# Planar Earthmover is not in $L_1$

Assaf Naor
Microsoft Research
anaor@microsoft.com

Gideon Schechtman*
Weizmann Institute and Microsoft Research
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 $\{0,1,\ldots,n\}^2 \subseteq \mathbb{R}^2$ 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

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

Observe that if $\mu$ and $\nu$ are the uniform probablity distribution over $k$-point subsets $A \subseteq X$ and $B \subseteq X$, respectively, then by the fact that all the extreme points of the $k\times k$ doubly stochastic matrices are permutation matrices,

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

This quantity is also known as the *minimum weight matching* between $A$ and $B$, corresponding to the weight function $d_X(\cdot,\cdot)$ (see [42]). Thus, the Earthmover distance is a natural measure of similarity between images [42, 15, 14]- 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 [42, 15, 14, 26], they are used as LP relaxations for classification problems such as 0-extension and metric labelling [9, 8, 2]. Transportation cost metrics are also prevalent in several areas of analysis and PDE (see the book [53] and the references therein).

Following extensive work on nearest neighbor search and data stream computations for $L_1$ metrics (see [24, 20, 19, 10, 22]), 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 $f : [0,1] \to \mathbb{R}$,

*Supported in part by the Israel Science Foundation.

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 [36, 23, 21] 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. [35]) whether the Earthmover distance embeds into $L_1$ with constant distortion (the best known upper bounds on the $L_1$ distortion were obtained in [8, 26], and will be discussed further below). In [30] 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 the Euclidean plane, as this corresponds to a natural similarity measure between images [14] (indeed, the case of the $L_1$ embeddability of planar Earthmover distance was explicitly asked in [35]). Here we settle this problem negatively by obtaining the first 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

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

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

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

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

We note that the best known upper bound for $c_1\left(\mathscr{P}_{\{0,1,\ldots,n\}^2}, \tau\right)$ is $O(\log n)$, as proved in [8, 26]. Later on (see Theorem 1.4) we will show a new embedding which achieves this bound.

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 [32]. 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 $n$-point subsets of $\left\{0, 1 \ldots, O\left(\sqrt{n \log\log n}\right)\right\}^2$, with $|\mathscr{Y}| \leq n^{O(\log\log n)}$, such that any $L_1$ embedding of $\mathscr{Y}$, equipped with the minimum weight matching metric $\tau$, 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, 31]). It turns out that planar Earthmover does not embed into any squared $L_2$ metric (see Remark 3.3 for a more general result):

**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 [37]). 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

$$\forall (a,b),(u,v) \in \mathbb{Z}_n^2,\ e_{uv}(a,b) \coloneqq e^{\frac{2\pi i(au+bv)}{n}}, \quad \text{and} \quad \forall (u,v)\in\mathbb{Z}_n^2,\ \widehat{\mu}(u,v) \coloneqq \frac{1}{n^2}\sum_{(a,b)\in\mathbb{Z}_n^2} \mu(a,b) e_{uv}(-a,-b).$$

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 [45]. Motivated in part by the results of [40] (see also [5, 41]), 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 $L_1\left(\mathbb{Z}_n^2\right) \oplus L_1\left(\mathbb{Z}_n^2\right)$ has distortion $O(\log n)$.*

The $O(\log n)$ distortion in Theorem 1.4 matches the best known distortion guarantee proved in [26, 8]. Our embedding has various nice features. First of all, it is a simple closed-form 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 (see the "Discussion and open problems" section).

## 2 Preliminaries and notation

For the necessary background on measure theory we refer to the book [46], 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 [53] 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 $\mu$ and $\nu$, i.e. all non-negative Borel measures $\pi$ on $X \times X$ such that for every measurable bounded $f : X \to \mathbb{R}$,

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

Observe that in the case of finitely supported measures, this condition translates to the standard formulation, in which we require that the marginals of $\pi$ are $\mu$ and $\nu$, i.e.

