Impossibility of Sketching of the 3D Transportation Metric with Quadratic Cost*

Alexandr Andoni† Assaf Naor‡ Ofer Neiman§

February 17, 2016

Abstract

Transportation cost metrics, also known as the Wasserstein distances W_p, are a natural choice for defining distances between two pointsets, or distributions, and have been applied in numerous fields. From the computational perspective, there has been an intensive research effort for understanding the W_p metrics over $\mathbb{R}^k$, with work on the W_1 metric (a.k.a *earth mover distance*) being most successful in terms of theoretical guarantees. However, the W_2 metric, also known as the *root-mean square* (RMS) bipartite matching distance, is often a more suitable choice in many application areas, e.g. in graphics. Yet, the geometry of this metric space is currently poorly understood, and efficient algorithms have been elusive. For example, there are no known non-trivial algorithms for nearest-neighbor search or sketching for this metric.

In this paper we take the first step towards explaining the lack of efficient algorithms for the W_2 metric, even over the three-dimensional Euclidean space $\mathbb{R}^3$. We prove that there are no meaningful embeddings of W_2 over $\mathbb{R}^3$ into a wide class of normed spaces, as well as that there are no efficient sketching algorithms for W_2 over $\mathbb{R}^3$ achieving constant approximation. For example, our results imply that: 1) any embedding into L_1 must incur a distortion of $\Omega(\sqrt{\log n})$ for pointsets of size n equipped with the W_2 metric; and 2) any sketching algorithm of size s must incur $\Omega\left(\sqrt{\log n}/\sqrt{s}\right)$ approximation. Our results follow from a more general statement, asserting that W_2 over $\mathbb{R}^3$ contains the 1/2-snowflake of *all* finite metric spaces with a uniformly bounded distortion. These are the first non-embeddability/non-sketchability results for W_2.

*Present manuscript overlaps with `http://arxiv.org/abs/1509.08677` and is intended for the theoretical computer science audience. In particular, here we omit some of the results from arXiv:1509.08677, but include the applications to TCS.

†Computer Science Department, Columbia University, 500 West 120th Street, Room 450, MC0401, New York, NY 10027, USA. Email: `andoni@cs.columbia.edu`

‡Mathematics Department, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA. Email: `naor@math.princeton.edu`. A. N. was supported in part by the BSF, the Packard Foundation and the Simons Foundation.

§Department of Computer Science, Ben-Gurion University of the Negev, P.O.B 653 Be'er Sheva 84105, Israel. Email: `neimano@cs.bgu.ac.il`. O. N. was supported in part by the ISF and the European Union's Seventh Framework Programme.

1 Introduction

Transportation metrics provide a natural distance on sets of points, or probability measures more generally, and as such have applications in numerous fields, such as computer science, as well as statistical physics, mathematical economics, automated control, shape optimization, applied probability, partial differential equations, metric geometry and many more, see [58, 53]. These metrics are also known as Wasserstein distance, Kantorovich-Rubinstein distance, Prokhorov distance, or the earth mover distance. We now recall basic notation and terminology from the theory of transportation cost metrics [73]. For a metric space (X, d_X) and $p \in (0, \infty)$, let $\mathcal{P}_p(X)$ denote the space of all (Borel) probability measures μ on X satisfying $\int_X d_X(x, x_0)^p \mathrm{d}\mu(x) < \infty$ for some (hence all) $x_0 \in X$. The *Wasserstein-p distance* between $\mu, \nu \in \mathcal{P}_p(X)$ is then

$$\mathsf{W}_p(\mu, \nu) \stackrel{\text{def}}{=} \inf_{\pi \in \Pi(\mu,\nu)} \left(\iint_{X \times X} d_X(x, y)^p \mathrm{d}\pi(x, y) \right)^{\frac{1}{p}},$$

where $\Pi(\mu, \nu)$ is the set of all couplings (matchings) π between (μ, ν) on X, i.e., probability measures π on $X \times X$ such that $\mu(A) = \pi(A \times X)$ and $\nu(A) = \pi(X \times A)$ for every $A \subseteq X$. W_p on $\mathcal{P}_p(X)$ is a metric whenever $p \geqslant 1$. Here we consider the classic setting of X being $\mathbb{R}^k$, for $k \geqslant 2$, endowed with the standard Euclidean distance.

In computer science, the transportation metrics on $\mathbb{R}^k$ play an important role in computer vision [74, 60, 28, 29, 35, 56, 51, 41], machine learning [25], information retrieval [59], and mechanism design [19], among others. For example, an image can be represented as a set of pixels in a color space $\mathbb{R}^3$; the transportation cost between such sets yields an accurate measure of dissimilarity between color characteristics of the images [61, 32].

These applications motivated a lot of research into the *computational* properties of transportation metrics. In particular, typical problems are to develop efficient algorithms for: computing the distance between two pointsets (finitely-supported measures), nearest neighbor search under these metrics, as well as problems in the streaming and sketching context.

So far, most of the rigorous algorithmic results have been developed for the W_1 metric, often refered to as the Earth Mover Distance (EMD). There is a long line of work on approximation algorithms for computing EMD between two pointsets in $\mathbb{R}^k$ [71, 2, 72, 1, 31, 64], culminating in a near-linear time algorithm achieving a $(1+\varepsilon)$-approximation [65, 3, 7]. Nearest neighbor search algorithms all proceed via either embedding EMD into L_1 or sketching. Understanding the embeddability of EMD over $\mathbb{R}^k$ into L_1 is a well-known open problem [38], and the best distortion is currently known [17, 32, 33, 50, 5] to be between $O(k \log n)$ and $\Omega(k + \sqrt{\log n})$ for pointsets in $[n]^k = \{1, 2, \ldots n\}^k$. Similarly, designing sketching algorithms for EMD over $\mathbb{R}^k$ is also a well-known open problem [54, 55]. Some of the sketching bounds for W_1 follow from the aforementioned L_1 embeddings, and some others are proved directly [4, 6].

Yet, in a number of applications the Wasserstein-2 distance W_2 is a *more natural* distance than Wasserstein-1 (EMD), and indeed other communities have paid more attention to W_2

[68]. Specifically, W_2 (a.k.a., *root-mean square bipartite matching distance*) corresponds to the "ℓ_2 error" between two pointsets, in contrast to the "ℓ_1 error" measured by W_1; as such they have better regularity properties and also have a differential interpretation [68]. See [42, 21] for a further discussion of why using W_2 gives results of a better quality than W_1. W_2 is used in graphics [66, 67, 69, 68], for shape interpolation [15], for barycenter computation [18, 14], shape reconstruction [22], blue noise generation [21], triangulations [42], among others.

Surprisingly, the algorithmic results for W_2 have been much more elusive. The best algorithms for computing W_2 distance between two pointsets follow from [57, 3], who obtain $\tilde{O}(n^2)$ time for exact and $\tilde{O}(n^{3/2})$ for approximate computation (in contrast to the near-linear time algorithms for W_1). Beyond these results, there are no known non-trivial algorithms for embedding, nearest neighbor search, or sketching for W_2! This discrepancy raises the question of why there has been such a dire lack of progress on algorithms for W_2.

Here we address this question by proving the first explicit lower bounds for W_2 over $\mathbb{R}^3$, establishing that it is a very rich space that cannot be represented faithfully even with weak guarantees in a large class of normed spaces (that includes all L_q spaces for finite q, and much more). In particular, focusing on W_2 on measures over $\mathbb{R}^3$ supported on at most n points, we show that $\Omega(\sqrt{\log n})$ distortion is required for either: 1) embedding of W_2 into L_1, and 2) constant-size sketching. To contrast these results to those known for W_1 over the same set of measures, while W_1 has a similar non-embeddability into L_1 [50], it does not translate into *sketching* lower bounds. In fact, it was only recently established [6] that the approximation for sketching W_1 must be super-constant (without giving an explicit bound). Besides stronger sketching lower bounds, our results for W_2 are stronger than any known W_1 non-embeddability results since they apply to a larger class of Banach space targets (nontrivial type), and also rule out embeddings that are much weaker than bi-Lipschitz, like coarse embeddings. Finally, our results also apply to W_p space for $p \in (1,2)$, yielding a $\Omega((\log n)^{1/p})$ distortion lower bound, which is asymptotically stronger than the distortion lower bound known for embedding W_1 into L_1.

Our results apply to measures over $\mathbb{R}^3$ only, and the validity of analogous results for measures over $\mathbb{R}^2$ remains an open question. The only progress has been obtained in the forthcoming work [8], where the authors establish the first lower bound for embedding $\mathsf{W}_2(\mathbb{R}^2)$ into L_1, showing that the distortion goes to infinity (without an explicit bound). However, [8] does not yield the full strength of our results in terms of ruling out embeddings into spaces with nontrivial type, as well as, say, coarse embeddings.

1.1 Main Results

We now present our results on non-existence of good embedding and sketching methods for W_2 over $\mathbb{R}^3$. We then show that these results follow from a more general principle: that W_2 over $\mathbb{R}^3$ is snowflake-universal, and hence, say, we can embed the square-root of a shortest path metric on an expander graph into it with distortion arbitrarily close to 1. Our results apply to all W_p for $p > 1$, but not to W_1.

Non-embeddability results. We now introduce the standard notion of embeddings.

