Planar Earthmover is not in L_1

Extended abstract

Assaf Naor
Microsoft Research
anaor@microsoft.com

Gideon Schechtman*
Weizmann Institute
gideon@weizmann.ac.il

Abstract

We show that any L_1 embedding of the transportation cost (a.k.a. Earthmover) metric on probability measures supported on the grid $\{0,1,\dots,n\}^2 \subseteq \mathbb{R}^2$ incurs distortion $\Omega\left(\sqrt{\log n}\right)$. We also use Fourier analytic techniques to construct a simple L_1 embedding of this space which has distortion $O(\log n)$.

1. Introduction

For a finite metric space (X, d_X) we denote by $\mathscr{P}_X$ the space of all probability measures on X. The transportation cost distance (also known as the Earthmover distance in the computer vision/graphics literature) between two probability measures $\mu, \nu \in \mathscr{P}_X$ is defined by

$$\tau(\mu,\nu) = \min\Bigg\{ \sum_{x,y\in X} d_X(x,y)\pi(x,y) : \forall x,y\in X,\ \pi(x,y)\ge 0, \sum_{z\in X}\pi(x,z)=\mu(x),\ \sum_{z\in X}\pi(z,y)=\nu(y)\Bigg\}.$$

Observe that if μ and ν are the uniform probability distribution over k-point subsets $A \subseteq X$ and $B \subseteq X$, respectively, then

$$\tau(\mu,\nu) = \min\left\{ \frac{1}{k}\sum_{a\in A} d_X(a,f(a)) : f : A\to B \text{ is a bijection}\right\}. \quad (1)$$

This quantity is also known as the *minimum weight matching* between A and B, corresponding to the weight function $d_X(\cdot,\cdot)$ (see [34]). Thus, the Earthmover distance is a natural measure of similarity between images [34, 11, 10]- the distance is the optimal way to match various features, where the cost of such a matching corresponds to the sum of the distances between the features that were matched. Indeed, such metrics occur in various contexts in computer science: Apart from being a popular distance measure in graphics and vision [34, 11, 10, 20], they are used as LP relaxations for classification problems such as 0-extension and metric labelling [8, 7, 2]. Transportation cost metrics are also prevalent in several areas of analysis and PDE (see the book [41] and the references therein).

Following extensive work on nearest neighbor search and data stream computations for L_1 metrics (see [19, 15, 14, 9, 17]), it became of great interest to obtain low distortion embeddings of useful metrics into L_1 (here, and in what follows, L_1 denotes the space of all Lebesgue measurable functions $f : [0,1] \to \mathbb{R}$, such that $\|f\|_1 := \int_0^1 |f(t)|dt < \infty$). Indeed, such embeddings can be used to construct approximate nearest neighbor databases, with an approximation guarantee depending on the *distortion* of the embedding (we are emphasizing here only one aspect of the algorithmic applications of low distortion embeddings into L_1- they are also crucial for the study of various cut problems in graphs, and we refer the reader to [29, 18, 16] for a discussion of these issues).

In the context of the Earthmover distance, nearest neighbor search (a.k.a. similarity search in the vision literature) is of particular importance. It was therefore asked (see, e.g. [28]) whether the Earthmover distance embeds into L_1 with constant distortion (the best known upper bounds on the L_1 distortion were obtained in [7, 20], and will be discussed further below). In [23] the case of the Hamming cube was settled negatively: It is shown there that any embedding of the Earthmover distance on $\{0,1\}^d$ (equipped with the L_1 metric) incurs distortion $\Omega(d)$. However, the most interesting case is that of the Earthmover distance on $\mathbb{R}^2$, as this corresponds to a natural similarity measure between images [10] (indeed, the case of the L_1 embeddability of planar Earthmover distance was explicitly asked in [28]). Here we settle this problem negatively by obtaining the first

*Supported in part by the Israel Science Foundation. Part of this work was carried out while G. Schechtman was visiting Microsoft Research.

super-constant lower bound on the L_1 distortion of the planar Earthmover distance. To state it we first recall some definitions.

Given two metric spaces (X, d_X) and (Y, d_Y), and a mapping $f : X \to Y$, we denote its Lipschitz constant by

$$\|f\|_{\mathrm{Lip}} := \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_Y(f(x), f(y))}{d_X(x,y)}.$$

If f is one to one then its distortion is defined as

$$\begin{aligned}\mathrm{dist}(f) &:= \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}\\ &= \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_Y(f(x), f(y))}{d_X(x,y)} \cdot \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_X(x,y)}{d_Y(f(x), f(y))}.\end{aligned}$$

The smallest distortion with which X can be embedded into Y is denoted $c_Y(X)$, i.e.,

$$c_Y(X) := \inf\left\{\mathrm{dist}(f) :\ f : X \hookrightarrow Y \text{ is one to one}\right\}.$$

When $Y = L_p$ we use the shorter notation $c_Y(X) = c_p(X)$. Thus, the parameter $c_2(X)$ is the Euclidean distortion of X and $c_1(X)$ is the L_1 distortion of X.

Our main result bounds from below the L_1 distortion of the space of probability measures on the n by n grid, equipped with the transportation cost distance.

Theorem 1.1. $c_1\left(\mathscr{P}_{\{0,1,\ldots,n\}^2}, \tau\right) = \Omega\left(\sqrt{\log n}\right)$.

After reducing the problem to a functional analytic question, our proof of Theorem 1.1 is a discretization of a theorem of Kislyakov from 1975 [25]. We attempted to make the presentation self contained by presenting here appropriate versions of the various functional anlaytic lemmas that are used in the proof.

For readers who are more interested in the minimum cost matching metric (1), we also prove the following lower bound:

Theorem 1.2 (Discretization). *For arbitrarily large integers n there is a family $\mathscr{Y}$ of disjoint n-point subsets of $\left\{0, 1\ldots, n^3\right\}^2$, with $|\mathscr{Y}| \le n^{O(\log\log n)}$, such that any L_1 embedding of $\mathscr{Y}$, equipped with the minimum weight matching metric τ, incurs distortion*

$$\Omega\left(\sqrt{\log\log\log n}\right) = \Omega\left(\sqrt{\log\log\log|\mathscr{Y}|}\right)$$

A metric spaces (X, d_X) is said to embed into squared L_2, or to be of negative type, if the metric space $\left(X, \sqrt{d_X}\right)$ is isometric to a subset of L_2. Squared L_2 metrics are important in various algorithmic applications since it is possible to efficiently solve certain optimization problems on them using semidefinite programming (see the discussion in [3, 24]). It turns out that planar Earthmover does not embed into any squared L_2 metric:

Theorem 1.3 (Nonembeddability into squared L_2). $\lim_{n\to\infty} c_2\left(\mathscr{P}_{\{0,\ldots,n\}^2}, \sqrt{\tau}\right) = \infty$.

