{i \neq j}\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| - 2C_1^\star\varepsilon_n$$

$$\geq (1 - C_1\varepsilon_n)\max_{i \neq j}\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| - 2C_1^\star\varepsilon_n$$

$$\geq \max_{i \neq j}\|T_n(z_i) - T_n(z_j)\| - \left(4C_1^\star + C_1\delta_n\right)\varepsilon_n.$$

By the triangle inequality and (17),

$$\gamma_n := \max_{i,j}\|T_n(z_i) - T_n(z_j)\| \geq \max_{i,j}\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| - 2C_1^\star\varepsilon_n$$

$$\geq c_0^\star\max_{i,j}\|z_i - z_j\| - C_0^\star\varepsilon_n - 2C_1^\star\varepsilon_n \geq c_0^\star r/8 - \left(C_0^\star + 2C_1^\star\right)\varepsilon_n.$$

Hence, we find that $T_n(z_0), \ldots, T_n(z_d)$ form a $C_2\varepsilon_n$-approximate regular simplex, where $C_2 := (4C_1^\star + C_1\delta_n)/(c_0^\star r/8 - (C_0^\star + 2C_1^\star)\varepsilon_n)$. Note that its maximum edge length is bounded as follows:

$$\gamma_n \leq \max_{i,j}\|\hat{\phi}_n(z_i) - \hat{\phi}_n(z_j)\| + 2C_1^\star\varepsilon_n \leq \delta_n + 2C_1^\star\varepsilon_n \leq 2\delta_n,$$

when $2C_1^\star\varepsilon_n \leq \delta_n$. By Lemma 19, there is a constant $C_3 > 0$ and an isometry $R_n^\star$, such that we have $\max_{x \in W}\|T_n(x) - \lambda_n R_n^\star(x)\| \leq C_3\lambda_n C_2\varepsilon_n$, where $\lambda_n := \gamma_n/h$. Because $r/8 \leq h \leq r$ and the bounds on γ_n above, there is a constant $C_2^\star \geq 1$ such that

$$1/C_2^\star \leq \lambda_n \leq C_2^\star. \tag{19}$$

This implies that

$$\left\|\hat{\phi}_n(x) - \lambda_n R_n^\star(x)\right\| \leq \left\|\hat{\phi}_n(x) - T_n(x)\right\| + \left\|T_n(x) - \lambda_n R_n^\star(x)\right\|$$

$$\leq \left(C_1^\star + C_3 C_2 C_2^\star\right)\varepsilon_n =: C_3^\star\varepsilon_n. \tag{20}$$

Covering and conclusion. (Reinitialize the constants $C_k, k \geq 1$.) Let $u_1 = u_\star$ and let $u_2, \ldots, u_K \in U$ be such that $u_1, \ldots, u_K$ form a maximal $(r/16)$-packing of U. (The number 16 is not essential here, but will play a role in the proof of Theorem 4.) Note that $U = U_1 \cup \cdots \cup U_K$ where $U_k := U \cap B(u_k, r/4)$, and note that $U_\star := U_1 \subset U$. For $u, u' \in U_k$, there are $w, w' \in U_\star$ such that $\|w - w'\| = \|u - u'\|$. Define $\tilde{\phi}_n = \hat{\phi}_n/\lambda_n$. By (18), and then (19)–(20), we have

$$\left\|\tilde{\phi}_n(u) - \tilde{\phi}_n(u')\right\| = \left\|\tilde{\phi}_n(w) - \tilde{\phi}_n(w')\right\| \pm C_0^\star\varepsilon_n/\lambda_n$$

$$= \left\|w - w'\right\| \pm \left(C_0^\star + C_3^\star\right)\varepsilon_n/C_2^\star =: \left\|w - w'\right\| \pm C_1\varepsilon_n.$$

Let

$$\xi_1 = \min_k \frac{\theta(U_k)}{\operatorname{diam}(U_k)}, \tag{21}$$

which is strictly positive. The result of [2], namely Lemma 18, gives a constant $C_2 \propto \xi_1$ and an isometry R_k such that $\max_{u \in U_k} \|\tilde{\phi}_n(u) - R_k(u)\| \leq C_2 \varepsilon_n$.