Definition 1. Fix two metric spaces (X, d_X) and (Y, d_Y), and $D \in [1, \infty]$. A mapping $f : X \to Y$ is an embedding with distortion at most D if there exists $s \in (0, \infty)$ such that every $x, y \in X$ satisfy $s \cdot d_X(x, y) \leqslant d_Y(f(x), f(y)) \leqslant Ds \cdot d_X(x, y)$. The infimum over those $D \in [1, \infty]$ for which this holds true is called the distortion of f and is denoted $\mathbf{dist}(f)$. If there exists a mapping $f : X \to Y$ with distortion at most D then we say that (X, d_X) embeds with distortion D into (Y, d_Y). The infimum of $\mathbf{dist}(f)$ over all $f : X \to Y$ is denoted $c_{(Y,d_Y)}(X, d_X)$, or $c_Y(X)$ if the metrics are clear from the context.

We prove the following theorem.

Theorem 2. *For any fixed $p \in (1, \infty)$ and $n \in \mathbb{N}$, consider the metric space X consisting of all the measures on $\mathbb{R}^3$ that are supported on at most n points, equipped with the W_p metric. Then any embedding of X into L_1 must incur distortion $\Omega(((p-1)\log n)^{1/p})$.*

Theorem 2 implies a $\Omega(\sqrt{\log n})$ approximation for any algorithmic approach proceeding via embedding W_2 over measures on $\mathbb{R}^3$ whose support is of size at most n into L_1. While embedding into L_1 is a common algorithmic technique for high-dimensional metric spaces, it is not the only one. In particular, despite non-embeddability into L_1, a metric could admit a better embedding into, say, $L_{1/2}$, which would imply efficient sketches and nearest neighbor search algorithms. We rule out such weaker embeddings as well.

In fact, our work actually yields impossibility results that are much stronger than the bi-Lipschitz nonembeddability statement that corresponds to Theorem 2. Our most general results are contained in the full version of this paper, but here is one illustrative example. Let X be either L_1 or a Banach space of nontrivial type.[1] Then for $p \in (1, \infty)$ there do not exist any nondecreasing functions $\alpha, \beta : [0, \infty) \to [0, \infty)$ with $\lim_{t\to\infty} \alpha(t) = \infty$ for which there is a mapping $f : \mathcal{P}_p(\mathbb{R}^3) \to X$ that satisfies

$$\forall\, \mu, \nu \in \mathcal{P}_p(\mathbb{R}^3), \qquad \alpha(\mathsf{W}_p(\mu, \nu)) \leqslant \|f(\mu) - f(\nu)\|_X \leqslant \beta(\mathsf{W}_p(\mu, \nu)).$$

Theorem 2 corresponds to the special case when the function α, β are linear and X is L_1. In common metric geometry jargon, the above statement asserts that $\mathcal{P}_p(\mathbb{R}^3)$ fails to admit a coarse embedding into any normed space of nontrivial type.

Sketching. We can also state our results using the language of the sketching algorithms. The notion of sketching is defined as follows [62].

Definition 3. Fix a metric (X, d_X), and approximation $D \geqslant 1$. We say (X, d_X) has sketching complexity $s \geqslant 1$ if, for any threshold $r > 0$, there exists a distribution over sketching maps $\mathrm{sk} : X \to \{0, 1\}^s$ and reconstruction algorithms $R : \{0, 1\}^s \times \{0, 1\}^s \to \{\mathrm{close}, \mathrm{far}\}$, satisfying the following. For any $x, y \in X$, with at least 2/3 probability of success:

[1] The correct class of Banach spaces here could even be all those Banach spaces that do not contain *every* finite metric space with distortion arbitrarily close to 1, but currently this stronger version of the ensuing statement holds true conditionally on a well-known open question in metric geometry; see the full version of this paper for more details.

- if $d_X(x,y) \leqslant r$, then $R(\mathrm{sk}(x),\mathrm{sk}(y)) = \text{close}$;
- if $d_X(x,y) > Dr$, then $R(\mathrm{sk}(x),\mathrm{sk}(y)) = \text{far}$.

We are now ready to state our sketching lower bound for W_p for $p > 1$.

Theorem 4. *Fix $p \in (1,\infty)$ and let $n,s \in \mathbb{N}$. Consider the metric space X consisting of all the measures on $\mathbb{R}^3$ that are supported on at most n points, equipped with the W_p metric. Then any sketching algorithm for X with sketching complexity s must have an approximation guarantee of $D = \Omega\left(\left(\frac{(p-1)\log n}{s}\right)^{1/p}\right)$.*

We note that, for comparison, standard ℓ_1, ℓ_2 metrics have constant sketching complexity [34, 62, 11]. Also, for W_1 over $\mathbb{R}^3$ (or $\mathbb{R}^2$), the only known lower bound is that $Ds = \omega(1)$, shown recently in [6], based on [50].

Snowflake universality. Our results follow from a more general phenomenon, captured by the following theorem.

Theorem 5. *If $p \in (1,\infty)$ then for every finite metric space (X,d_X) we have*

$$c_{(\mathcal{P}_p(\mathbb{R}^3),\mathsf{W}_p)}\left(X, d_X^{\frac{1}{p}}\right) = 1.$$

For a metric space (X,d_X) and $\theta \in (0,1]$, the metric space (X,d_X^θ) is commonly called the θ-*snowflake* of (X,d_X); see e.g. [20]. Thus Theorem 5 asserts that the θ-snowflake of any finite metric space (X,d_X) embeds with distortion $1+\varepsilon$ into $\mathcal{P}_p(\mathbb{R}^3)$ for every $\varepsilon \in (0,\infty)$ and $\theta \in (0,1/p]$.[2] Our techniques fall short of proving a longstanding conjecture of Bourgain [16], who asked whether $(\mathcal{P}_1(\mathbb{R}^2),\mathsf{W}_1)$ is not universal (i.e., does not contain all finite metrics).[3] Bourgain proved in [16] that $(\mathcal{P}_1(\ell_1),\mathsf{W}_1)$ is universal (despite the fact that ℓ_1 is not universal), but it remains an intriguing open question to determine whether or not $(\mathcal{P}_1(\mathbb{R}^k),\mathsf{W}_1)$ is universal for any finite $k \in \mathbb{N}$, the case $k = 2$ being most challenging.

Theorem 6 below implies that Theorem 5 is sharp if $p \in (1,2]$, and yields a nontrivial, though probably non-sharp, restriction on the embeddability of snowflakes into $\mathcal{P}_p(\mathbb{R}^3)$ also for $p \in (2,\infty)$.

Theorem 6. *For arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space (X_n, d_{X_n}) such that for every $\alpha \in (0,1]$ we have*

$$c_{(\mathcal{P}_p(\mathbb{R}^3),\mathsf{W}_p)}(X_n, d_{X_n}^\alpha) \gtrsim \begin{cases} (\log n)^{\alpha-\frac{1}{p}} & \text{if } p \in (1,2], \\ (\log n)^{\alpha+\frac{1}{p}-1} & \text{if } p \in (2,\infty). \end{cases}$$

[2] Formally, Theorem 5 makes this assertion when $\theta = 1/p$, but for general $\theta \in (0,1/p]$ one can then apply Theorem 5 to the metric space $(X, d_X^{\theta p})$ to deduce the seemingly more general statement.

[3] Bourgain actually formulated this question as asking whether a certain Banach space (namely, the dual of the Lipschitz functions on the square $[0,1]^2$) has finite Rademacher cotype, but this is equivalent to the above formulation.

Here, and in what follows, we use standard asymptotic notation, i.e., for $a, b \in [0,\infty)$ the notation $a \gtrsim b$ (respectively $a \lesssim b$) stands for $a \geqslant cb$ (respectively $a \leqslant cb$) for some universal constant $c \in (0,\infty)$. The notation $a \asymp b$ stands for $(a \lesssim b) \wedge (b \lesssim a)$.

The rest of the paper is organized as follows. We give the proof of Theorem 5 in Section 2, and its consequences, Theorem 2 and 4, in Section 2.1. We then present some future research directions suggested by our results in Section 3. Finally, we prove the sharpness of Theorem 5, namely Theorem 6, in Appendix A.

2 Proof of Theorem 5

To establish the theorem, we will construct an explicit embedding of an n-point metric into $\mathsf{W}_2(\mathbb{R}^3)$. In what follows fix $n \in \mathbb{N}$ and an n-point metric space (X, d_X).