Motivated by the proof of Theorem 1.1, we also construct simple low-distortion embeddings of the space $\left(\mathscr{P}_{\{0,1,\ldots,n\}^2}, \tau\right)$ into L_1. It is convenient to work with probability measures on the torus $\mathbb{Z}_n^2$ instead of the grid $\{0, 1, \ldots, n\}^2$. One easily checks that $\{0, \ldots, n\}^2$ embeds with constant distortion into $\mathbb{Z}_{2n}^2$ (see e.g. Lemma 6.12 in [30]). Every $\mu \in \mathscr{P}_{\mathbb{Z}_n^2}$ can be written in the Fourier basis as

$$\mu = \sum_{(u,v)\in\mathbb{Z}_n^2} \widehat{\mu}(u,v) e_{uv}, \tag{2}$$

where

$$\forall (a,b), (u,v) \in \mathbb{Z}_n^2,\ e_{uv}(a,b) := e^{\frac{2\pi i(au+bv)}{n}},$$

and

$$\forall (u,v) \in \mathbb{Z}_n^2,\ \widehat{\mu}(u,v) := \frac{1}{n^2} \sum_{(a,b)\in\mathbb{Z}_n^2} \mu(a,b) e_{uv}(-a,-b).$$

Observe that for $n = 2^k + 1$, $k \in \mathbb{N}$, the decomposition (2) can be computed in time $O\left(n^2 \log n\right)$ using the fast Fourier transform [37]. Motivated in part by the results of [32] (see also [5, 33]), we define

$$A\mu = \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{e^{\frac{2\pi i u}{n}} - 1}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{\mu}(u,v) \cdot e_{uv}, \tag{3}$$

and

$$B\mu = \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{e^{\frac{2\pi i v}{n}} - 1}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{\mu}(u,v) \cdot e_{uv}. \tag{4}$$

Theorem 1.4. *The mapping $\mu \mapsto (A\mu, B\mu)$ from $\left(\mathscr{P}_{\mathbb{Z}_n^2}, \tau\right)$ to $L_1\left(\mathbb{Z}_n^2\right) \oplus L_1\left(\mathbb{Z}_n^2\right)$ is bi-Lipschitz, with distortion $O(\log n)$.*

The $O(\log n)$ distortion in Theorem 1.4 matches the best known distortion guarantee proved in [20, 7]. But, our embedding has various new features. First of all, it is a *linear* mapping into a low dimensional L_1 space, which is based on the computation of the Fourier transform. It is thus very fast to compute, and is versatile in the sense that it might behave better on images whose Fourier transform is sparse (we do not study this issue here). Thus there is scope to apply the embedding on certain subsets of the frequencies, and this might improve the performance in practice. This is an interesting "applied" question which should be investigated further. For lack of space, the proof of Theorem 1.4 will be deferred to the full version of this paper.

2. Preliminaries and notation

For the necessary background on measure theory we refer to the book [38], however, in the setting of the present paper, our main results will deal with finitely supported measures, in which case no background and measurabilty assumptions are necessary. We also refer to the book [41] for background on the theory of optimal transportation of measures. Let (X, d_X) be a metric space. We denote by $\mathscr{M}_X$ the space of all Borel measures on X with bounded total variation, and by $\mathscr{P}_X \subseteq \mathscr{M}_X$ the set of all Borel *probability* measures on X. We also let $\mathscr{M}_X^+ \subseteq \mathscr{M}_X$ be the space of *non-negative* measures on X with finite total mass, and we denote by $\mathscr{M}_X^0 \subseteq \mathscr{M}_X$ the space of all measures $\mu \in \mathscr{M}_X$ with $\mu(X) = 0$. Given a measure $\mu \in \mathscr{M}_X$, we can decompose it in a unique way as $\mu = \mu^+ - \mu^-$, where $\mu^+, \mu^- \in \mathscr{M}_X^+$ are disjointly supported. If $\mu, \nu \in \mathscr{M}_X^+$ have the same total mass, i.e. $\mu(X) = \nu(X) < \infty$, then we let $\Pi(\mu, \nu)$ be the space of all *couplings* of μ and ν, i.e. all non-negative Borel measures π on $X \times X$ such that for every measurable bounded $f : X \to \mathbb{R}$,

$$\int_{X\times X} f(x)d\pi(x,y) = \int_X f(x)d\mu(x),$$

and

$$\int_{X\times X} f(y)d\pi(x,y) = \int_X f(y)d\nu(y).$$

Observe that in the case of finitely supported measures, this condition translates to the standard formulation, in which we require that the marginals of π are μ and ν, i.e.

$$\forall x, y \in X,\ \sum_{z\in X} \pi(x,z) = \mu(x) \ \wedge \ \sum_{z\in X} \pi(z,y) = \nu(y).$$

The *transportation cost distance* between μ and ν, denoted here by $\tau(\mu,\nu) = \tau_{(X,d_X)}(\mu.\nu)$ (and also referred to in the literature as the Wasserstein 1 distance, Monge-Kantorovich distance, or the Earthmover distance), is

$$\tau(\mu,\nu) := \inf\left\{\int_{X\times X} d_X(x,y)\,d\pi(x,y) :\ \pi \in \Pi(\mu,\nu)\right\}. \quad (5)$$

For $\mu \in \mathscr{M}_X^0$, $\mu^+(X) = \mu^-(X)$, so we may write $\|\mu\|_\tau := \tau(\mu^+, \mu^-)$. This is easily seen to be a norm on the vector space $\mathscr{M}_{X,\tau}^0 := \left\{\mu \in \mathscr{M}_X^0 :\ \|\mu\|_\tau < \infty\right\}$.

Fix some $x_0 \in X$, and let $\mathrm{Lip}_0(X) = \mathrm{Lip}_{x_0}(X)$ be the linear space of all Lipschitz mappings $f : X \to \mathbb{R}$ with $f(x_0) = 0$, equipped with the norm $\|\cdot\|_{\mathrm{Lip}}$ (i.e. the norm of a function equals its Lipschitz constant). Any $\mu \in \mathscr{M}_{X,\tau}^0$ can be thought of as a bounded linear functional on $\mathrm{Lip}_0(X)$, given by $f \mapsto \int_X f d\mu$. The famous *Kantorovich duality theorem* (see Theorem 1.14 in [41]) implies that $\mathrm{Lip}_0(X)^* = \mathscr{M}_{X,\tau}^0$, in the sense that every bounded linear functional on $\mathrm{Lip}_0(X)$ is obtained in this way, and for every $\mu \in \mathscr{M}_{X,\tau}^0$,

$$\|\mu\|_\tau = \|\mu\|_{\mathrm{Lip}_0(X)^*} := \sup\left\{\int_X f d\mu :\ f \in \mathrm{Lip}_0(X),\ \|f\|_{\mathrm{Lip}} \le 1\right\}.$$

(We note that this identity amounts to duality of linear programming.)

3. Proof of Theorem 1.1

Fix an integer $n \ge 2$ and let $X = \{0, 1, \dots, n-1\}^2$, equipped with the standard Euclidean metric. In what follows, for concreteness, $\mathrm{Lip}_0 := \mathrm{Lip}_0(X)$ is defined using the base point $x_0 = (0,0)$. Also, for ease of notation we denote $\mathscr{M} = \mathscr{M}_{X,\tau}^0$. Observe that Lip_0 and $\mathscr{M}$ are vector spaces of dimension $n^2 - 1$, and by Kantorovich duality, $\mathrm{Lip}_0^* = \mathscr{M}$ and $\mathscr{M}^* = \mathrm{Lip}_0$.

Assume that $F : \mathscr{P}_X \to L_1$ is a bi-Lipschitz embedding, satisfying for all two probability measures $\mu, \nu \in \mathscr{P}_X$,

$$\tau(\mu,\nu) \le \|F(\mu) - F(\nu)\|_1 \le L \cdot \tau(\mu,\nu). \quad (6)$$

Our goal is to bound L from below. We begin by reducing the problem to the case of *linear mappings*. Recall that given two normed spaces $(Z, \|\cdot\|_Z)$ and $(W, \|\cdot\|_W)$, the norm of a linear mapping $T : Z \to W$ is defined as $\|T\| = \sup_{z\in Z\setminus\{0\}} \frac{\|Tz\|_W}{\|z\|_Z}$ (observe that in this case $\|T\| = \|T\|_{\mathrm{Lip}}$).