Let

$$\xi_2 = \frac{1}{2} \min\{\rho(U_k \cap U_{k'}) : U_k \cap U_{k'} \neq \varnothing\}. \tag{22}$$

Take $k, k' \in [K]$ such that $U_k \cap U_{k'} \neq \varnothing$, so that there is $u \in U$ such that $B(u, \xi_2) \subset U_k \cap U_{k'}$. Since

$$\max_{x \in B(u, \xi_2)} \|R_k(x) - R_{k'}(x)\| \leq \max_{x \in U_k \cap U_{k'}} \|R_k(x) - \tilde{\phi}_n(x)\| + \|\tilde{\phi}_n(x) - R_{k'}(x)\|$$

$$\leq \max_{x \in U_k} \|R_k(x) - \tilde{\phi}_n(x)\| + \max_{x \in U_{k'}} \|\tilde{\phi}_n(x) - R_{k'}(x)\| \leq 2 C_2 \varepsilon_n,$$

we have $\|R_k(x) - R_{k'}(x)\| \leq (2\|x - u\|/\xi_2 + 1) 2 C_2 \varepsilon_n$ for all $x \in \mathbb{R}^d$, by Lemma 20. Hence, $\|R_k(x) - R_{k'}(x)\| \leq (2 \operatorname{diam}(U)/\xi_2 + 1) 2 C_2 \varepsilon_n =: C_3 \varepsilon_n$ for all $x \in U$. If instead $U_k \cap U_{k'} = \varnothing$, we do as follows. Since U is connected, there is a sequence $k_0 = k, k_1, \ldots, k_m = k'$ in $[K]$, such that $U_{k_i} \cap U_{k_{i+1}} \neq \varnothing$. We thus have $\max_{x \in U} \|R_{k_i}(x) - R_{k_{i+1}}(x)\| \leq C_3 \varepsilon_n$. By the triangle inequality, we conclude that $\max_{x \in U} \|R_k(x) - R_{k'}(x)\| \leq K C_3 \varepsilon_n$ for any $k, k' \in [K]$. Noting that $R_1 = R_n^\star$ (since $U_1 = U_\star$), for any $k \in [K]$ and $x \in U_k$,

$$\|\tilde{\phi}_n(x) - R_n^\star(x)\| \leq \|R_k(x) - R_1(x)\| + C_2 \varepsilon_n \leq (K C_3 + C_2) \varepsilon_n.$$

We conclude that, for any $x \in U$,

$$\|\hat{\phi}_n(x) - \lambda_n R_n^\star(x)\| \leq (K C_3 + C_2) \lambda_n \varepsilon_n \leq (K C_3 + C_2) C_2^\star \varepsilon_n =: C_4 \varepsilon_n. \tag{23}$$

This concludes the proof when $d \geq 2$.

A refinement of the constant. Assume now that $U = U^h$ for some $h > 0$. Tracking the constants above, we see that they all depend only on $(d, \rho(U), \operatorname{diam}(U), \operatorname{diam}(Q))$, as well as ξ_1 and ξ_2 defined in (21) and (22), respectively. We note that $\operatorname{diam}(U_k) \leq r$ and $\rho(U_k) \geq \min(r/2, h)$ by Lemma 21, so that $\xi_1 \geq \min(r/2, h)/r$. To bound ξ_2, we can do as we did at the beginning of this section, so that at the end of that section, we can restrict our attention to chains $k_0, \ldots, k_m$ where $\|u_{k_j} - u_{k_{j+1}}\| \leq 2r/16 = r/8$. To be sure, fix $k, k' \in [K]$ and let $\gamma : [0, 1] \mapsto U$ be a curve such that $\gamma(0) = u_k$ and $\gamma(1) = u_{k'}$. Define $s_0 = 0$ and then $s_{j+1} = \inf\{s > s_j : \|\gamma(s) - u_{k_j}\| > r/16\}$, and let $k_{j+1} \in [K]$ be such that $\|\gamma(s_{j+1}) - u_{k_{j+1}}\| \leq r/16$, which is well-defined since $(u_k, k \in [K])$ is a $(r/16)$-packing of U. We then have

$$\|u_{k_j} - u_{k_{j+1}}\| \leq \|u_{k_j} - \gamma(s_j)\| - \|\gamma(s_{j+1}) - u_{k_{j+1}}\| \leq r/16 + r/16 = r/8.$$