We start by presenting the intuition behind the construction. In particular, let us demonstrate a fundamental difference between W_1 and W_p for $p > 1$ for a simple transportation instance. We will exploit this construction in our embedding. Fix a positive integer k, and consider the optimal transport between the sets $A = \{0, \frac{1}{k}, \frac{2}{k}, \dots, \frac{k-1}{k}\}$ and $B = \{\frac{1}{k}, \frac{2}{k}, \dots, 1\}$. While under the W_1 metric the optimal cost is simply 1, under W_p the optimal transport would send every $x \in A$ to $x + \frac{1}{k} \in B$, which incurs a cost of $\left(\sum_{i=1}^{k}\left(\frac{1}{k}\right)^p\right)^{1/p} = k^{1/p-1} \xrightarrow[k\to\infty]{} 0$. Note that for any $0 \leqslant \varepsilon < 1$, we can increase the transport cost to ε by introducing a "gap" of size εk. E.g., for some i, define $A = \{0, \frac{1}{k}, \dots, \frac{i}{k}, \frac{i+\varepsilon k}{k}, \frac{i+\varepsilon k+1}{k}, \dots, \frac{k-1}{k}\}$ and $B = A \setminus \{0\} \cup \{1\}$. Then the optimal transport cost under W_p would be

$$\left(\left(\frac{\varepsilon k}{k}\right)^p + \sum_{i=1}^{k-\varepsilon k}\left(\frac{1}{k}\right)^p\right)^{1/p} \xrightarrow[k\to\infty]{} \varepsilon \ .$$

We shall use the fact that any graph, in particular the complete graph, can be realized in $\mathbb{R}^3$, so that if every edge is represented by a wire, there are no wire crossings (except at vertices). Imagine that each wire is replaced by a set of points with distances $1/k$ between neighboring points. We then introduce a gap of length proportional to $d_X(u,v)^{1/p}$ on the wire connecting u and v. The embedding of $u \in X$ will be into a uniform measure over the point realizing u, and all the points in all the wires. Then the transport from u to v must move the mass at u to the mass of v. By the simple example above, this can be done at cost proportional to $d_X(u,v)^{1/p}$, when k is sufficiently large. The trickier part is showing no better transport exist. To this end, we require that all the wires are sufficiently far apart, so any transport plan that does not move along the wires will have a huge cost. Finally, the triangle inequality ensures that the cost of a plan using the wires between the points $u = u_0, u_1, \dots, u_q = v$ is at least $d_X(u,v)^{1/p}$ (this is the reason why we make the gaps proportional to the p-th roots).

We now proceed with the formal proof of the theorem. Write $X = \{x_1, \dots, x_n\}$ and fix $\phi : \{1,\dots,n\} \times \{1,\dots,n\} \to \{1,\dots,n^2\}$ to be an arbitrary bijection between $\{1,\dots,n\} \times$

$\{1,\ldots,n\}$ and $\{1,\ldots,n^2\}$. Below it will be convenient to use the following notation.

$$m \stackrel{\mathrm{def}}{=} \min_{\substack{x,y\in X\\ x\neq y}} d_X(x,y)^{\frac{1}{p}} \qquad \text{and} \qquad M \stackrel{\mathrm{def}}{=} \max_{x,y\in X} d_X(x,y)^{\frac{1}{p}}. \tag{1}$$

Fix $K\in\mathbb{N}$. Denoting the standard basis of $\mathbb{R}^3$ by $e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3=(0,0,1)$, for every $i,j\in\{1,\ldots,n\}$ with $i<j$ define five families of points in $\mathbb{R}^3$ by setting for $s\in\{0,\ldots,K\}$,

$$Q_s^1(i,j) \stackrel{\mathrm{def}}{=} \frac{Mi}{m}e_1+\frac{M\phi(i,j)s}{mK}e_2, \tag{2}$$

$$Q_s^2(i,j) \stackrel{\mathrm{def}}{=} \frac{Mi}{m}e_1+\frac{M\phi(i,j)}{m}e_2+\frac{Ms}{mK}e_3, \tag{3}$$

$$Q_s^3(i,j) \stackrel{\mathrm{def}}{=} \frac{M(s(j-i)+Ki)+(K-s)d_X(x_i,x_j)^{\frac{1}{p}}}{mK}e_1+\frac{M\phi(i,j)}{m}e_2+\frac{M}{m}e_3, \tag{4}$$

$$Q_s^4(i,j) \stackrel{\mathrm{def}}{=} \frac{Mj}{m}e_1+\frac{M\phi(i,j)}{m}e_2+\frac{M(K-s)}{mK}e_3, \tag{5}$$

$$Q_s^5(i,j) \stackrel{\mathrm{def}}{=} \frac{Mj}{m}e_1+\frac{M(K-s)\phi(i,j)}{mK}e_2. \tag{6}$$

Then $Q_K^1(i,j)=Q_0^2(i,j)$, $Q_K^3(i,j)=Q_0^4(i,j)$ and $Q_K^4(i,j)=Q_0^5(i,j)$, so the total number of points thus obtained equals $5(K+1)-3=5K+2$.

Define $\mathcal{B}\subseteq\mathbb{R}^3$ by setting

$$\mathcal{B} \stackrel{\mathrm{def}}{=} \bigcup_{\substack{i,j\in\{1,\ldots,n\}\\ i<j}} \mathcal{B}_{ij}, \tag{7}$$

where for every $i,j\in\{1,\ldots,n\}$ with $i<j$ we write

$$\mathcal{B}_{ij} \stackrel{\mathrm{def}}{=} \bigcup_{s=0}^{K}\left\{Q_s^1(i,j),Q_s^2(i,j),Q_s^3(i,j),Q_s^4(i,j),Q_s^5(i,j)\right\}. \tag{8}$$

Hence $|\mathcal{B}_{ij}|=5K+2$. We also define $\mathcal{C}\subseteq\mathbb{R}^3$ by

$$\mathcal{C} \stackrel{\mathrm{def}}{=} \mathcal{B}\smallsetminus\left\{\frac{Mi}{m}e_1:\ i\in\{1,\ldots,n\}\right\}. \tag{9}$$

Note that by (2) we have $(Mi/m)e_1=Q_0^1(i,j)$ if $i,j\in\{1,\ldots,n\}$ satisfy $i<j$, and by (6) we have $(Mi/m)e_1=Q_K^5(\ell,i)$ if $\ell,i\in\{1,\ldots,n\}$ satisfy $\ell<i$. Thus $\mathcal{C}$ corresponds to removing from $\mathcal{B}$ those points that lie on the x-axis. In what follows, we denote $N=|\mathcal{C}|+1$. Finally, for every $i\in\{1,\ldots,n\}$ we define $\mathcal{C}_i\subseteq\mathbb{R}^3$ by

$$\mathcal{C}_i \stackrel{\mathrm{def}}{=} \mathcal{C}\cup\left\{\frac{Mi}{m}e_1\right\}. \tag{10}$$

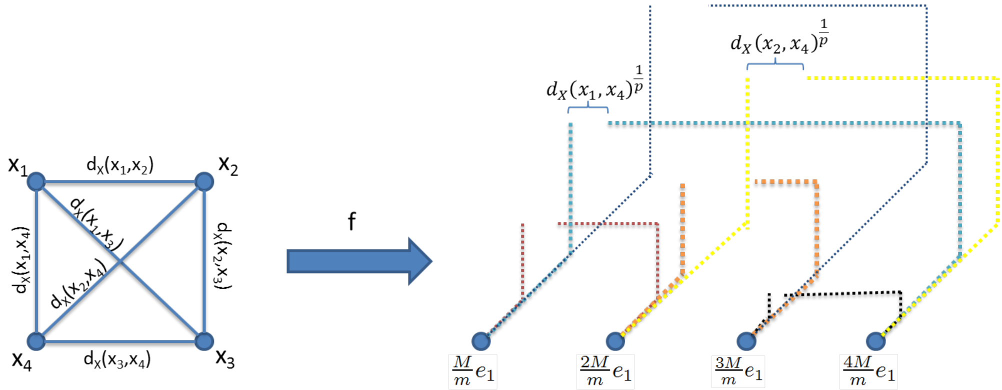

Figure 1: A schematic depiction of the embedding $f : X \to \mathcal{P}_p(\mathbb{R}^3)$ for a four-point metric space $(X, d_X) = (\{x_1, x_2, x_3, x_4\}, d_X)$. Here the x-axis is the horizontal direction, the z-axis is the vertical direction and the y-axis is perpendicular to the page plane. Recall that m and M are defined in (1).

Hence $|\mathcal{C}_i| = N$. Our embedding $f : X \to \mathcal{P}_p(\mathbb{R}^3)$ will be given by

$$\forall\, j \in \{1, \ldots, n\}, \qquad f(x_j) \stackrel{\mathrm{def}}{=} \frac{1}{N} \sum_{u \in \mathcal{C}_j} \delta_u, \tag{11}$$

where, as usual, δ_u is the point mass at u. Thus $f(x_j)$ is the uniform probability measure over $\mathcal{C}_j$. A schematic depiction of the above construction appears in Figure 1 below.

Lemma 7 below estimates the distortion of f, proving Theorem 5.

Lemma 7. *Fix $\varepsilon \in (0,1)$ and $p \in (1,\infty)$. Let $f : X \to \mathcal{P}_p(\mathbb{R}^3)$ be the mapping appearing in (11), considered as a mapping from the snowflaked metric space $(X, d_X^{1/p})$ to the metric space $(\mathcal{P}_p(\mathbb{R}^3), \mathsf{W}_p)$. Then, recalling the definitions of m and M in (1), we have*

$$K \geqslant \left(\frac{5M^p n^{2p}}{pm^p\varepsilon}\right)^{\frac{1}{p-1}} \implies \mathbf{dist}(f) \leqslant 1 + \varepsilon. \tag{12}$$

Proof. We shall show that under the assumption on K that appears in (12) we have

$$\forall\, i,j \in \{1, \ldots, n\}, \qquad \left(\frac{d_X(x_i,x_j)}{m^p N}\right)^{\frac{1}{p}} \leqslant \mathsf{W}_p(f(x_i), f(x_j)) \leqslant (1+\varepsilon)\left(\frac{d_X(x_i,x_j)}{m^p N}\right)^{\frac{1}{p}}, \tag{13}$$

where we recall that we defined N to be equal to $|\mathcal{C}| + 1$ for $\mathcal{C}$ given in (9). Clearly (13) implies that $\mathbf{dist}(f) \leqslant 1 + \varepsilon$, as required.