Lemma 3.1 (Reduction to a linear embedding of $\mathscr{M}$ into ℓ_1^N). *Under the assumption of an existence of an embedding $F : \mathscr{P}_X \to L_1$ satisfying* (6)*, there exists an integer N, and an invertible linear operator $T : \mathscr{M} \to \ell_1^N$, with $\|T\| \le 2L$ and $\|T(\mu)\|_1 \ge \|\mu\|_\tau$ for all $\mu \in \mathscr{M}$ (the factor 2 can be replaced by $1+\varepsilon$ for every $\varepsilon > 0$, but this is irrelevant for us here).*

Proof. By translation we may assume that F maps the uniform measure on X to 0. For $\mu \in \mathscr{M}$ denote $\|\mu\|_\infty := \max_{x\in X} |\mu(x)|$. Observe that it is always the case that $\|\mu\|_\infty \le \|\mu\|_\tau$. Indeed, if $\pi \in \Pi(\mu^+, \mu^-)$ then

$$\int_{X\times X} \|x-y\|_2 d\pi(x,y) \ge \int_{X\times X} d\pi(x,y) = \mu^+(X) = \mu^-(X) \ge \|\mu\|_\infty.$$

Let $B_{\mathscr{M}}$ denote the unit ball of $\mathscr{M}$. Define for $\mu \in B_{\mathscr{M}}$ a probability measure $\psi(\mu) \in \mathscr{P}(X)$ by $\psi(\mu)(x) := \frac{\mu(x)+1}{n^2}$. Then every $\mu, \nu \in \mathscr{M}$ satisfy $\|\mu - \nu\|_\tau = \frac{\|\psi(\mu)-\psi(\nu)\|_\tau}{n^2}$.

The mapping $h{:=}n^2\cdot F\circ\psi : B_{\mathscr{M}} \to L_1$ satisfies $h(0)=0$, $\|h\|_{\mathrm{Lip}}\le L$, and $\|h(\mu)-h(\nu)\|_1 \ge \|\mu-\nu\|_\tau$. This implies that there exists a map $\tilde{h}:\mathscr{M}\to L_1$ satisfying the same inequalities. We shall present two arguments establishing this fact: The first is a soft non-constructive proof, using the notion of ultraproducts, and the second argument is more elementary, but does not preserve the Lipschitz constant.

Let $\mathscr{U}$ be a free ultrafilter on $\mathbb{N}$, and denote by $(L_1)_{\mathscr{U}}$ the corresponding ultrapower of L_1 (see [12] for the necessary background on ultrapowers of Banach spaces. In particular, it is shown there that $(L_1)_{\mathscr{U}}$ is isometric to an $L_1(\sigma)$ space, for some measure σ). Define for $\mu\in\mathscr{M}$, $\tilde{h}(\mu)=(j\cdot h(\mu/j))_{j=1}^{\infty}/\mathscr{U}$, where we set, say, $h(\nu)=0$ for $\nu\in\mathscr{M}\setminus B_{\mathscr{M}}$. Then, by standard arguments, $\|\tilde{h}\|_{\mathrm{Lip}}\le L$ and $\|\tilde{h}^{-1}\|_{\mathrm{Lip}}\le 1$. Moreover, $\tilde{h}(\mathscr{M})$ spans a separable subspace of $(L_1)_{\mathscr{U}}$, and thus we may assume without loss of generality that $\tilde{h}$ takes values in L_1.

An alternative proof (for those of us who don't mind losing a constant factor), proceeds as follows. For every $f\in L_1$ let $\chi(f):[0,1]\times\mathbb{R}\to\{-1,0,1\}$ be the function given by

$$\chi(f)(s,t)=\mathrm{sign}(f(s))\cdot\mathbf{1}_{\left[0,|f(s)|\right]}(t) = \begin{cases}1 & f(s)>0,\ 0\le t\le f(s),\\ -1 & f(s)<0,\ 0\le t\le -f(s),\\ 0 & \text{otherwise.}\end{cases}$$

Then $\|f-g\|_1=\|\chi(f)-\chi(g)\|_{L_1([0,1]\times\mathbb{R})}$ for every $f,g\in L_1$ (We note here that the space $L_1([0,1]\times\mathbb{R})$ is isometric to L_1.) Define $\tilde{h}:\mathscr{M}\to L_1([0,1]\times\mathbb{R})$ by setting $\tilde{h}(\mu)=\|\mu\|_\tau\cdot\chi\circ h(\mu/\|\mu\|_\tau)$ for $\mu\in\mathscr{M}\setminus\{0\}$, and $\tilde{h}(0)=0$. Since for every $f\in L_1$, $\chi(f)$ takes values in $\{-1,0,1\}$, we have the following pointwise identity for every $\mu,\nu\in\mathscr{M}$ with $\|\mu\|_\tau\ge\|\nu\|_\tau$:

$$\left|\tilde{h}(\mu)-\tilde{h}(\nu)\right| = \|\nu\|_\tau\cdot\left|\chi\circ h\left(\frac{\mu}{\|\mu\|_\tau}\right)-\chi\circ h\left(\frac{\nu}{\|\nu\|_\tau}\right)\right| + (\|\mu\|_\tau-\|\nu\|_\tau)\cdot\left|\chi\circ h\left(\frac{\mu}{\|\mu\|_\tau}\right)\right|.$$

Thus

$$\begin{aligned}
&\left\|\tilde{h}(\mu)-\tilde{h}(\nu)\right\|_{L_1([0,1]\times\mathbb{R})}\\
&= \|\nu\|_\tau\cdot\left\|h\left(\frac{\mu}{\|\mu\|_\tau}\right)-h\left(\frac{\nu}{\|\nu\|_\tau}\right)\right\|_1 + (\|\mu\|_\tau-\|\nu\|_\tau)\cdot\left\|h\left(\frac{\mu}{\|\mu\|_\tau}\right)\right\|_1 \qquad (7)\\
&\ge \|\nu\|_\tau\cdot\left\|\frac{\mu}{\|\mu\|_\tau}-\frac{\nu}{\|\nu\|_\tau}\right\|_\tau+\|\mu\|_\tau-\|\nu\|_\tau\\
&\ge \|\nu-\mu\|_\tau-\left\|\mu-\frac{\|\nu\|_\tau}{\|\mu\|_\tau}\mu\right\|_\tau+\|\mu\|_\tau-\|\nu\|_\tau\\
&= \|\nu-\mu\|_\tau.
\end{aligned}$$

It also follows from the identity (7) that

$$\begin{aligned}
&\left\|\tilde{h}(\mu)-\tilde{h}(\nu)\right\|_{L_1([0,1]\times\mathbb{R})}\\
&\le L\|\nu\|_\tau\cdot\left\|\frac{\mu}{\|\mu\|_\tau}-\frac{\nu}{\|\nu\|_\tau}\right\|_\tau+L\|\mu-\nu\|_\tau\\
&\le L\|\mu-\nu\|_\tau+L\|\nu\|_\tau\|\mu\|_\tau\cdot\left|\frac{1}{\|\mu\|_\tau}-\frac{1}{\|\nu\|_\tau}\right|\\
&\quad +L\|\mu-\nu\|_\tau\\
&\le 3L\|\mu-\nu\|_\tau.
\end{aligned}$$

We are now in position to use a Theorem of Ribe [36] (see also [13], and Corollary 7.10 in [4], for softer proofs), which implies that there is an into linear isomorphism $S:\mathscr{M}\to L_1^{**}$ satisfying $\|S\|\le L$ and $\|S^{-1}\|\le 1$. Since $\mathscr{M}$ is finite dimensional, by the principle of local reflexivity [26] (alternatively by Kakutani's representation theorem [21, 27]), and a simple approximation argument, we get that there exists an integer N and an into linear isomorphism $T:\mathscr{M}\to\ell_1^N$ satisfying $\|T\|\le 2L$ and $\|T^{-1}\|\le 1$ (the value of N is irrelevant for us here, and indeed it is possible to conclude the proof without passing to a finite dimensional L_1 space, but this slightly simplifies some of the ensuing arguments. For completeness we note here that using a theorem of Talagrand [39] we can ensure that $N=O(n^2\log n)$). $\square$

From now on let $T:\mathscr{M}\to\ell_1^N$ be the linear operator guaranteed by Lemma 3.1. Since T is an isomorphism, the adjoint operator $T^*:\ell_\infty^N\to\mathscr{M}^*=\mathrm{Lip}_0$ is a quotient mapping, i.e. $\|T^*\|\le 2L$ and the image of the unit ball of ℓ_∞^N under T^* contains the unit ball of Lip_0.