We can therefore redefine ξ_2 in (22) as $\frac{1}{2} \min\{\rho(U_k \cap U_{k'}) : \|u_k - u_{k'}\| \leq r/8\}$. Because $U = U^h$, for each $k \in [K]$, there is v_k such that $u_k \in B(v_k, \min(r/16, h)) \subset U$. By the triangle inequality,

$B(v_k, \min(r/16, h)) \subset U_{k'}$ when $\|u_k - u_{k'}\| \leq r/8$, so that $\xi_2 \geq \min(r/16, h)$. So we see that everything depends on $(d, h, \rho(U), \mathrm{diam}(U), \mathrm{diam}(Q))$. The second part of the theorem now follows by invariance considerations.

3.4. Proof of Lemma 5

Let $c = \mathrm{ess\,inf}_U f$ and $C = \mathrm{ess\,sup}_U f$, which by assumption belong to $(0, \infty)$. Fix $i \in [n]$ and let $N_i = \#\{j \neq i : \|x_j - x_i\| \leq r\}$. For $j \neq i$, $p_i(j) := \mathbb{P}(\|x_j - x_i\| \leq r) = \int_{B(x_i, r)} f(u)\,\mathrm{d}u$. For an upper bound, we have

$$p_i(j) \leq C \, \mathrm{Vol}\big(B(x_i, r) \cap U\big) \leq C \, \mathrm{Vol}\big(B(x_i, r)\big) = C \zeta_d r^d =: Q,$$

where Vol denotes the Lebesgue measure in $\mathbb{R}^d$ and ζ_d is the volume of the unit ball in $\mathbb{R}^d$. Hence, $\mathbb{P}(N_i > 2(n-1)Q) \leq \mathbb{P}(\mathrm{Bin}(n-1, Q) > 2(n-1)Q) \leq e^{-(n-1)Q/3}$ by Bennett's inequality for the binomial distribution. By the union bound, we conclude that $\max_i N_i \leq 2(n-1)Q$ with probability at least $1 - ne^{-(n-1)Q/3}$, which tends to 1 if $nr^d \geq C_0 \log n$ and $C_0 > 0$ is sufficiently large.

For a lower bound, we use the following lemma.

Lemma 21. *Suppose $U \subset \mathbb{R}^d$ is open and such that $U = U^h$ for some $h > 0$. Then for any $x \in U$ and any $r > 0$, $B(x, r) \cap U$ contains a ball of radius $\min(r, h)/2$. Moreover, the closure of that ball contains x.*

Proof. By definition, there is $y \in U$ such that $x \in B(y, h) \subset U$. We then have $B(x, r) \cap U \supset B(x, r) \cap B(y, h)$, so it suffices to show that the latter contains a ball of radius $\min(r, h)/2$. By symmetry, we may assume that $r \leq h$. If $\|x - y\| \leq r/2$, then $B(x, r/2) \subset B(y, h)$ and we are done. Otherwise, let $z = (1 - t)x + ty$ with $t := r/2\|x - y\| \in (0, 1)$, and note that $B(z, r/2) \subset B(x, r) \cap B(y, h)$ and $x \in \partial B(z, r/2)$. □

Now that Lemma 21 is established, we apply it to get

$$p_i(j) \geq c \, \mathrm{Vol}\big(B(x_i, r) \cap U\big) \geq c \zeta_d \big(\min(r, h)/2\big)^d =: q.$$

Hence, $\mathbb{P}(N_i < (n-1)q/2) \leq \mathbb{P}(\mathrm{Bin}(n-1, q) < (n-1)q/2) \leq e^{-(6/7)(n-1)q}$. By the union bound, we conclude that $\min_i N_i \geq (n-1)q/2$ with probability at least $1 - ne^{-(6/7)(n-1)q}$, which tends to 1 if $nr^d \geq C_1 \log n$ and $C_1 > 0$ is sufficiently large. (Recall that h is fixed.)