To prove the right hand inequality in (13), suppose that $i, j \in \{1, \ldots, n\}$ satisfy $i < j$ and consider the coupling $\pi \in \Pi(f(x_i), f(x_j))$ given by

$$\pi \stackrel{\mathrm{def}}{=} \frac{1}{N}\left(\sum_{t=1}^{5}\sum_{s=0}^{K-1} \delta_{\left(Q_s^t(i,j), Q_{s+1}^t(i,j)\right)} + \delta_{\left(Q_K^2(i,j), Q_0^3(i,j)\right)} + \sum_{u \in \mathcal{C} \smallsetminus \mathcal{B}_{ij}} \delta_{(u,u)}\right), \tag{14}$$

where for (14) recall (8) and (9). The meaning of (14) is simple: the supports of $f(x_i)$ and $f(x_j)$ equal $\mathcal{C}_i$ and $\mathcal{C}_j$, respectively, where we recall (10). Note that $\mathcal{C}_i \smallsetminus \mathcal{C}_j = \{Q_0^1(i,j)\}$ and $\mathcal{C}_j \smallsetminus \mathcal{C}_i = \{Q_K^5(i,j)\}$, where we recall (2) and (6). So, the coupling π in (14) corresponds to shifting the points in $\mathcal{B}_{ij}$ from the support of $f(x_i)$ to the support of $f(x_j)$ while keeping the points in $\mathcal{C} \smallsetminus \mathcal{B}_{ij}$ unchanged.

Now, recalling the definitions (2), (3), (4), (5) and (6),

$$\begin{aligned}\mathsf{W}_p(f(x_i), f(x_j))^p &\leqslant \iint_{\mathbb{R}^3\times\mathbb{R}^3} \|x-y\|_2^p \mathrm{d}\pi(x,y)\\ &= \frac{1}{N}\sum_{t=1}^{5}\sum_{s=0}^{K-1} \left\|Q_s^t(i,j) - Q_{s+1}^t(i,j)\right\|_2^p + \frac{\|Q_K^2(i,j)-Q_0^3(i,j)\|_2^p}{N}.\end{aligned} \tag{15}$$

Note that if $s \in \{0,\ldots,K-1\}$ then by (2), (3), (5), (6) we have

$$\begin{aligned} t\in\{1,5\} &\implies \left\|Q_s^t(i,j) - Q_{s+1}^t(i,j)\right\|_2 = \frac{M\phi(i,j)}{mK} \leqslant \frac{Mn^2}{mK},\\ t\in\{2,4\} &\implies \left\|Q_s^t(i,j) - Q_{s+1}^t(i,j)\right\|_2 = \frac{M}{mK}.\end{aligned} \tag{16}$$

Also, by (3) and (4) we have

$$\left\|Q_K^2(i,j) - Q_0^3(i,j)\right\|_2 = \frac{d_X(x_i,x_j)^{\frac{1}{p}}}{m}. \tag{17}$$

Finally, by (4) for every $s \in \{0,\ldots,K-1\}$ we have

$$\left\|Q_s^3(i,j) - Q_{s+1}^3(i,j)\right\|_2 = \frac{M(j-i)}{mK} - \frac{d_X(x_i,x_j)^{\frac{1}{p}}}{mK} \leqslant \frac{Mn}{mK}, \tag{18}$$

where we used the fact that $M(j-i) - d_X(x_i,x_j)^{1/p} \geqslant 0$, which holds true by the definition of M in (1) because $j-i \geqslant 1$. A substitution of (16), (17) and (18) into (15) yields the estimate

$$\begin{aligned}\mathsf{W}_p(f(x_i), f(x_j))^p &\leqslant \frac{d_X(x_i,x_j)}{m^pN} + \frac{5K}{N}\left(\frac{Mn^2}{mK}\right)^p\\ &= \left(1 + \frac{5M^pn^{2p}}{K^{p-1}d_X(x_i,x_j)}\right)\frac{d_X(x_i,x_j)}{m^pN} \leqslant (1+p\varepsilon)\frac{d_X(x_i,x_j)}{m^pN},\end{aligned}$$

where we used the fact that by the definition of m in (1) we have $m^p \leqslant d_X(x_i,x_j)$, and the lower bound on K that is assumed in (12). This implies the right hand inequality in (13) because $1+p\varepsilon \leqslant (1+\varepsilon)^p$.

Passing now to the proof of the left hand inequality in (13), we need to prove that for every $i,j \in \{1,\ldots,n\}$ with $i<j$ we have

$$\forall\, \pi \in \Pi(f(x_i), f(x_j)), \qquad \iint_{\mathbb{R}^3\times\mathbb{R}^3} \|x-y\|_2^p \mathrm{d}\pi(x,y) \geqslant \frac{d_X(x_i,x_j)}{m^pN}. \tag{19}$$

Note that we still did not use the triangle inequality for d_X, but this will be used in the proof of (19). Also, the reason why we are dealing with $\mathcal{P}_p(\mathbb{R}^3)$ rather than $\mathcal{P}_p(\mathbb{R}^2)$ will become clear in the ensuing argument.

Recall that the measures $f(x_i)$ and $f(x_j)$ are uniformly distributed over sets of the same size, and their supports $\mathcal{C}_i$ and $\mathcal{C}_j$ (respectively) satisfy $\mathcal{C}_i \triangle \mathcal{C}_j = \{(Mi/m)e_1, (Mj/m)e_1\}$. Since the set of all doubly stochastic matrices is the convex hull of the permutation matrices, and every permutation is a product of disjoint cycles, it follows that it suffices to establish the validity of (19) when $\pi = \frac{1}{N}\sum_{\ell=1}^{L} \delta_{(u_{\ell-1},u_\ell)}$ for some $L \in \{1,\ldots,n\}$ and $u_1,\ldots u_{L-1} \in \mathcal{C}$, where we set $u_0 = (Mi/m)e_1$ and $u_L = (Mj/m)e_1$. With this notation, our goal is to show that

$$\frac{1}{N}\sum_{\ell=1}^{L} \|u_\ell - u_{\ell-1}\|_2^p \geqslant \frac{d_X(x_i,x_j)}{m^p N}. \tag{20}$$

For every $a \in \{1,\ldots,n\}$ define $\mathcal{S}_a \subseteq \mathbb{R}^3$ by $\mathcal{S}_a \stackrel{\mathrm{def}}{=} \mathcal{S}_a^1 \cup \mathcal{S}_a^2$, where

$$\mathcal{S}_a^1 \stackrel{\mathrm{def}}{=} \bigcup_{b=a+1}^{n} \bigcup_{s=0}^{K} \left\{Q_s^1(a,b), Q_s^2(a,b)\right\}, \tag{21}$$

and

$$\mathcal{S}_a^2 \stackrel{\mathrm{def}}{=} \bigcup_{c=1}^{a-1} \bigcup_{s=0}^{K} \left\{Q_s^3(c,a), Q_s^4(c,a), Q_s^5(c,a)\right\}. \tag{22}$$

Thus, recalling (7), the sets $\mathcal{S}_1,\ldots,\mathcal{S}_n$ form a partition of $\mathcal{B}$ and $a \in \mathcal{S}_a$ for every $a \in \{1,\ldots,n\}$. For every $\ell \in \{0,\ldots,L\}$ let $a(\ell)$ be the unique element of $\{1,\ldots,n\}$ for which $u_\ell \in \mathcal{S}_{a(\ell)}$. Then $a(0) = i$ and $a(L) = j$. The left hand side of (20) can be bounded from below as follows

$$\frac{1}{N}\sum_{\ell=1}^{L} \|u_\ell - u_{\ell-1}\|_2^p \geqslant \frac{1}{N}\sum_{\ell=1}^{L} \min_{\substack{u\in\mathcal{S}_{a(\ell-1)}\\ v\in\mathcal{S}_{a(\ell)}}} \|u-v\|_2^p. \tag{23}$$

We shall show that

$$\forall\, a,b \in \{1,\ldots,n\},\ \forall (u,v) \in \mathcal{S}_a \times \mathcal{S}_b, \qquad \|u-v\|_2^p \geqslant \frac{d_X(x_a,x_b)}{m^p}. \tag{24}$$

The validity of (24) implies the required estimate (20) because, by (23), it follows from (24) and the triangle inequality for d_X that

$$\frac{1}{N}\sum_{\ell=1}^{L} \|u_\ell - u_{\ell-1}\|_2^p \geqslant \frac{1}{N}\sum_{\ell=1}^{L} \frac{d_X\big(x_{a(\ell-1)}, x_{a(\ell)}\big)}{m^p} \geqslant \frac{d_X(x_i,x_j)}{m^p N}.$$

It remains to justify (24). Suppose that $a,b \in \{1,\ldots,n\}$ satisfy $a<b$ and $(u,v) \in \mathcal{S}_a \times \mathcal{S}_b$. Write $u = Q_s^t(c,d)$ and $v = Q_\sigma^\tau(\gamma,\delta)$ for some $s,\sigma \in \{0,\ldots,K\}$, $t,\tau \in \{1,\ldots,5\}$ and $c,d,\gamma,\delta \in \{1,\ldots,n\}$.

We shall check below, via a direct case analysis, that the absolute value of one of the three coordinates of $u-v$ is either at least M/m or at least $d_X(x_a,x_b)^{1/p}/m$. Since by the definition of M in (1) we have $M \geqslant d_X(x_a,x_b)^{1/p}$, this assertion will imply (24).