The rest of the proof follows Kislyakov's [25] and is a discretization of his argument. The idea is to compose T^* with a map $\mathscr{F}$ which is the imaginary part of the discrete two dimensional Fourier transform (see the exact definition below), seen as a map from Lip_0 to $\ell_2(X)$, and to prove two properties of the composed map: Using the fact that $\|T^*\|\le 2L$ we shall show that $\mathscr{F}\circ T^*$ is *order bounded* with good bound, that is,

$$\mathscr{F}\left(T^*(B_{\ell_\infty^N})\right)\subseteq\{y\in\ell_2(X):\ |y|\le x\},$$

for some $x\in\ell_2(X)$ such that $\|x\|_2\le 4Ln$. Then, using the quotient property of T^*, we find a family of functions $\{\phi_i\in B_{\ell_\infty^N}\}_{i\in I}$ such that if $\mathscr{F}(T^*(\phi_i))\le x$ for all $i\in I$ then $\|x\|_2\ge cn\sqrt{\log n}$, for some universal $c>0$.

We now define two more auxiliary linear operators. The first is the formal identity $\mathrm{Id}:\mathrm{Lip}_0\to W$, where W is the space of all functions $f:X\to\mathbb{R}$ with $f(0)=0$, equipped with the (discrete Sobolev) norm

$$\begin{aligned}\|f\|_W \;&:=\; \sum_{i=0}^{n-1}\sum_{j=0}^{n-2}|f(i,j+1)-f(i,j)| \\ &+\sum_{j=0}^{n-1}\sum_{i=0}^{n-2}|f(i+1,j)-f(i,j)| \\ &+n\sum_{i=0}^{n-2}|f(i+1,0)-f(i,0)| \\ &+n\sum_{j=0}^{n-2}|f(0,j+1)-f(0,j)|.\end{aligned}$$

The second operator is also a formal identity (discrete Sobolev embedding) $S : W \to \ell_2(X)$, where the Euclidean norm on $\ell_2(X)$ is taken with respect to the counting measure on X. The final operator that we will use is the imaginary part of the Fourier operator, already referred to above, which we denote by $\mathscr{F} : \ell_2(X) \to \ell_2(X)$. It is defined for $f : X \to \mathbb{R}$ by

$$\begin{aligned}\mathscr{F}(f)(u,v) &:= \Im\left(\frac{1}{n^2}\sum_{(k,\ell)\in X} f(k,\ell)e^{\frac{2\pi i(uk+v\ell)}{n}}\right) \\ &= \frac{1}{n^2}\sum_{(k,\ell)\in X} f(k,\ell)\sin\left(\frac{2\pi(uk+v\ell)}{n}\right).\end{aligned}$$

The following lemma summarizes known estimates on the norms of these operators.

Lemma 3.2 (Operator norm bounds). *The following operator norm bounds hold true:*
• $\|\mathrm{Id}\| \le 4n(n-1)$. • $\|S\| \le \frac{1}{2}$. • $\|\mathscr{F}\| \le \frac{1}{n}$.

Proof. The first statement means that for every $f : X \to \mathbb{R}$ with $f(0) = 0$, $\|f\|_W \le 4n(n-1)\|f\|_{\mathrm{Lip}}$, which is obvious from the definitions. The second assertion is that $\|f\|_2 \le \frac{1}{2}\|f\|_W$. This is a discrete version of Sobolev's inequality [33] (with non-optimal constant), which can be proved as follows. First of all, since $f(0) = 0$, for every $(u,v) \in X$,

$$\begin{aligned}|f(u,v)| \;&=\; \left|\sum_{k=0}^{u-1}[f(k+1,v)-f(k,v)] + \sum_{\ell=0}^{v-1}[f(0,\ell+1)-f(0,\ell)]\right| \\ &\le\; \sum_{k=0}^{n-2}|f(k+1,v)-f(k,v)| + \sum_{\ell=0}^{n-2}|f(0,\ell+1)-f(0,\ell)| \\ &:=\; A(v). \qquad (8)\end{aligned}$$

Analogously,

$$\begin{aligned}|f(u,v)| \;&\le\; \sum_{\ell=0}^{n-2}|f(u,\ell+1)-f(u,\ell)| + \sum_{k=0}^{n-2}|f(k+1,0)-f(k,0)| \\ &:=\; B(u). \qquad (9)\end{aligned}$$

Multiplying (8) and (9), and summing over X, we see that

$$\begin{aligned}\|f\|_2^2 \;&\le\; \sum_{(u,v)\in X} A(v)B(u) \\ &=\; \left(\sum_{v=0}^{n-1}A(v)\right)\cdot\left(\sum_{u=0}^{n-1}B(u)\right) \\ &\le\; \frac{1}{4}\left(\sum_{v=0}^{n-1}A(v)+\sum_{u=0}^{n-1}B(u)\right)^2 \\ &=\; \frac{1}{4}\|f\|_W^2.\end{aligned}$$

The final assertion follows from the fact that the system of functions $\left\{(k,\ell)\mapsto e^{\frac{2\pi i(uk+v\ell)}{n}}\right\}_{(u,v)\in X}$ are orthogonal in $\ell_2^{\mathbb{C}}(X)$ (the space of complex valued functions on X), and have norms bounded by n. □

We now recall some facts related to absolutely summing operators on Banach spaces (we refer the interested reader to [40, 42] for more information on this topic). Given two Banach spaces Y and Z, the π_1 norm of an operator $A : Y \to Z$, denoted $\pi_1(A)$, is defined to be the smallest constant $K > 0$ such that for every $m \in \mathbb{N}$ and every $y_1,\ldots,y_m \in Y$ there exists a norm 1 linear functional $y^* \in Y^*$ satisfying

$$\sum_{j=1}^{m}\|Ay_j\|_Z \le K\sum_{j=1}^{m}|y^*(y_j)|. \qquad (10)$$

This defines an *ideal norm* in the sense that it is a norm, and for every two operators $P : W \to Y$ and $Q : Z \to V$ we have $\pi_1(QAP) \le \|Q\|\cdot\pi_1(A)\cdot\|P\|$. Observe that it is always the case that $\pi_1(A) \ge \|A\|$.