3.5. More auxiliary results

We list here a few additional of auxiliary results that will be used in the proof of Theorem 4.

For $V \subset \mathbb{R}^d$ and $x, x' \in V$, define the intrinsic metric

$$\delta_V(x, x') = \sup\big\{L : \exists \gamma : [0, L] \mapsto V, \ 1\text{-Lipschitz, with } \gamma(0) = x, \gamma(L) = x'\big\},$$

where γ is 1-Lipschitz if $\|\gamma(s) - \gamma(t)\| \leq |s - t|$ for all $s, t \in [0, L]$. If no such curve exists, set $\delta_V(x, x') = \infty$. The intrinsic diameter of V is defined as $\sup\{\delta_V(x, x') : x, x' \in V\}$. We note that, if $L := \delta_V(x, x') < \infty$, then there is a curve $\gamma \subset \bar{V}$ with length L joining x and x'. Recall that a curve with finite length is said to be rectifiable. See [5] for a detailed account of intrinsic metrics.

For $U \subset \mathbb{R}^d$ and $h > 0$, let $U^{\ominus h} = \{x \in U : B(x, h) \subset U\}$. This is referred to as an erosion (of the set U) in mathematical morphology.

Lemma 22. *If $U \subset \mathbb{R}^d$ is open and connected, then for each pair of points $x, x' \in U$, there is $h > 0$ and a rectifiable curve within $U^{\ominus h}$ joining x and x'.*

Proof. Take $x, x' \in U$. By taking an intersection with an open ball that contains x, x', if needed, we may assume without loss of generality that U is bounded. Since every connected open set in a Euclidean space is also path-connected [26], Example 2.5.13, there is a continuous curve $\gamma : [0, 1] \mapsto U$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. A priori, γ could have infinite length. However, $\gamma \ (\equiv \gamma([0, 1]))$ is compact. For each $t \in [0, 1]$, let $r(t) > 0$ be such that $B_t := B(\gamma(t), r(t)) \subset U$. Since $\gamma \subset \bigcup_{t \in [0,1]} B_t$, there is $0 \leq t_1 < \cdots < t_m \leq 1$ such that $\gamma \subset \bigcup_{j \in [m]} B_{t_j}$. Since γ is connected, necessarily, for all $j \in [m-1]$ there is $s_j \in [t_j, t_{j+1}]$ such that $\gamma(s_j) \in B_{t_j} \cap B_{t_{j+1}}$. Let $s_0 = 0$ and $s_m = 1$. Then $[\gamma(s_j)\gamma(s_{j+1})] \subset B_{t_{j+1}} \subset U$ for all $j \in 0, \ldots, m-1$, and therefore the polygonal line defined by $x = \gamma(s_0), \gamma(s_1), \ldots, \gamma(s_{m-1}), \gamma(s_m) = x'$ is inside $\bigcup_{j \in [m]} B_{t_j} \subset U^{\ominus r}$ where $r := \min_{j \in [m]} r(t_j) > 0$. By construction, this polygonal line joins x and x', and is also rectifiable since it has a finite number of vertices. $\qquad\square$

Lemma 23. *Suppose $U \subset \mathbb{R}^d$ is bounded, connected, and such that $U = U^h$ for some $h > 0$. Then there is $h_{\ddagger} > 0$ such that, for all $h' \in [0, h_{\ddagger}]$, the intrinsic diameter of $U^{\ominus h'}$ is finite.*

Proof. Let $V = U^{\ominus h}$. By assumption, for all $x \in U$, there is $y \in V$ such that $x \in B(y, h) \subset U$. In particular, $U \supset V \neq \varnothing$.