Suppose first that $t,\tau \in \{1,2,4,5\}$. By comparing (21), (22) with (2), (3), (4), (5) we see that $\langle u, e_1\rangle = Ma/m$ and $\langle v, e_1\rangle = Mb/m$. Since $b-a \geqslant 1$, this implies that $\langle u-v, e_1\rangle \geqslant M/m$, as required.

If $t=\tau=3$ then by (22) we necessarily have $d=a$ and $\delta=b$. Hence $(c,d)\neq(\gamma,\delta)$ and therefore $|\phi(c,d)-\phi(\gamma,\delta)| \geqslant 1$, since ϕ is a bijection between $\{1,\ldots,n\}\times\{1,\ldots,n\}$ and $\{1,\ldots,n^2\}$. By (4) we therefore have $|\langle u-v, e_2\rangle| \geqslant M/m$, as required.

It remains to treat the case $t\neq\tau$ and $3\in\{t,\tau\}$. If $\{t,\tau\}\subseteq\{1,3,5\}$ then by contrasting (4) with (2) and (6) we see that the third coordinate of one of the vectors u,v vanishes while the third coordinate of the other vector equals M/m. Therefore $|\langle u-v,e_3\rangle| \geqslant M/m$, as required. The only remaining case is $\{t,\tau\}\subseteq\{2,3,4\}$. In this case $|\langle u-v,e_2\rangle| = M|\phi(c,d)-\phi(\gamma,\delta)|/m$, by (4), (3), (5). So, if $(c,d)\neq(\gamma,\delta)$ then $|\phi(c,d)-\phi(\gamma,\delta)|\geqslant 1$, and we are done. We may therefore assume that $c=\gamma$ and $d=\delta$. Observe that by (22) if $\{t,\tau\}=\{3,4\}$ then $\{d,\delta\}=\{a,b\}$, which contradicts $d=\delta$. So, we also necessarily have $\{t,\tau\}=\{2,3\}$, in which case, since $a<b$, by (21) and (22) we see that $c=\gamma=a$ and $d=\delta=b$. By interchanging the labels s and σ if necessary, we may assume that $u=Q^2_\sigma(a,b)$ and $v=Q^3_s(a,b)$. By (3) and (4) we therefore have

$$\langle v-u, e_1\rangle = \frac{M(s(b-a)+Ka)}{mK} + \frac{(K-s)d_X(x_a,x_b)^{\frac{1}{p}}}{mK} - \frac{Ma}{m}$$
$$= \frac{d_X(x_a,x_b)^{\frac{1}{p}}}{m} + \frac{sM(b-a)-sd_X(x_a,x_b)^{\frac{1}{p}}}{mK} \geqslant \frac{d_X(x_a,x_b)^{\frac{1}{p}}}{m},$$

where we used the fact that by (1) we have $M\geqslant d_X(x_a,x_b)^{1/p}$, and that $b-a\geqslant 1$. This concludes the verification of the remaining case of (24), and hence the proof of Lemma 7 is complete. □

2.1 Implications: Theorems 2 and 4

Theorem 2 follows from the fact that the shortest path metric on an expander graph on N nodes has $\Omega(\log N)$ distortion lower bound for embedding it into L_1 [36]. Note that in the proof above we obtain measures supported on n points where $n \leqslant N^{O(1)}\cdot\left(\frac{5M^pN^{2p}}{pm^p}\right)^{\frac{1}{p-1}}$ for a $1+\varepsilon=2$ approximation. Hence, any embedding of W_p on $\mathbb{R}^3$ pointsets of size n into L_1 has a distortion lower bound of $\Omega((\log N)^{1/p}) = \Omega(((p-1)\log n)^{1/p})$.

Similarly, Theorem 4 follows by considering X to be the N-point subset of $(\mathcal{P}_1(\{0,1\}^{O(\log N)}), \mathsf{W}_1)$ introduced in [33, Section 3]. Any sketching algorithm for this metric X requires $\Omega(\frac{\log N}{s})$ approximation for sketching complexity s [5, Theorem 4.1]. Since we can embed X into the square of W_2 with constant distortion, we obtain a $\Omega\left(\left(\frac{(p-1)\log n}{s}\right)^{1/p}\right)$ approximation lower bound for any W_p sketch with sketching complexity s.

3 Future Directions

As discussed in the Introduction, it seems plausible that Theorem 5 and Theorem 6 are not sharp when $p \in (2,\infty)$. Specifically, we conjecture that there exist $D_p \in [1,\infty)$ such that for every finite metric space (X, d_X) we have

$$c_{\mathcal{P}_p(\mathbb{R}^3)}\Big(X, \sqrt{d_X}\Big) \leqslant D_p. \tag{25}$$

Perhaps (25) even holds true with $D_p = 1$. Since L_2 admits an isometric embedding into L_p (see e.g. [75]), the perceived analogy between Wasserstein p spaces and L_p spaces makes it natural to ask whether or not $(\mathcal{P}_2(\mathbb{R}^3), \mathsf{W}_2)$ admits a bi-Lipschitz embedding into $(\mathcal{P}_p(\mathbb{R}^3), \mathsf{W}_p)$. If the answer to this question were positive then (25) would hold true by virtue of the case $p = 2$ of Theorem 5. We also conjecture that the lower bound of Theorem 6 could be improved when $p > 2$ to state that for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space (Y, d_Y) such that for every $\alpha \in (1/2, 1]$,

$$c_{(\mathcal{P}_p(\mathbb{R}^3),\mathsf{W}_p)}(Y, d_Y^\alpha) \gtrsim_p (\log n)^{\alpha-\frac{1}{2}}. \tag{26}$$

It was shown in [48] that L_p has Markov type 2 for every $p \in (2,\infty)$. We therefore ask whether or not $(\mathcal{P}_p(\mathbb{R}^3), \mathsf{W}_p)$ has Markov type 2 for every $p \in (2,\infty)$. A positive answer to this question would imply that the lower bound (26) is indeed achievable. For this purpose it would also suffice to show that for every $p \in (2,\infty)$ and $k \in \mathbb{N}$ we have

$$M_p((\mathcal{P}_p(\mathbb{R}^3), \mathsf{W}_p); 2^k) \lesssim_p 2^{k\left(\frac{1}{2}-\frac{1}{p}\right)}. \tag{27}$$

Proving (27) may be easier than proving that $M_2(\mathcal{P}_p(\mathbb{R}^3), \mathsf{W}_p) < \infty$, since the former involves arguing about the pth powers of Wasserstein p distances while the latter involves arguing about Wasserstein p distances squared. Note that $M_p(L_p; m) \lesssim \sqrt{p} m^{1/2-1/p}$ by [48] (see also [45, Theorem 4.3]), so the L_p-version of (27) is indeed valid.

Another natural direction to pursue concerns with the distortion of embedding finite metric spaces into Wasserstein spaces.

Question 1. Is it true that for $p \in (1, 2]$ and $n \in \mathbb{N}$ every n-point metric space (X, d_X) satisfies

$$c_{\mathcal{P}_p(\mathbb{R}^3)}(X) \lesssim_p (\log n)^{1-\frac{1}{p}}?$$

A positive answer to Question (1) would resolve the *metric cotype dichotomy problem* [39] (see the full version for more details). We believe that Question 1 is an especially intriguing challenge in embedding theory (for a concrete and natural target space) because a positive answer would require an interesting new construction, and a negative answer would require devising a new bi-Lipschitz invariant that would serve as an obstruction for embeddings into Wasserstein spaces.

Focusing for concreteness on the case $p = 2$, Question 1 asks whether $c_{\mathcal{P}_2(\mathbb{R}^3)}(X) \lesssim \sqrt{\log n}$ for every n-point metric space (X, d_X). Note that Theorem 5 implies that (X, d_X) embeds into $\mathcal{P}_2(X)$ with distortion at most the square root of the *aspect ratio* of (X, d_X), i.e.,

$$c_{(\mathcal{P}_2(\mathbb{R}^3),\mathsf{W}_2)}(X, d_X) \leqslant \sqrt{\frac{\operatorname{diam}(X, d_X)}{\min_{\substack{x,y\in X\\ x\neq y}} d_X(x, y)}}, \tag{28}$$

but we are asking here for the largest possible growth rate of the distortion of X into $\mathcal{P}_2(X)$ in terms of the cardinality of X. While for certain embedding results there are standard methods (see e.g. [12, 30, 40]) for replacing the dependence on the aspect ratio of a finite metric space by a dependence on its cardinality, these methods do not seem to apply to our embedding in (28). See the full version for further discussion.

References

[1] Pankaj Agarwal and Kasturi Varadarajan. A near-linear constant-factor approximation for euclidean bipartite matching? In *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, SCG '04, pages 247–252, New York, NY, USA, 2004. ACM.

[2] Pankaj K. Agarwal, Alon Efrat, and Micha Sharir. Vertical decomposition of shallow levels in 3-dimensional arrangements and its applications. *SIAM Journal on Computing*, 29(3):912–953, 2000.

[3] Pankaj K. Agarwal and R. Sharathkumar. Approximation algorithms for bipartite matching with metric and geometric costs. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 555–564, New York, NY, USA, 2014. ACM.

[4] Alexandr Andoni, Khanh Do Ba, Piotr Indyk, and David Woodruff. Efficient sketches for Earth-Mover Distance, with applications. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2009.

[5] Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Earth mover distance over high-dimensional spaces. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 343–352, 2008. Previously ECCC Report TR07-048.