Lemma 3.3. *Using the above notation we have that* $\pi_1(\mathrm{Id}) \le 4n(n-1)$. *Therefore, Lemma 3.2 implies that*

$$\pi_1(\mathscr{F}\circ S\circ \mathrm{Id}\circ T^*) \le 4nL.$$

Proof. Fix m functions $f_1,\ldots,f_m : X \to \mathbb{R}$ such that

$f_1(0) = \cdots = f_m(0) = 0$. Then

$$\begin{aligned}
\sum_{j=1}^{m} \|f_j\|_W = & \sum_{s=1}^{n-1}\sum_{t=0}^{n-2}\sum_{j=1}^{m} \big(|f_j(s,t+1) - f_j(s,t)| \\
& +|f_j(t+1,s) - f_j(t,s)|\big) \\
& +(n+1)\sum_{t=0}^{n-2}\sum_{j=1}^{m} \big(|f_j(0,t+1) - f_j(0,t)| \\
& +|f_j(t+1,0) - f_j(t,0)|\big) \\
\leq \; & 4n(n-1) \\
& \cdot \max\Bigg\{ \max_{\substack{0\le s\le n-1\\ 0\le t\le n-2}} \sum_{j=1}^{m} |f_j(s,t+1) - f_j(s,t)|, \\
& \max_{\substack{0\le s\le n-1\\ 0\le t\le n-2}} \sum_{j=1}^{m} |f_j(t+1,s) - f_j(t,s)| \Bigg\}.
\end{aligned}$$

Assume without loss of generality that the maximum above equals $\sum_{j=1}^{m} |f_j(s_0,t_0+1) - f_j(s_0,t_0)|$, for some $0 \le s_0 \le n-1$ and $0 \le t_0 \le n-2$. Consider the measure $\mu = \delta_{(s_0,t_0+1)} - \delta_{(s_0,t_0)} \in \mathscr{M} = \mathrm{Lip}_0^*$. One checks that $\|\mu\|_\tau = 1$, and $\sum_{j=1}^{m} |f_j(s_0,t_0+1) - f_j(s_0,t_0)| = \sum_{j=1}^{m} |\mu(f_j)|$, implying the required result. □

The fundamental property of the π_1 norm is the Pietsch Factorization Theorem (see [40]), a special (particularly easy) case of which is the following lemma. We present a proof for the sake of completeness.

Lemma 3.4 (Pietsch factorization). *Let Y be a Banach space, and fix a linear operator $A : \ell_\infty^N \to Y$. Then there exists a probability measure σ on $\{1,\dots,N\}$ and a linear operator $R : L_1(\sigma) \to Y$ such that $A = R \circ I$, where I is the formal identity from ℓ_∞^N to $L_1(\sigma)$, and $\|R\| = \pi_1(A)$.*

Proof. Recall that $A : \ell_\infty^N \to Y$ satisfies for all $x_1,\dots,x_m \in \ell_\infty^N$,

$$\begin{aligned}
\sum_{j=1}^{m} \|Ax_j\| \;&\le\; \pi_1(A) \cdot \sup_{\substack{x^*\in(\ell_\infty^N)^*\\ \|x^*\|=1}} \sum_{j=1}^{m} |x^*(x_j)| \\
&=\; \pi_1(A) \cdot \max_{1\le k\le N} \sum_{j=1}^{m} |x_j(k)|, \qquad (11)
\end{aligned}$$

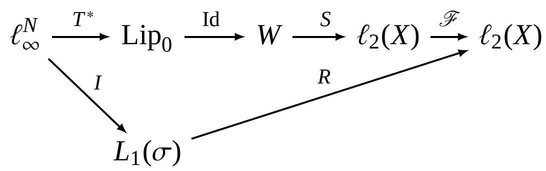
where the last equality follows from the fact that the evaluation functionals $x \mapsto x(k)$ are the extreme points of the unit ball of $\ell_1^N = \left(\ell_\infty^N\right)^*$.

Denoting by $e_1,\dots,e_N$ the standard basis of $\mathbb{R}^N$ we deduce from (11) that $\pi_1(A) \ge \sum_{j=1}^{N} \|Ae_j\|$. Define a probability measure σ on $\{1,\dots,N\}$ by $\sigma(k) = \frac{\|Ae_k\|}{\sum_{j=1}^{N}\|Ae_j\|}$. Then for every $x \in \ell_\infty^N$ we see that

$$\begin{aligned}
\|Ax\| \;&=\; \left\|\sum_{k=1}^{N} x(k)Ae_k\right\| \\
&\le\; \sum_{k=1}^{N} |x(k)| \cdot \|Ae_k\| \\
&=\; \left(\sum_{j=1}^{N} \|Ae_j\|\right) \int_{\{1,\dots,N\}} |x(k)| d\sigma(k) \\
&\le\; \pi_1(A) \int_{\{1,\dots,N\}} |x(k)| d\sigma(k).
\end{aligned}$$

Defining $Rx = Ax$, this implies the required result. □

From now on let R and σ be the operator and probability measure corresponding to $A = \mathscr{F}\circ S\circ \mathrm{Id}\circ T^*$ in Lemma 3.4. Thus $R\circ I = \mathscr{F}\circ S\circ\mathrm{Id}\circ T^*$ and $\|R\| \le 4nL$. Schematically, we have the following commuting diagram:



We need only one more simple result from classical Banach space theory. This is a special case of a more general theorem, but we shall prove here only what is needed to conclude the proof of Theorem 1.1.

Lemma 3.5. *Let $R : L_1(\sigma) \to \ell_2$ be a linear operator. Fix $f : \{0,\dots,N\} \to [0,\infty)$. Then there is $x \in \ell_2$ with non-negative coordinates such that*

$$\begin{aligned}
& R\left(\{g : \{0,\dots,N\} \to \mathbb{R} : \forall j,\ |g(j)| \le f(j)\}\right) \\
& \qquad\qquad \subseteq \{y \in \ell_2 : \forall j,\ |y_j| \le x_j\},
\end{aligned}$$

and $\|x\|_2 \le \|R\| \cdot \|f\|_{L_1(\sigma)}$.

Proof. R is given by a matrix $(R_{ij} : i = 1,\dots,N,\ j \in \mathbb{N})$. In other words, for every j, $(Rf)_j = \sum_{i=1}^{N} R_{ij} f(i)$. Observe that using this notation,

$$\|R\| = \max_{1\le i\le N} \left(\frac{1}{\sigma(i)^2}\sum_{j=1}^{\infty} R_{ij}^2\right)^{1/2}. \qquad (12)$$

Fix $g \in L_1(\sigma)$ such that for all $i \in \{1,\dots,N\}$, $|g(i)| \le f(i)$. Then for all j,

$$|(Rg)_j| \le \sum_{i=1}^{N} |R_{ij}| f(i) := x_j.$$

Now,

$$\begin{aligned}
\|x\|_2 &= \left[\sum_{j=1}^{\infty}\left(\sum_{i=1}^{N}|R_{ij}|f(i)\right)^2\right]^{1/2}\\
&\le \sum_{i=1}^{N}\left(\sum_{j=1}^{\infty}|R_{ij}|^2 f(i)^2\right)^{1/2}\\
&= \sum_{i=1}^{n}\sigma(i)f(i)\left(\frac{1}{\sigma(i)^2}\sum_{j=1}^{\infty}R_{ij}^2\right)^{1/2}\\
&\le \|R\|\cdot\|f\|_{L_1(\sigma)},
\end{aligned}$$

where we have used (12). □

We are now in position to conclude the proof of Theorem 1.1.