Let V_1 be a connected component of V. Pick $y_1 \in V_1$ and note that $B_1 := B(y_1, h) \subset U$ by definition, and also $B_1 \subset V_1$ because B_1 is connected. Let ζ_d be the volume of the unit ball in $\mathbb{R}^d$. Since the connected components are disjoint and each has volume at least $\zeta_d h^d$ while U has volume at most $\zeta_d (\mathrm{diam}(U)/2)^d$, V can have at most $\lceil (\mathrm{diam}(U)/2h)^d \rceil$ connected components, which we now denote by $V_1, \ldots, V_K$. Pick $y_k \in V_k$ for each $k \in [K]$. Applying Lemma 22, for each pair of distinct $k, k' \in [K]$, there is a rectifiable (i.e., finite-length) path $\gamma_{k,k'} \subset U$ joining y_k and $y_{k'}$. By Lemma 22, the length of $\gamma_{k,k'}$, denoted $D_{k,k'}$, is finite, and there is $h_{k,k'} > 0$ such that $\gamma_{k,k'} \subset U^{\ominus h_{k,k'}}$. Let $D_{\ddagger} = \max_{k,k' \in [K]} D_{k,k'}$ and $h_{\ddagger} = \min_{k,k' \in [K]} h_{k,k'}$.

We now show that each connected component V_k has finite diameter in the intrinsic metric of $V' := U^{\ominus h/2}$. Since V_k is bounded, there is $x_1, \ldots, x_m \in V_k$ such that $V_k \subset \bigcup_{j \in [m_k]} Q_j$, where $Q_j := B(x_j, h/2) \subset V'$. Take any $x, x' \in V_k$. Let $j, j' \in [m_k]$ be such that $x \in Q_j$ and $x' \in Q_{j'}$. Since V_k is connected, there is a sequence $j = j_0, j_1, \ldots, j_{S_k} = j' \in [m_k]$ such that $Q_{j_s} \cap Q_{j_{s+1}} \neq \varnothing$ for all $s = 0, \ldots, S_k$. Choose $z_s \in Q_{j_s} \cap Q_{j_{s+1}}$ and let $z_0 = x$ and $z_{S_k} = x'$. Then $[z_s z_{s+1}] \subset Q_{j_{s+1}}$ for all s. Let L be the polygonal line formed by $z_0, \ldots, z_{S_k}$. By construction, $L \subset \bigcup_{s=0}^{S_k} Q_{j_s} \subset V'$, it joins x and x', and has length at most $(S_k + 1)2h$. Hence, $\delta_{V'}(x, x') \leq$

$(S_k + 1)2h \le 2(m_k + 1)h$. This being valid for all $x, x' \in V_k$, we proved that V_k has diameter at most $D_k := 2(m_k + 1)h$ in the intrinsic metric of V'. Let $D_\star = \max_{k \in [K]} D_k$.

Now take $h_\dagger \in [0, h_\ddagger]$ and any $x, x' \in U^{\ominus h_\dagger}$. Let $y, y' \in V$ be such that $x \in B(y, h)$ and $x' \in B(y', h)$. Let $k, k' \in [K]$ be such that $y \in V_k$ and $y' \in V_{k'}$. There are curves $\gamma, \gamma' \subset V'$ of length at most $D_\star$ such that γ joins y and y_k, while γ' joins y' and $y_{k'}$. We then join y_k and $y_{k'}$ with $\gamma_{k,k'}$ All together, we have the curve $[xy] \cup \gamma \cup \gamma_{k,k'} \cup \gamma' \cup [y'x']$, which joins x and x', lies entirely in $U^{\ominus h_\dagger}$, and has length bounded by $h + D_\star + D_\ddagger + D_\star + h =: D$. And this is true for any pair of such points. $\qquad \square$

Lemma 24. *Suppose that $S_1, S_2 : \mathbb{R}^d \mapsto \mathbb{R}^d$ are two affinities such that $\max_j \|S_1(z_j) - S_2(z_j)\| \le \varepsilon$, where $z_0, \ldots, z_d$ form in a η-approximate regular simplex with minimum edge length at least λ. There is $C > 0$ depending only on d such that, if $\eta \le 1/C$, then $\|S_1(x) - S_2(x)\| \le C\varepsilon\|x - z_0\|/\lambda + \varepsilon$ for all $x \in \mathbb{R}^d$.*