[6] Alexandr Andoni, Robert Krauthgamer, and Ilya Razenshteyn. Sketching and embedding are equivalent for norms. In *Proceedings of the Symposium on Theory of Computing (STOC)*, 2015. Full version at `http://arxiv.org/abs/1411.2577`.

[7] Alexandr Andoni, Aleksandar Nikolov, Krzysztof Onak, and Grigory Yaroslavtsev. Parallel algorithms for geometric graph problems. In *Proceedings of the Symposium on Theory of Computing (STOC)*, 2014. Full version at `http://arxiv.org/abs/1401.0042`.

[8] T. Austin and A. Naor. On the bi-Lipschitz structure of Wasserstein spaces. In preparation, 2016.

[9] Tim Austin, Assaf Naor, and Yuval Peres. The wreath product of $\mathbb{Z}$ with $\mathbb{Z}$ has Hilbert compression exponent $\frac{2}{3}$. *Proc. Amer. Math. Soc.*, 137(1):85–90, 2009.

[10] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[11] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. Previously in FOCS'02.

[12] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[13] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.

[14] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[15] Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. In *ACM Transactions on Graphics (TOG)*, volume 30, page 158. ACM, 2011.

[16] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[17] Moses Charikar. Similarity estimation techniques from rounding. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 380–388, 2002.

[18] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *Proceedings of The 31st International Conference on Machine Learning*, pages 685–693, 2014.

[19] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Mechanism design via optimal transport. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 269–286, New York, NY, USA, 2013. ACM.

[20] Guy David and Stephen Semmes. *Fractured fractals and broken dreams*, volume 7 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1997. Self-similar geometry through metric and measure.

[21] Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. Blue noise through optimal transport. *ACM Transactions on Graphics (TOG)*, 31(6):171, 2012.

[22] Fernando De Goes, David Cohen-Steiner, Pierre Alliez, and Mathieu Desbrun. An optimal transport approach to robust reconstruction and simplification of 2d shapes. In *Computer Graphics Forum*, volume 30, pages 1593–1602. Wiley Online Library, 2011.

[23] Jian Ding, James R. Lee, and Yuval Peres. Markov type and threshold embeddings. *Geom. Funct. Anal.*, 23(4):1207–1229, 2013.

[24] P. Enflo. Uniform homeomorphisms between Banach spaces. In *Séminaire Maurey-Schwartz (1975–1976), Espaces, L^p, applications radonifiantes et géométrie des espaces de Banach, Exp. No. 18*, page 7. Centre Math., École Polytech., Palaiseau, 1976.

[25] Norm Ferns, Pablo Samuel Castro, Doina Precup, and Prakash Panangaden. Methods for computing state similarity in markov decision processes. In *UAI '06, Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence, Cambridge, MA, USA, July 13-16, 2006*, 2006.

[26] Thomas Foertsch and Viktor Schroeder. Hyperbolicity, CAT(-1)-spaces and the Ptolemy inequality. *Math. Ann.*, 350(2):339–356, 2011.

[27] D. J. H. Garling. *Inequalities: a journey into linear analysis*. Cambridge University Press, Cambridge, 2007.

[28] Kristen Grauman and Trevor Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Beijing, China, October 2005.

[29] Kristen Grauman and Trevor Darrell. Approximate correspondences in high dimensions. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2006.

[30] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput.*, 35(5):1148–1184 (electronic), 2006.

[31] Piotr Indyk. A near linear time constant factor approximation for euclidean bichromatic matching (cost). In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007.

[32] Piotr Indyk and Nitin Thaper. Fast color image retrieval via embeddings. *Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[33] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006.

[34] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474, 2000. Preliminary version appeared in STOC'98.

[35] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[36] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[37] N. Linial, A. Magen, and A. Naor. Girth and Euclidean distortion. *Geom. Funct. Anal.*, 12(2):380–394, 2002.

[38] Jiří Matoušek and Assaf Naor. Open problems on embeddings of finite metric spaces. August 2011. Available at `http://kam.mff.cuni.cz/~matousek/metrop.ps`.

[39] Manor Mendel and Assaf Naor. Metric cotype. *Ann. of Math. (2)*, 168(1):247–298, 2008.

[40] Manor Mendel and Assaf Naor. Maximum gradient embeddings and monotone clustering. *Combinatorica*, 30(5):581–615, 2010.

[41] David M. Mount, Nathan S. Netanyahu, and San Ratanasanya. New approaches to robust, point-based image registration. In Jacqueline Le Moigne, Nathan S. Netanyahu, and Roger D. Eastman, editors, *Image Registration for Remote Sensing*, pages 179–199. Cambridge University Press, 2011. Cambridge Books Online.

[42] Patrick Mullen, Pooran Memari, Fernando de Goes, and Mathieu Desbrun. Hot: Hodge-optimized triangulations. In *ACM Transactions on Graphics (TOG)*, volume 30, page 103. ACM, 2011.

[43] Assaf Naor. A phase transition phenomenon between the isometric and isomorphic extension problems for Hölder functions between L_p spaces. *Mathematika*, 48(1-2):253–271 (2003), 2001.

[44] Assaf Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[45] Assaf Naor. Comparison of metric spectral gaps. *Anal. Geom. Metr. Spaces*, 2:1–52, 2014.

[46] Assaf Naor and Yuval Peres. Embeddings of discrete groups and the speed of random walks. *Int. Math. Res. Not. IMRN*, pages Art. ID rnn 076, 34, 2008.

[47] Assaf Naor and Yuval Peres. L_p compression, traveling salesmen, and stable walks. *Duke Math. J.*, 157(1):53–108, 2011.

[48] Assaf Naor, Yuval Peres, Oded Schramm, and Scott Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.

[49] Assaf Naor and Gideon Schechtman. Remarks on non linear type and Pisier's inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.

[50] Assaf Naor and Gideon Schechtman. Planar earthmover is not in L_1. *SIAM J. Comput. (SICOMP)*, 37(3):804–826, 2007. An extended abstract appeared in FOCS'06.

[51] Kangyu Ni, Xavier Bresson, Tony F. Chan, and Selim Esedoglu. Local histogram based segmentation using the wasserstein distance. *International Journal of Computer Vision*, 84(1):97–111, 2009.

[52] Shin-Ichi Ohta. Markov type of Alexandrov spaces of non-negative curvature. *Mathematika*, 55(1-2):177–189, 2009.

[53] Yann Ollivier, Herve Pajot, and Cedric Villani. *Optimal Transport, Theory and Applications*. Cambridge University Press, New York, NY, USA, 2014.

[54] List of open problems in sublinear algorithms: Problem 7. `http://sublinear.info/7`.

[55] List of open problems in sublinear algorithms: Problem 49. `http://sublinear.info/49`.

[56] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009*, pages 460–467, 2009.

[57] Jeff M. Phillips and Pankaj K. Agarwal. On bipartite matching under the RMS distance. In *Proceedings of the 18th Annual Canadian Conference on Computational Geometry, CCCG 2006, August 14-16, 2006, Queen's University, Ontario, Canada*, 2006.

[58] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[59] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. A metric for distributions with applications to image databases. In *ICCV*, pages 59–66, 1998.

[60] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, 40(2):99–121, November 2000.

[61] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[62] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 360–369, 2002.

[63] I. J. Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.

[64] R. Sharathkumar and Pankaj K. Agarwal. Algorithms for the transportation problem in geometric settings. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 306–317, 2012.

[65] R. Sharathkumar and Pankaj K. Agarwal. A near-linear time -approximation algorithm for geometric bipartite matching. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 385–394, 2012.

[66] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4):66:1–66:11, July 2015. In SIGGRAPH'15.

[67] Justin Solomon, Leonidas Guibas, and Adrian Butscher. Dirichlet energy for analysis and synthesis of soft maps. *Computer Graphics Forum*, 32(5):197–206, 2013. In Eurographics'13.

[68] Justin Solomon, Raif Rustamov, Leonidas Guibas, and Adrian Butscher. Earth mover's distances on discrete surfaces. *ACM Trans. Graph.*, 33(4):67:1–67:12, July 2014.

[69] Justin Solomon, Raif Rustamov, Leonidas Guibas, and Adrian Butscher. Wasserstein propagation for semi-supervised learning. In *Proceedings of The 31st International Conference on Machine Learning*, pages 306–314, 2014.

[70] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[71] Pravin M. Vaidya. Geometry helps in matching. *SIAM Journal on Computing*, 18(6):1201–1225, 1989.

[72] Kasturi R. Varadarajan and Pankaj K. Agarwal. Approximation algorithms for bipartite and non-bipartite matching in the plane. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, pages 805–814. Society for Industrial and Applied Mathematics, 1999.

[73] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[74] Michael Werman, Shmuel Peleg, and Azriel Rosenfeld. A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing*, 32(3):328 – 336, 1985.

[75] P. Wojtaszczyk. *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1991.

A Sharpness of Theorem 5

Here we prove Theorem 6, that shows the sharpness of Theorem 5 whenever $p \in (1,2]$, and in addition show a nontrivial, though probably non-sharp, restriction on the embeddability of snowflakes into $\mathcal{P}_p(\mathbb{R}^3)$ also for $p \in (2,\infty)$.