Proof of Theorem 1.1. For $(u,v)\in\{1,\ldots,n-1\}^2$ define $\varphi_{u,v}: X\to\mathbb{R}$ by

$$\varphi_{u,v}(k,\ell):=\frac{1}{u+v}\cdot\sin\left(\frac{2\pi(uk+v\ell)}{n}\right).$$

Then $\varphi_{u,v}(0)=0$ and one computes that $\|\varphi_{u,v}\|_{\mathrm{Lip}}<\frac{4\pi}{n}$. By the fact that T^* maps the unit ball of ℓ_∞^N onto the unit ball of Lip_0, it follows that there is $\phi_{u,v}\in\ell_\infty^N$ with $\|\phi_{u,v}\|_\infty\le\frac{4\pi}{n}$ and $T^*\phi_{u,v}=\varphi_{u,v}$. Now, the functions $|I(\phi_{u,v})|\in L_1(\sigma)$ are point-wise bounded by the constant $\frac{4\pi}{n}$, so by Lemma 3.5 there exists $x\in\ell_2(X)$ of norm at most $\frac{4\pi}{n}\|R\|\le 16\pi L$ such that $|R(I(\phi_{u,v}))|$ is bounded pointwise by x.

Note that

$$\begin{aligned}
&R\circ I(\phi_{u,v})(u,v)=\mathscr{F}\circ S\circ\mathrm{Id}\circ T^*(\phi_{u,v})(u,v)\\
&= \mathscr{F}(\varphi_{u,v})(u,v)\\
&= \frac{1}{n^2}\sum_{(k,\ell)\in X}\frac{1}{u+v}\cdot\sin^2\left(\frac{2\pi(uk+v\ell)}{n}\right)\\
&= \frac{1}{n^2(u+v)}\sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1}\left(\frac12-\frac14\cdot e^{2\pi i\cdot\frac{2uk}{n}}\cdot e^{2\pi i\cdot\frac{2v\ell}{n}}\right.\\
&\qquad\left.-\frac14\cdot e^{-2\pi i\cdot\frac{2uk}{n}}\cdot e^{-2\pi i\cdot\frac{2v\ell}{n}}\right)\\
&= \begin{cases}\frac{1}{2(u+v)} & (u,v)\neq\left(\frac n2,\frac n2\right),\\ 0 & (u,v)=\left(\frac n2,\frac n2\right).\end{cases}
\end{aligned}$$

But

$$\begin{aligned}
(16\pi L)^2 &\ge \|x\|_2^2\\
&\ge \sum_{u,v=1}^{n-1}x_{u,v}^2\\
&\ge \sum_{u,v=1}^{n-1}\Big[R\circ I(\phi_{u,v})(u,v)\Big]^2\\
&\ge \frac18\sum_{u,v=1}^{n-1}\frac{1}{(u+v)^2}\\
&\ge \frac{\log n}{16},
\end{aligned}$$

where the last bound follows from comparison with the appropriate integrals. The proof of Theorem 1.1 is complete. □

4. Discretization and minimum weight matching

In this section we deduce Theorem 1.2 from Theorem 1.1. The main tool is the following theorem of Bourgain [6], which gives a quantitative version of Ribe's theorem [36].

Theorem 4.1 (Bourgain's quantitative version of Ribe's theorem [6]). *There exists a universal constant C with the following property. Let Y and Z be Banach spaces, $\dim(Y)=d$. Assume that $\mathscr{Y}$ is an ε-net in the unit ball of Y, $f:\mathscr{Y}\to Z$ satisfies $\mathrm{dist}(f)\le D$, and that $\log\log\frac{1}{\varepsilon}\ge Cd\log D$. Then there exists an invertible linear operator $T: Y\to Z$ satisfying $\|T\|\cdot\|T^{-1}\|\le C\cdot D$.*

Proof of Theorem 1.2. Observe that for every $\mu\in\mathscr{M}$, the measure $\frac{1}{\mu^+(X)}\cdot(\mu^+\otimes\mu^-)$ is in $\Pi(\mu^+,\mu^-)$. Thus

$$\begin{aligned}
\|\mu\|_\tau &\le \frac{1}{\mu^+(X)}\int_{X\times X}\|x-y\|_2 d\mu^+(x)d\mu^-(y)\\
&\le \sqrt{2}\cdot(n-1)\cdot\mu^+(X)\\
&\le 2n\cdot|\mathrm{supp}(\mu^+)|\cdot\|\mu\|_\infty\\
&\le 2n^3\|\mu\|_\infty.
\end{aligned}$$

On the other hand, as we have seen in the proof of Lemma 3.1, for every $\mu\in\mathscr{M}$, $\|\mu\|_\infty\le\|\mu\|_\tau$. It follows from these consideration, and Theorems 1.1 and 4.1, that for every integer $N\ge e^{e^{C'n^2\log\log n}}$, the set of probability measures $\mathscr{Y}\subseteq\mathscr{P}_X$ consisting of measures $\mu\in\mathscr{P}_X$ such that for all $x\in X$, $\mu(x)=k/N$ for some $k\in\{0,\ldots,N\}$, satisfies $c_1(\mathscr{Y},\tau)=\Omega\left(\sqrt{\log n}\right)$. We pass to a family of subsets as follows. Let M be an integer which will be determined later. For every $\mu\in\mathscr{Y}$ we assign a subset

$S_\mu \subseteq \{0,\dots,nM\}^2$ as follows. For every $(u,v)\in X=\{0,\dots,n-1\}^2$, if $\mu(u,v)=k/N$, where $k\in\{0,\dots,N\}$, then S_μ will contain arbitrary k distinct points from the set $(uM,vM)+\left\{0,\dots,\left\lceil\sqrt{N}\right\rceil\right\}^2$. Provided $M\ge 4\sqrt{N}$, the sets $\{S_\mu\}_{\mu\in\mathscr{Y}}$ thus obtained are disjoint N point subsets of $\{0,\dots,nM\}^2$, and it is straightforward to check that the minimum weight matching metric on $\{S_\mu\}_{\mu\in\mathscr{Y}}$ is bi-Lipschitz equivalent to $(\mathscr{Y},\tau)$ with constant distortion. □

5. Uniform and coarse nonembeddability into Hilbert space

In this section we prove Theorem 1.3. We shall prove, in fact, that the space $\mathscr{M}_{[0,1]^2,\tau}$ does not embed uniformly or coarsely into L_2. We first recall the defintions of these important notions (see [4, 30] and the references therein for background on these concepts). Let (X,d_X) and (Y,d_Y) be metric spaces. For $f:X\to Y$ and $t>0$ we define

$$\Omega_f(t)=\sup\{d_Y(f(x),f(y));\ d_X(x,y)\le t\},$$

and

$$\omega_f(t)=\inf\{d_Y(f(x),f(y));\ d_N(x,y)\ge t\}.$$

Clearly Ω_f and ω_f are non-decreasing, and for every $x,y\in X$,

$$\omega_f\left(d_X(x,y)\right)\le d_Y(f(x),f(y))\le\Omega_f\left(d_X(x,y)\right).$$

With these definitions, f is uniformly continuous if $\lim_{t\to0}\Omega_f(t)=0$, and f is said to be a uniform embedding if f is invertible and both f and f^{-1} are uniformly continuous. Also, f is said to be a coarse embedding if $\Omega_f(t)<\infty$ for all $t>0$ and $\lim_{t\to\infty}\omega_f(t)=\infty$.

In what follows we will use the following standard notation: Given a sequence of Banach spaces $\left\{(Z_j,\|\cdot\|_{Z_j})\right\}_{j=1}^\infty$ the Banach space $\left(\bigoplus_{j=1}^\infty Z_j\right)_1$ is the space of all sequences $\overline{z}=(z_j)_{j=1}^\infty\in\prod_{j=1}^\infty Z_j$ such that $\|\overline{z}\|:=\sum_{j=1}^\infty\|z_j\|_{Z_j}<\infty$. If for every $j\in\mathbb{N}$, $Z_j=Z_1$, we write $\ell_1(Z_1)=\left(\bigoplus_{j=1}^\infty Z_j\right)_1$.