Proof. Note that this is closely related to Lemma 20. By translation and scale invariance, assume that $z_0 = 0$ and $\lambda = 1$. Let $L_i = S_i - S_i(0)$. We have $\|L_1(z_j) - L_2(z_j)\| \le \|S_1(z_j) - S_2(z_j)\| + \|S_1(0) - S_2(0)\| \le 2\varepsilon$. Let Z denote the matrix with columns $z_1, \ldots, z_d$. In matrix notation, we have

$$\left\|(L_1 - L_2)Z\right\|_F = \sqrt{\sum_j \left\|(L_1 - L_2)z_j\right\|^2} \le 2\sqrt{d}\varepsilon.$$

We also have $\|(L_1 - L_2)Z\|_F \ge \|(L_1 - L_2)Z\| \ge \sigma_d(Z)\|L_1 - L_2\|$, and by Lemma 12, $\sigma_d(Z) = \sigma_d([z_0, Z]) \ge 1/C_1$ when $\eta \le 1/C_1$, where C_1 depends only on d. In that case, $\|L_1 - L_2\| \le C_2\varepsilon$ for another constant C_2. Equivalently, for $x \in \mathbb{R}^d$, $\|L_1(x) - L_2(x)\| \le C_2\varepsilon\|x\|$, which in turn implies that $\|S_1(x) - S_2(x)\| \le \|L_1(x) - L_2(x)\| + \|S_1(0) - S_2(0)\| \le C_2\varepsilon\|x\| + \varepsilon$. $\qquad \square$

3.6. Proof of Theorem 4

Because ϕ_n is bounded independently of n, we may assume without loss of generality that $C_0\varepsilon_n \le r_n$ and $C_0 r_n \le h$ for all n, where $C_0 \ge 1$ will be chosen large enough later on.

Take $y \in U$ and let $\Omega_y = \Omega_n \cap B(y, r_n)$ and $Q_y = \phi_n(\Omega_y)$. We first show that there is $C_1 \propto \text{diam}(Q)/\rho(U)$ such that, for any $y \in U$, $\text{diam}(Q_y) \le C_1 r_n$. For this, we mimic the proof of Lemma 3. Take $x, x' \in \Omega_y$ such that $\xi := \|\phi_n(x) - \phi_n(x')\| = \text{diam}(Q_y)$. Let u be such that $B(u, \rho(U)) \subset U$. Let $y_1, \ldots, y_m$ be an $(r_n + 2\varepsilon_n)$-packing of $B(u, \rho(U))$ with $m \ge A_1(\rho(U)/r_n)^d$ for some $A_1 \propto 1$. Then let $\{x_{i_s} : s \in [m]\} \subset \Omega_n$ be such that $\max_{s \in [m]} \|y_s - x_{i_s}\| \le \varepsilon_n$. By the triangle inequality, for all $s \ne t$, we have $\|x_{i_s} - x_{i_t}\| \ge \|y_s - y_t\| - 2\varepsilon_n \ge r_n > \|x - x'\|$. By (10), we have $\|\phi_n(x_{i_s}) - \phi_n(x_{i_t})\| \ge \xi$, so that $\phi_n(x_{i_1}), \ldots, \phi_n(x_{i_m})$ form a ξ-packing. Therefore, $m \le A_2(\text{diam}(Q)/\xi)^d$ for some $A_2 \propto 1$. We conclude that $\xi \le (A_2/A_1)^{1/d}(\text{diam}(Q)/\rho(U))r_n =: C_1 r_n$.

We apply Theorem 3 to $U_y := B(y, r_n)$ and Ω_y. With the fact that $\delta_H(\Omega_y, U_y) \le 2\varepsilon_n$ – as we saw in the proof of (17) – and invariance considerations, we obtain a constant $C \propto 1$ and a similarity S_y such that $\max_{x \in \Omega_y} \|\phi_n(x) - S_y(x)\| \le C(\text{diam}(Q_y)/r_n)\varepsilon_n \le CC_1\varepsilon_n =: C_2\varepsilon_n$. (Note that all the quantities with subscript y depend also on n, but this will be left implicit.)