The results of this section rely crucially on K. Ball's notion [10] of Markov type. We shall start by recalling the relevant background on this important invariant of metric spaces, including variants and notation from [45] that will be used below. Let $\{Z_t\}_{t=0}^\infty$ be a Markov chain on the state space $\{1,\ldots,n\}$ with transition probabilities $a_{ij} = \Pr[Z_{t+1} = j | Z_t = i]$ for every $i,j \in \{1,\ldots,n\}$. $\{Z_t\}_{t=0}^\infty$ is said to be stationary if $\pi_i = \Pr[Z_t = i]$ does not depend on $t \in \{1,\ldots,n\}$ and it is said to be reversible if $\pi_i a_{ij} = \pi_j a_{ji}$ for every $i,j \in \{1,\ldots,n\}$.

Let $\{Z'_t\}_{t=0}^\infty$ be the Markov chain that starts at Z_0 and then evolves independently of $\{Z_t\}_{t=0}^\infty$ with the same transition probabilities. Thus $Z'_0 = Z_0$ and conditioned on Z_0 the random variables Z_t and Z'_t are independent and identically distributed. We note for future use that if $\{Z_t\}_{t=0}^\infty$ as above is stationary and reversible then for every symmetric function $\psi : \{1,\ldots,n\} \times \{1,\ldots,n\} \to \mathbb{R}$ and every $t \in \mathbb{N}$ we have

$$\mathbb{E}\big[\psi(Z_t, Z'_t)\big] = \mathbb{E}\big[\psi(Z_{2t}, Z_0)\big]. \tag{29}$$

This is a consequence of the observation that, by stationarity and revesibility, conditioned on the random variable Z_t the random variables Z_0 and Z_{2t} are independent and identically distributed. Denoting $A = (a_{ij}) \in M_n(\mathbb{R})$, the validity of (29) can be alternatively checked directly as follows.

$$\begin{aligned}\mathbb{E}\left[\psi(Z_t, Z'_t)\right] = \mathbb{E}\big[\mathbb{E}\left[\psi(Z_t, Z'_t)|Z_0\right]\big] &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \pi_i A^t_{ij} A^t_{ik} \psi(j,k) \\ &\overset{(\star)}{=} \sum_{j=1}^n \sum_{k=1}^n \pi_j \left(\sum_{i=1}^n A^t_{ji} A^t_{ik} \right) \psi(j,k) = \sum_{j=1}^n \sum_{k=1}^n \pi_j A^{2t}_{jk} \psi(j,k), \end{aligned} \tag{30}$$

where $(\star)$ uses the reversibility of the Markov chain $\{Z_t\}_{t=0}^\infty$ through the validity of $\pi_i A^t_{ij} = \pi_j A^t_{ji}$ for every $i,j \in \{1,\ldots,n\}$. The final term in (30) equals the right hand side of (29), as required.

Given $p \in [1,\infty)$, a metric space (X, d_X) and $m \in \mathbb{N}$, the Markov type p constant of (X, d_X) at time m, denoted $M_p(X, d_X; m)$ (or simply $M_p(X; m)$ if the metric is clear from the context) is defined to be the infimum over those $M \in (0,\infty)$ such that for every $n \in \mathbb{N}$, every stationary reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ with state space $\{1,\ldots,n\}$, and every $f : \{1,\ldots,n\} \to X$ we have

$$\mathbb{E}\big[d_X(f(Z_m), f(Z_0))^p\big] \leqslant M^p m \mathbb{E}\big[d_X(f(Z_1), f(Z_0))^p\big].$$

Observe that by the triangle inequality we always have

$$M_p(X;m) \leqslant m^{1-\frac{1}{p}}.$$

As we shall explain below, any estimate of the form $M_p(X;m) \lesssim_X m^{\theta}$ for $\theta < 1 - 1/p$ is a nontrivial obstruction to the embeddability of certain metric spaces into X, but it is especially important (e.g. for Lipschitz extension theory [10]) to single out the case when $M_p(X;m) \lesssim_X 1$. Specifically, (X, d_X) is said to have Markov type p if

$$M_p(X, d_X) \stackrel{\mathrm{def}}{=} \sup_{m \in \mathbb{N}} M_p(X, d_X; m) < \infty.$$

$M_p(X, d_X)$ is called the Markov type p constant of (X, d_X), and it is often denoted simply $M_p(X)$ if the metric is clear from the context.

The Markov type of many important classes of metric spaces is satisfactorily understood, though some fundamental questions remain open; see Section 4 of the survey [44] and the references therein, as well as more recent progress in e.g. [23]. Here we study this notion in the context of Wasserstein spaces. The link of Markov type to the nonembeddability of snowflakes is simple, originating in an idea of [37]. This is the content of the following lemma.

Lemma 8. *Fix a metric space (Y, d_Y), $m \in \mathbb{N}$, $K, p \in [1,\infty)$ and $\theta \in [0,1]$. Suppose that*

$$M_p(Y;m) \leqslant K m^{\frac{\theta(p-1)}{p}}. \tag{31}$$

Denote $n = 2^{4m}$. Then there exists an n-point metric space (X, d_X) such that

$$\alpha \in \left[\frac{1+\theta(p-1)}{p}, 1\right] \implies c_Y(X, d_X^{\alpha}) \gtrsim \frac{1}{K}(\log n)^{\alpha - \frac{1+\theta(p-1)}{p}}.$$

Proof. Take $(X, d_X) = (\{0,1\}^{4m}, \|\cdot\|_1)$, i.e., X is the $4m$-dimensional discrete hypercube, equipped with the Hamming metric. Thus $|X| = n$. Let $\{Z_t\}_{t=0}^{\infty}$ be the standard random walk on X, with Z_0 distributed uniformly over X. Suppose that $f : X \to Y$ satisfies

$$\forall\, x, y \in X, \qquad s\|x-y\|_1^{\alpha} \leqslant d_Y(f(x), f(y)) \leqslant Ds\|x-y\|_1^{\alpha} \tag{32}$$

for some $s, D \in (0,\infty)$. Our goal is to bound D from below. By the definition of $M_p(Y;m)$,

$$\mathbb{E}\big[d_Y(f(Z_m), f(Z_0))^p\big] \stackrel{(31)}{\leqslant} K^p m^{1+\theta(p-1)} \mathbb{E}\big[d_Y(f(Z_1), f(Z_0))^p\big]. \tag{33}$$

By the right hand inequality in (32) we have

$$\mathbb{E}\big[d_Y(f(Z_1),f(Z_0))^p\big]\leqslant D^ps^p\mathbb{E}\big[\|Z_1-Z_0\|_1^{\alpha p}\big]=D^ps^p. \tag{34}$$

At the same time, it is simple to see (and explained explicitly in e.g. [49] or [44, Section 9.4]) that $\mathbb{E}\big[\|Z_m-Z_0\|_1^{\alpha p}\big]\geqslant(\eta m)^{\alpha p}$ for some universal constant $\eta\in(0,1)$. Hence,

$$\mathbb{E}\big[d_Y(f(Z_m),f(Z_0))^p\big]\overset{(32)}{\geqslant}s^p\mathbb{E}\big[\|Z_m-Z_0\|_1^{\alpha p}\big]\gtrsim s^p(\eta m)^{\alpha p}. \tag{35}$$

The only way for (34) and (35) to be compatible with (33) is if

$$D\gtrsim\frac{1}{K}m^{\alpha-\frac{1+\theta(p-1)}{p}}\asymp\frac{1}{K}(\log n)^{\alpha-\frac{1+\theta(p-1)}{p}}.$$

□

Remark 9. In Lemma 8 we took the metric space X to be a discrete hypercube, but similar conclusions apply to snowflakes of expander graphs and graphs with large girth [37], as well as their subsets [13] and certain discrete groups [9, 46, 47] (see also [44, Section 9.4]). We shall not attempt to state here the wider implications of the assumption (31) to the nonembeddability of snowflakes, since the various additional conclusions follow mutatis mutandis from the same argument as above, and Lemma 8 as currently stated suffices for the proof of Theorem 6.

Remark 10. Since the proof of Lemma 8 applied the Markov type p assumption (31) to the discrete hypercube, it would have sufficed to work here with a classical weaker bi-Lipschitz invariant due to Enflo [24], called Enflo type. Such an obstruction played a role in ruling out certain snowflake embeddings in [26] (in a different context), though the fact that the argument of [26] could be cast in the context of Enflo type was proved only later [52, Proposition 5.3]. Here we work with Markov type rather than Enflo type because the proof below for Wasserstein spaces yields this stronger conclusion without any additional effort.

The following lemma is a variant of [52, Lemma 4.1].