Theorem 5.1. *The spaces $\left\{\mathscr{M}^0_{\{0,\dots,n\}^2,\tau}\right\}_{n=1}^\infty$ do not admit a uniform or coarse embedding into L_2 with moduli uniformly bounded in n, i.e., there do not exist increasing functions $\omega,\Omega:[0,\infty)\to[0,\infty)$ which either satisfy $\lim_{t\to0}\omega(t)=\lim_{t\to0}\Omega(t)=0$, or $\lim_{t\to\infty}\omega(t)=\infty$, and mappings $f_n:\mathscr{M}^0_{\{0,\dots,n\}^2}\to L_2$, such that $\omega(\|\mu-\nu\|_\tau)\le\|f_n(\mu)-f_n(\nu)\|_2\le\Omega(\|\mu-\nu\|_\tau)$ for all $\mu,\nu\in\mathscr{M}^0_{\{0,\dots,n\}^2}$ and all n.*

Proof. If this is not the case then by passing to a limit along an ultrafilter we easily deduce that $\mathscr{M}^0_{[0,1]^2,\tau}$ uniformly or coarsely embeds in an ultraproduct of Hilbert spaces and thus in L_2 (see [12, 13]). By a theorem of Aharoni, Maurey and Mityagin [1] in the case of uniform embeddings, and a result of Randrianarivony [35] in the case of coarse embeddings, this implies that $\mathscr{M}^0_{[0,1]^2}$ is linearly isomorphic to a subspace of L_0. By a theorem of Nikišin [31] it follows that $\mathscr{M}^0_{[0,1]^2}$ is isomorphic to a subspace of $L_{1-\varepsilon}$ for any $\varepsilon\in(0,1)$. We recall that it is an open problem posed by Kwapien (see the discussion in [22, 4]) whether a Banach space which linearly embed into L_0 is linearly isomorphic to a subspace of L_1. If this were the case, we would have finished by Theorem 1.1. Since the solution of Kwapien's problem is unknown, we proceed as follows.

Let $\{S_j\}_{j=1}^\infty$ be a sequence of disjoint squares in $[0,1]^2$ with

$$d(S_j,S_k)=\min_{a\in S_j,\ b\in S_k}\|a-b\|_2 > \max\left\{\mathrm{diam}S_j,\mathrm{diam}S_k\right\}. \quad (13)$$

Consider the linear subspace Y of $\mathscr{M}^0_{[0,1]^2}$ consisting of all measures μ satisfying $\mathrm{supp}(\mu)\subseteq\bigcup_{j=1}^\infty S_j$ and $\mu(S_j)=0$ for all j. It is intuitively clear that in the computation of $\|\mu\|_\tau$ for $\mu\in Y$ the best transportation leaves each of the S_j invariant; i.e., it is enough to take the infimum in (5) only over measures $\pi\in\Pi(\mu,\nu)$ which are supported on $\bigcup_{j=1}^\infty(S_j\times S_j)$. This is proved formally as follows: Fix $\mu\in Y$ and write $\mu=\sum_{j=1}^\infty\mu_j$, where $\mathrm{supp}(\mu_j)\subseteq S_j$ and $\mu_j(S_j)=0$ for all $j\in\mathbb{N}$. We claim that

$$\|\mu\|_{[0,1]^2,\tau}=\sum_{j=1}^\infty\|\mu_j\|_{S_j,\tau}. \quad (14)$$

If $\pi_j\in\Pi(\mu_j^+,\mu_j^-)$ then $\pi:=\sum_{j=1}^\infty\pi_j\in\Pi(\mu^+,\mu^-)$. Thus $\|\mu\|_{[0,1]^2,\tau}\le\sum_{j=1}^\infty\|\mu_j\|_{S_j,\tau}$. To prove the reverse inequality take $\pi\in\Pi(\mu^+,\mu^-)$. For every $j=1,2,\dots$ define a measure σ_j on S_j as follows: For $A\subseteq S_j$ set $\sigma_j(A):=\pi\left(A\times\bigcup_{k\ne j}S_k\right)$. Thus, in particular, by our assumption (13) for every $y\in S_j$,

$$\int_{S_j}\|x-y\|_2d\sigma_j(x)=\int_{S_j\times\bigcup_{k\ne j}S_k}\|x-y\|_2d\pi(x,z)\le\int_{S_j\times\bigcup_{k\ne j}S_k}\|x-z\|_2d\pi(x,z). \quad (15)$$

Writing

$$\widetilde{\pi}:=\pi\cdot\mathbf{1}_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}+\sum_{j=1}^{\infty}\frac{1}{\sigma_j(S_j)}\cdot\sigma_j\otimes\sigma_j$$
$$=\pi\cdot\mathbf{1}_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}+\sum_{j=1}^{\infty}\frac{1}{\pi\left(S_j\times\bigcup_{k\neq j}S_k\right)}\cdot\sigma_j\otimes\sigma_j,$$

it follows from our definitions that $\widetilde{\pi}\in\Pi(\mu^+,\mu^-)$ and $\widetilde{\pi}$ is supported on $\bigcup_{j=1}^{\infty}(S_j\times S_j)$. Moreover, for each j, $\widetilde{\pi}_j:=\widetilde{\pi}|_{S_j}\in\Pi(\mu_j^+,\mu_j^-)$, so that

$$\begin{aligned}\sum_{j=1}^{\infty}\|\mu_j\|_{S_j,\tau}&\leq\sum_{j=1}^{\infty}\int_{S_j\times S_j}\|x-y\|_2d\widetilde{\pi}_j(x,y)\\ &=\int_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}\|x-y\|_2d\pi(x,y)\\ &\quad+\sum_{j=1}^{\infty}\frac{1}{\pi\left(S_j\times\bigcup_{k\neq j}S_k\right)}\\ &\quad\cdot\int_{S_j\times S_j}\|x-y\|_2d\sigma_j(x)d\sigma_j(y)\\ &\overset{(15)}{\leq}\int_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}\|x-y\|_2d\pi(x,y)\\ &\quad+\sum_{j=1}^{\infty}\int_{S_j\times\bigcup_{k\neq j}S_k}\|x-z\|_2d\pi(x,z)\\ &=\int_{\left(\bigcup_{j=1}^{\infty}S_j\right)\times\left(\bigcup_{j=1}^{\infty}S_j\right)}\|x-y\|_2d\pi(x,y).\end{aligned}$$

This concludes the proof of (14). It follows that Y is isometric to $\left(\bigoplus_{n=1}^{\infty}\mathscr{M}_{S_n,\tau}^0\right)_1$, which in turn is isometric to $\ell_1\left(\mathscr{M}_{[0,1]^2,\tau}^0\right)$. Now, Kalton proved in [22] that if for some Banach space X, $\ell_1(X)$ is isomorphic to a subspace of L_0, then X is isomorphic to a subspace of L_1 and we finish by Theorem 1.1. □

Proof of Theorem 1.3. Assume for the sake of contradiction that there exists $C<\infty$ such that for all $n\in\mathbb{N}$, $c_2\left(\mathscr{P}_{\{0,\dots,n\}^2},\sqrt{\tau}\right)<C$. By the proof of Lemma 3.1 we know that the unit ball of $\mathscr{M}_{\{0,\dots,n\}^2,\tau}$ is isometric to a subset of $(\mathscr{P}_{\{0,\dots,n\}^2},\tau)$. Thus by our assumption there exist mappings $f_n:\mathscr{M}_{\{0,\dots,n\}^2}\to L_2$ such that for every $\mu,\nu\in\mathscr{M}_{\{0,\dots,n\}^2}$ with $\|\mu\|_\tau,\|\nu\|_\tau\leq 1$,

$$\sqrt{\|\mu-\nu\|_\tau}\leq\|f_n(\mu)-f_n(\nu)\|_2\leq C\cdot\sqrt{\|\mu-\nu\|_\tau}\ . \quad (16)$$