Fix $y_\star \in U^{\ominus r_n}$. For $x \in \Omega_n$, there is $y \in U^{\ominus r_n}$ such that $x \in U_y$. Assume γ is parameterized by arc length and let $h_\ddagger$ be given by Lemma 23 and let D denote the intrinsic diameter of $U^{\ominus h_\ddagger}$. Then assuming $r_n \le h_\ddagger$, there is a curve $\gamma \subset U^{\ominus r_n}$ of length $L \le D$ joining $y_\star$ and y. Let $y_0 = y_\star$, $y_j = \gamma(j r_n)$ for $j = 0, \dots, J := \lfloor L/r_n \rfloor$, and then $y_{J+1} = y$. We have $\max_{z \in U_{y_j} \cap U_{y_{j+1}}} \|S_{y_j}(z) - S_{y_{j+1}}(z)\| \le 2C_2 \varepsilon_n$ by the triangle inequality. We also have $\rho(U_{y_j} \cap U_{y_{j+1}}) \ge r_n$, because $\|y_j - y_{j+1}\| \le r_n$. Let v_j be such that $B(v_j, r_n/2) \subset U_{y_j} \cap U_{y_{j+1}}$. Fix j and let $v_{j,0}, \dots, v_{j,d}$ denote a regular simplex inscribed in the ball $B(v_j, r_n/4)$. Let $\lambda_n \propto r_n$ denote its edge length. Then let $x_{j,0}, \dots, x_{s,d} \in \Omega_n$ be such that $\max_k \|x_{j,k} - v_{j,k}\| \le \varepsilon_n$. When C_0 is large enough, $x_{j,0}, \dots, x_{j,d} \in B(v_j, r_n/2)$ by the triangle inequality. Moreover, $\max_{k,l} \|x_{j,k} - x_{j,l}\| \le \lambda_n + 2\varepsilon_n$, as well as $\min_{k \ne l} \|x_{j,k} - x_{j,l}\| \ge \lambda_n - 2\varepsilon_n$. When C_0 is large enough, $F_j := \{x_{j,0}, \dots, x_{j,d}\}$ is therefore an η-approximate regular simplex, with $\eta \propto \varepsilon_n/r_n$, and minimum edge length $\propto r_n$. Now, since $\max_k \|S_{y_j}(x_{j,k}) - S_{y_{j+1}}(x_{j,k})\| \le 2C_2 \varepsilon_n$, by Lemma 24, for all $z \in \mathbb{R}^d$, $\|S_{y_j}(z) - S_{y_{j+1}}(z)\| \le CC_2 \varepsilon_n \|z - x_{j,0}\|/r_n + 2C_2 \varepsilon_n$ for some $C \propto 1$, assuming $\varepsilon_n/r_n \le 1/C$. In particular, by the fact that $\|x - x_{j,0}\| \le \operatorname{diam}(U)$, this gives $\|S_{y_j}(x) - S_{y_{j+1}}(x)\| \le C_3 \varepsilon_n/r_n$ for some $C_3 \propto \operatorname{diam}(U) C_2$. Hence,

$$\|S_{y_\star}(x) - S_y(x)\| \le (J+1) C_3 \varepsilon_n/r_n \le C_4 \varepsilon_n/r_n^2,$$

since $J \le L/r_n \le D/r_n$.

This being true for any arbitrary $x \in \Omega_n$, we conclude that

$$\max_{x \in \Omega_n} \|\phi_n(x) - S_{y_\star}(x)\| \le C_4 \varepsilon_n/r_n^2 + C_2 \varepsilon_n \le C_5 \varepsilon_n/r_n^2.$$

4. Discussion

This paper builds on [14] to provide some theory on ordinal embedding, an important problem in multivariate statistics (aka unsupervised learning). We leave open two main problems:

- What are the optimal rates of convergence for ordinal embedding with all triple and quadruple comparisons?
- What is the minimum size of $K = K_n$ for consistency of ordinal embedding based on the K-nearest neighbor distance comparisons?