Lemma 11. *Fix $p\in[1,\infty)$ and $\theta\in[1/p,1]$. Suppose that (X,d_X) is a metric space such that for every two X-valued independent and identically distributed finitely supported random variables Z,Z' and every $x\in X$ we have*

$$\mathbb{E}\big[d_X(Z,Z')^p\big]\leqslant 2^{\theta p}\mathbb{E}\big[d_X(Z,x)^p\big]. \tag{36}$$

Then for every $k\in\mathbb{N}$ we have

$$M_p(X;2^k)\leqslant 2^{k\left(\theta-\frac{1}{p}\right)}. \tag{37}$$

Proof. Fix $n\in\mathbb{N}$, a stationary reversible Markov chain $\{Z_t\}_{t=0}^\infty$ with state space $\{1,\ldots,n\}$, and $f:\{1,\ldots,n\}\to X$. Recalling (29) with $\psi(i,j)=d_X(f(i),f(j))^p$, for every $t\in\mathbb{N}$ we have

$$\begin{aligned}\mathbb{E}\big[d_X(Z_{2t},Z_0)^p\big]\overset{(29)}{=}\mathbb{E}\big[d_X(Z_t,Z_t')^p\big]&\overset{(36)}{\leqslant}2^{\theta p}\mathbb{E}\big[d_X(Z_t,Z_0)^p\big]\\&\leqslant 2^{\theta p-1}M_p(X;t)^p\cdot 2t\mathbb{E}\big[d_X(Z_1,Z_0)^p\big],\end{aligned} \tag{38}$$

where the last step of (38) uses the definition of $M_p(X;t)$. By the definition of $M_p(X;2t)$, we have thus proved that

$$M_p(X;2t) \leqslant 2^{\theta-\frac{1}{p}} M_p(X;t),$$

so (37) follows by induction on k. □

Corollary 12 below follows from Lemma 8 and Lemma 11. Specifically, under the assumptions and notation of Lemma 11, use Lemma 8 with m replaced by 2^k and θ replaced by $(\theta p - 1)/(p-1)$.

Corollary 12. *Fix $p \in [1,\infty)$ and $\theta \in [1/p, 1]$. Suppose that (X, d_X) is a metric space that satisfies the assumptions of Lemma 11. Then for arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space (Y, d_Y) such that for every $\alpha \in [\theta, 1]$ we have*

$$c_X\left(Y, d_Y^\alpha\right) \gtrsim (\log n)^{\alpha-\theta}.$$

The link between the above discussion and embeddings of snowflakes of metrics into Wasserstein spaces is explained in the following lemma, which is a variant of [70, Proposition 2.10].

Lemma 13. *Fix $p \in [1,\infty)$ and $\theta \in [1/p, 1]$. Suppose that (X, d_X) is a metric space that satisfies the assumptions of Lemma 11, i.e., inequality (36) holds true for X-valued random variables. Then the same inequality holds true in the metric space $(\mathcal{P}_p(X), \mathsf{W}_p)$ as well, i.e., for every two $\mathcal{P}_p(X)$-valued and identically distributed finitely supported random variables $\mathfrak{M}, \mathfrak{M}'$ and every $\mu \in \mathcal{P}_p(X)$,*

$$\mathbb{E}\big[\mathsf{W}_p(\mathfrak{M}, \mathfrak{M}')^p\big] \leqslant 2^{\theta p}\mathbb{E}\big[\mathsf{W}_p(\mathfrak{M}, \mu)^p\big].$$

Proof. Suppose that the distribution of $\mathfrak{M}$ equals $\sum_{i=1}^n q_i \delta_{\mu_i}$ for some $\mu_1, \ldots, \mu_n \in \mathcal{P}_p(X)$ and $q_1, \ldots, q_n \in [0,1]$ with $\sum_{i=1}^n q_i = 1$. Our goal is to show that

$$\sum_{i=1}^n \sum_{j=1}^n q_i q_j \mathsf{W}_p(\mu_i, \mu_j)^p \leqslant 2^{\theta p} \sum_{i=1}^n q_i \mathsf{W}_p(\mu_i, \mu)^p. \tag{39}$$

The finitely supported probability measures are dense in $(\mathcal{P}_p(X), \mathsf{W}_p)$ (see [58, 73]), so it suffices to prove (39) when there exists $N \in \mathbb{N}$ and points $x_{ik}, x_k \in X$ for every $(i,k) \in \{1, \ldots, n\} \times \{1, \ldots, N\}$ such that we have $\mu = \frac{1}{N}\sum_{k=1}^N \delta_{x_k}$ and $\mu_i = \frac{1}{N}\sum_{k=1}^N \delta_{x_{ik}}$ for every $i \in \{1, \ldots, n\}$. Let $\{\sigma_i\}_{i=1}^N \subseteq S_N$ be permutations of $\{1, \ldots, N\}$ that induce optimal couplings of the pairs (μ, μ_i), i.e.,

$$\forall\, i \in \{1, \ldots, n\}, \qquad \mathsf{W}_p(\mu_i, \mu)^p = \frac{1}{N}\sum_{k=1}^N d_X(x_{i\sigma_i(k)}, x_k)^p. \tag{40}$$

Since the measure $\frac{1}{N}\sum_{k=1}^N \delta_{(x_{i\sigma_i(k)}, x_{j\sigma_j(k)})}$ is a coupling of (μ_i, μ_j),

$$\forall\, i,j \in \{1, \ldots, n\}, \qquad \mathsf{W}_p(\mu_i, \mu_j)^p \leqslant \frac{1}{N}\sum_{k=1}^N d_X(x_{i\sigma_i(k)}, x_{j\sigma_j(k)})^p. \tag{41}$$

Consequently,

$$\sum_{i=1}^n\sum_{j=1}^n q_iq_j\mathsf{W}_p(\mu_i,\mu_j)^p \overset{(41)}{\leqslant} \frac{1}{N}\sum_{k=1}^N\sum_{i=1}^n\sum_{j=1}^n q_iq_jd_X(x_{i\sigma_i(k)},x_{j\sigma_j(k)})^p$$
$$\overset{(36)}{\leqslant} \frac{2^{\theta p}}{N}\sum_{k=1}^N\sum_{i=1}^n\sum_{j=1}^n q_iq_jd_X(x_{i\sigma_i(k)},x_k)^p \overset{(40)}{=} 2^{\theta p}\sum_{i=1}^n q_i\mathsf{W}_p(\mu_i,\mu)^p. \qquad \square$$

Proof of Theorem 6. Let (Ω,μ) be a probability space. For $p\in[1,\infty]$ define $T: L_p(\mu)\to L_p(\mu\times\mu)$ by $Tf(x,y)=f(x)-f(y)$. Then clearly $\|T\|_{L_p(\mu)\to L_p(\mu\times\mu)}\leqslant 2$ for $p\in\{1,\infty\}$ and

$$\forall\, f\in L_2(\mu),\qquad \|Tf\|^2_{L_2(\mu\times\mu)}=2\|f\|^2_{L_2(\mu)}-2\Big(\int_\Omega f\mathrm{d}\mu\Big)^2\leqslant 2\|f\|^2_{L_2(\mu)}.$$

Or $\|T\|_{L_2(\mu)\to L_2(\mu\times\mu)}\leqslant\sqrt{2}$. So, by the Riesz–Thorin theorem (e.g. [27]),

$$p\in[1,2]\implies \|T\|_{L_p(\mu)\to L_p(\mu\times\mu)}\leqslant 2^{\frac{1}{p}}, \tag{42}$$

and

$$p\in[2,\infty]\implies \|T\|_{L_p(\mu)\to L_p(\mu\times\mu)}\leqslant 2^{1-\frac{1}{p}}. \tag{43}$$

Switching to probabilistic terminology, the estimates (42) and (43) say that if Z,Z' are i.i.d. random variables then $\mathbb{E}\big[|Z-Z'|^p\big]\leqslant 2\mathbb{E}\big[|Z|^p\big]$ when $p\in[1,2]$ and $\mathbb{E}\big[|Z-Z'|^p\big]\leqslant 2^{p-1}\mathbb{E}\big[|Z|^p\big]$ when $p\in[2,\infty)$. By applying this to the random variables $Z-a, Z'-a$ for every $a\in\mathbb{R}$, we deduce that the real line (with its usual metric) satisfies (36) with

$$\theta=\theta_p\overset{\mathrm{def}}{=}\max\left\{\frac{1}{p},1-\frac{1}{p}\right\}. \tag{44}$$

Invoking this statement coordinate-wise shows that $\ell_p^3=(\mathbb{R}^3,\|\cdot\|_p)$ satisfies (36) with $\theta=\theta_p$. Lemma 13 therefore implies that $(\mathcal{P}_p(\ell_p^3),\mathsf{W}_p)$ also satisfies (36) with $\theta=\theta_p$. Hence, by Corollary 12 for arbitrarily large $n\in\mathbb{N}$ there exists an n-point metric space (Y,d_Y) such that for every $\alpha\in(\theta_p,1]$,

$$c_{(\mathcal{P}_p(\ell_p^3),\mathsf{W}_p)}(Y,d_Y^\alpha)\gtrsim(\log n)^{\alpha-\theta_p}=\begin{cases}(\log n)^{\alpha-\frac{1}{p}} & \text{if } p\in(1,2],\\ (\log n)^{\alpha+\frac{1}{p}-1} & \text{if } p\in(2,\infty).\end{cases}$$

Since the ℓ_p norm on $\mathbb{R}^3$ is $\sqrt{3}$-equivalent to the ℓ_2 norm on $\mathbb{R}^3$,

$$c_{(\mathcal{P}_p(\ell_p^3),\mathsf{W}_p)}(Y,d_Y^\alpha)\asymp c_{(\mathcal{P}_p(\ell_2^3),\mathsf{W}_p)}(Y,d_Y^\alpha),$$

thus completing the proof of Theorem 6. $\square$

Remark 14. In the proof of Theorem 6 we chose to check the validity of (36) with $\theta=\theta_p$ given in (44) using an interpolation argument since it is very short. But, there are different proofs of this fact: when $p\in[1,2)$ one could start from the trivial case $p=2$, and then pass to general $p\in[1,2)$ by invoking the classical fact [63] that the metric space $(\mathbb{R},|x-y|^{p/2})$ admits an isometric embedding into Hilbert space. Alternatively, in [43, Lemma 3] this is proved via a direct computation.