Let $\mathscr{U}$ be a free ultrafilter on $\mathbb{N}$. Define $\widetilde{f}_n:\mathscr{M}_{\{0,\dots,n\}^2}\to(L_2)_{\mathscr{U}}$ by $\widetilde{f}_n(\mu)=\left(\sqrt{j}\cdot f_n(\mu/j)\right)_{j=1}^{\infty}/\mathscr{U}$. Inequalities (16) imply that all $\mu,\nu\in\mathscr{M}_{\{0,\dots,n\}^2}$ satisfy $\sqrt{\|\mu-\nu\|_\tau}\leq\|\widetilde{f}_n(\mu)-\widetilde{f}_n(\nu)\|_{(L_2)_{\mathscr{U}}}\leq C\cdot\sqrt{\|\mu-\nu\|_\tau}$. Since the ultrapower $(L_2)_{\mathscr{U}}$ is isometric to a Hilbert space (see [12]), we arrive at a contradiction with Theorem 5.1. □

Remark 5.1. We believe that Theorem 1.3 can be made quantitative, i.e. one can give explicit quantitative estimates on the rate with which $c_2\left(\mathscr{P}_{\{0,\dots,n\}^2},\sqrt{\tau}\right)$ tends to infinity. This would involve obtaining quantitative versions of the proofs in [1, 22, 35], which seems easy but somewhat tedious. We did not attempt to obtain such bounds.

Remark 5.2. We do not know whether $\left(\mathscr{P}_{[0,1]^2},\tau\right)$ admits a uniform embedding into Hilbert space. The proof above actually gives that for all $\alpha\in(0,1]$, $\left(\mathscr{P}_{[0,1]^2,\tau},\tau^\alpha\right)$ does not embed bi-Lipschitzly into Hilbert space. But, our proof exploits the homogeneity of the function $t\mapsto t^\alpha$ in an essential way, so it does not apply to the case of more general moduli.

References

[1] I. Aharoni, B. Maurey, and B. S. Mityagin. Uniform embeddings of metric spaces and of Banach spaces into Hilbert spaces. *Israel J. Math.*, 52(3):251–265, 1985.

[2] A. Archer, J. Fakcharoenphol, C. Harrelson, R. Krauthgamer, K. Talwar, and E. Tardos. Approximate classification via earthmover metrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1079–1087. Society for Industrial and Applied Mathematics, 2004.

[3] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 553–562, New York, NY, USA, 2005. ACM Press.

[4] Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[5] E. Berkson, J. Bourgain, A. Pełczynski, and M. Wojciechowski. Canonical Sobolev projections of weak type $(1,1)$. *Mem. Amer. Math. Soc.*, 150(714):viii+75, 2001.

[6] J. Bourgain. Remarks on the extension of Lipschitz maps defined on discrete sets and uniform homeomorphisms. In *Geometrical aspects of functional analysis (1985/86)*, volume 1267 of *Lecture Notes in Math.*, pages 157–167. Springer, Berlin, 1987.

[7] M. S. Charikar. Similarity estimation techniques from rounding algorithms. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 380–388. ACM Press, 2002.

[8] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *SODA '01: Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 109–118. Society for Industrial and Applied Mathematics, 2001.

[9] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *SoCG '04: Proceedings of the Twentieth Annual Symposium on Computational Geometry*, pages 253–262, New York, NY, USA, 2004. ACM Press.

[10] L. J. Guibas, Y. Rubner, and C. Tomassi. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[11] L. J. Guibas, Y. Rubner, and C. Tomassi. A metric for distributions with applications to image databases. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, pages 59–66, 2003.

[12] S. Heinrich. Ultraproducts in Banach space theory. *J. Reine Angew. Math.*, 313:72–104, 1980.

[13] S. Heinrich and P. Mankiewicz. Applications of ultrapowers to the uniform and Lipschitz classification of Banach spaces. *Studia Math.*, 73(3):225–251, 1982.

[14] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE Computer Society, 2001.

[15] P. Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *41st Annual Symposium on Foundations of Computer Science*, pages 189–197. IEEE Computer Society, 2001.

[16] P. Indyk. Algorithms for dynamic geometric problems over data streams. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 373–380, New York, NY, USA, 2004. ACM Press.

[17] P. Indyk. Nearest neighbors in high-dimensional spaces. In *Handbook of discrete and computational geometry, second edition*, pages 877–892. CRC Press, Inc., Boca Raton, FL, USA, 2004.

[18] P. Indyk and J. Matoušek. Low distortion embeddings of finite metric spaces. In *Handbook of discrete and computational geometry, second edition*, pages 177–196. CRC Press, Inc., Boca Raton, FL, USA, 2004.

[19] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 604–613, New York, NY, USA, 1998. ACM Press.

[20] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *ICCV '03: Proceedings of the 3rd International Workshop on Statistical and Computational Theories of Vision*, 2003.

[21] S. Kakutani. Concrete representation of abstract (L)-spaces and the mean ergodic theorem. *Ann. Math. (2)*, 42:523–537, 1941.

[22] N. J. Kalton. Banach spaces embedding into L_0. *Israel J. Math.*, 52(4):305–319, 1985.

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

[24] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems, and embeddability of negative type metrics into L_1. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 53–62.

[25] S. Kislyakov. Sobolev imbedding operators and the nonisomorphism of certain Banach spaces. *Funct. Anal. Appl.*, 9:290–294, 1975.

[26] J. Lindenstrauss and H. P. Rosenthal. The $\mathscr{L}_p$ spaces. *Israel J. Math.*, 7:325–349, 1969.

[27] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1979. Function spaces.

[28] J. Matoušek. Open problems on embeddings of finite metric spaces. Available at `http://kam.mff.cuni.cz/$\sim$matousek/metrop.ps.gz`.

[29] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[30] M. Mendel and A. Naor. Metric cotype. To appear in Ann. Math., 2005.

[31] E. M. Nikišin. A resonance theorem and series in eigenfunctions of the Laplace operator. *Izv. Akad. Nauk SSSR Ser. Mat.*, 36:795–813, 1972.

[32] A. Pelczynski. Boundedness of the canonical projection for Sobolev spaces generated by finite families of linear differential operators. In *Analysis at Urbana. Vol. 1: Analysis in function spaces, Proc. Spec. Year Mod. Anal./Ill. 1986-87, Lond. Math. Soc. Lect. Note Ser. 137, 395-415* . 1989.

[33] A. Pełczyński and M. Wojciechowski. Sobolev spaces. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1361–1423. North-Holland, Amsterdam, 2003.

[34] S. Peleg, M. Werman, and H. Rom. A unified approach to the change of resolution: space and gray-level. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):739–742, 1989.

[35] N. L. Randrianarivony. Characterization of quasi-Banach spaces which coarsely embed into a Hilbert space. *Proc. Amer. Math. Soc.*, 134(5):1315–1317 (electronic), 2006.

[36] M. Ribe. On uniformly homeomorphic normed spaces. *Ark. Mat.*, 14:237–244, 1976.

[37] D. N. Rockmore. Efficient computation of Fourier inversion for finite groups. *J. Assoc. Comput. Mach.*, 41(1):31–66, 1994.

[38] W. Rudin. *Real and complex analysis. 3rd ed.* New York, NY: McGraw-Hill., 1987.

[39] M. Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[40] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow, 1989.

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

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