We note that we only studied the large sample behavior of exact embedding methods. In particular, we did not discuss or proposed any methodology for producing such an embedding. For this, we refer the reader to [1,4,23] and references therein. In fact, the practice of ordinal embedding raises a number of other questions in terms of theory, for instance:

- How many flawed comparisons can be tolerated?

Acknowledgements

References

[1] Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D.J. and Belongie, S. (2007). Generalized non-metric multidimensional scaling. In *International Conference on Artificial Intelligence and Statistics* 11–18.

[2] Alestalo, P., Trotsenko, D.A. and Väisälä, J. (2001). Isometric approximation. *Israel J. Math.* **125** 61–82. MR1853806

[3] Aumann, R.J. and Kruskal, J.B. (1958). The coefficients in an allocation problem. *Naval Res. Logist. Quart.* **5** 111–123. MR0096550

[4] Borg, I. and Groenen, P.J.F. (2005). *Modern Multidimensional Scaling*: *Theory and Applications*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2158691

[5] Burago, D., Burago, Y. and Ivanov, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Providence, RI: Amer. Math. Soc.. MR1835418

[6] Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329. MR2977397

[7] Davenport, M.A. (2013). Lost without a compass: Nonmetric triangulation and landmark multidimensional scaling. In *Computational Advances in Multi-Sensor Adaptive Processing* (*CAMSAP*), 2013 *IEEE 5th International Workshop on* 13–16. IEEE.

[8] De Silva, V. and Tenenbaum, J.B. (2004). Sparse multidimensional scaling using landmark points. Technical report, Stanford Univ.

[9] Ellis, D.P., Whitman, B., Berenzweig, A. and Lawrence, S. (2002). The quest for ground truth in musical artist similarity. In *Proceedings of the International Symposium on Music Information Retrieval* (*ISMIR*) 170–177.

[10] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078

[11] Horn, R.A. and Johnson, C.R. (1990). *Matrix Analysis*. Cambridge: Cambridge Univ. Press. MR1084815

[12] Jamieson, K.G. and Nowak, R.D. (2011). Low-dimensional embedding using adaptively selected ordinal data. In *Communication*, *Control*, *and Computing* (*Allerton*), 2011 49*th Annual Allerton Conference on* 1077–1084. IEEE.

[13] Kelley, J.L. (1975). *General Topology*. New York: Springer. MR0370454

[14] Kleindessner, M. and Von Luxburg, U. (2014). Uniqueness of ordinal embedding. In *Proceedings of the* 27*th Conference on Learning Theory* 40–67.

[15] Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29** 1–27. MR0169712

[16] McFee, B. and Lanckriet, G. (2011). Learning multi-modal similarity. *J. Mach. Learn. Res.* **12** 491–523. MR2783175

[17] Nhat, V.D.M., Vo, N., Challa, S. and Lee, S. (2008). Nonmetric MDS for sensor localization. In 3*rd International Symposium on Wireless Pervasive Computing* (*ISWPC*) 396–400.

[18] Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* **27** 125–140. MR0140376

[19] Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* **27** 219–246. MR0173342

[20] Shepard, R.N. (1966). Metric structures in ordinal data. *J. Math. Psych.* **3** 287–315.

[21] Sikorska, J. and Szostok, T. (2004). On mappings preserving equilateral triangles. *J. Geom.* **80** 209–218. MR2176582

[22] Suppes, P. and Winet, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Sci.* **1** 259–270. MR0073099

[23] Terada, Y. and Von Luxburg, U. (2014). Local ordinal embedding. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*-14) 847–855.

[24] Vestfrid, I.A. (2003). Linear approximation of approximately linear functions. *Aequationes Math.* **66** 37–77. MR2003454

[25] Von Luxburg, U. and Alamgir, M. (2013). Density estimation from unweighted k-nearest neighbor graphs: A roadmap. In *Advances in Neural Information Processing Systems* 225–233.

[26] Waldmann, S. (2014). *Topology*: *An Introduction*. Cham: Springer. MR3308923

[27] Young, F.W. and Hamer, R.M., eds. (1987). *Multidimensional Scaling*: *History*, *Theory*, *and Applications*. Lawrence Erlbaum Associates.