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

The *transportation cost distance* between $\mu$ and $\nu$, 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

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

For $\mu \in \mathscr{M}_X^0$, $\mu^+(X) = \mu^-(X)$, so we may write $\|\mu\|_\tau \coloneqq \tau(\mu^+, \mu^-)$. This is easily seen to be a norm on the vector space $\mathscr{M}_{X,\tau}^0 \coloneqq \left\{\mu \in \mathscr{M}_X^0 : \ \|\mu\|_\tau < \infty\right\}$. It is easy to check that for any two non-negative measures $\mu, \nu$ with the same total mass, and for every non-negative measure $\sigma$, we have $\tau(\mu+\sigma, \nu+\sigma) = \tau(\mu,\nu)$. It follows that

$$\|\mu - \nu\|_\tau = \tau\left((\mu-\nu)^+, (\mu-\nu)^-\right) = \tau\left((\mu-\nu)^+ + \min\{\mu,\nu\}, (\mu-\nu)^- + \min\{\mu,\nu\}\right) = \tau(\mu,\nu). \tag{6}$$

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 [53]) 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 $\mu \in \mathscr{M}_{X,\tau}^0$,

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

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

# 3 Proof of Theorem 1.1

Fix an integer $n \ge 2$ and denote $X = \{0, 1, \ldots, n-1\}^2$, equipped with the standard Euclidean metric. In what follows, for concreteness, $\mathrm{Lip}_0 \coloneqq \mathrm{Lip}_0(X)$ 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 $\mathrm{Lip}_0$ 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 any two probability measures $\mu, \nu \in \mathscr{P}_X$,

$$\tau(\mu,\nu) \le \|F(\mu) - F(\nu)\|_1 \le L \cdot \tau(\mu,\nu). \tag{7}$$

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 $\ell_1^N$)**.** *Under the assumption of an existence of an embedding $F : \mathscr{P}_X \to L_1$ satisfying* (7)*, 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 $\psi(\mu)(x) := \frac{\mu(x)+1}{n^2}$. It is clear that for every $\mu, \nu \in \mathscr{M}$, $\|\mu - \nu\|_\tau = n^2 \|\psi(\mu) - \psi(\nu)\|_\tau$. The mapping $h := \frac{1}{n^2} \cdot F \circ \psi : B_{\mathscr{M}} \to L_1$ satisfies $h(0) = 0$, $\|h\|_{\mathrm{Lip}} \le L$, and $\|h(\mu) - h(\nu)\|_1 \ge \|\mu - \nu\|_\tau$, where we have used (7) and (6). 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 [16] 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 $\sigma$). 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

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

It is straightforward to check that $\|\chi(f) - \chi(g)\|_{L_1([0,1]\times\mathbb{R})} = \|f - g\|_1$ 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

$$\begin{aligned}
\left\|\tilde{h}(\mu)-\tilde{h}(\nu)\right\|_{L_1([0,1]\times\mathbb{R})} &= \|\nu\|_\tau \cdot \left\| h\left(\frac{\mu}{\|\mu\|_\tau}\right) - h\left(\frac{\nu}{\|\nu\|_\tau}\right)\right\|_1 + (\|\mu\|_\tau - \|\nu\|_\tau)\cdot\left\|h\left(\frac{\mu}{\|\mu\|_\tau}\right)\right\|_1 \qquad (8)\\
&\geq \|\nu\|_\tau \cdot \left\|\frac{\mu}{\|\mu\|_\tau} - \frac{\nu}{\|\nu\|_\tau}\right\|_\tau + \|\mu\|_\tau - \|\nu\|_\tau\\
&\geq \|\nu-\mu\|_\tau - \left\|\mu - \frac{\|\nu\|_\tau}{\|\mu\|_\tau}\mu\right\|_\tau + \|\mu\|_\tau - \|\nu\|_\tau\\
&= \|\nu-\mu\|_\tau.
\end{aligned}$$

It also follows from the identity (8) that

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

We are now in position to use a Theorem of Ribe [44] (see also [17], 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 $\|S\| \leq L$ and $\|S^{-1}\| \leq 1$. Since $\mathscr{M}$ is finite dimensional, by the principle of local reflexivity [33] (alternatively by Kakutani's representation theorem [27, 34]), 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 $\|T\| \leq 2L$ and $\|T^{-1}\| \leq 1$ (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 [50] we can ensure that $N = O(n^2 \log n)$). □

**Remark 3.1.** The first argument above is not special for $L_1$, and can be generalized to show that for every finite dimensional Banach space $Y$, if the unit ball of $Y$ admits a bi-Lipschitz embedding into a Banach space $Z$, then $Y$ embeds linearly with the same distortion into $Z$.

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. $\|T^*\| \leq 2L$ and the image of the unit ball of $\ell_\infty^N$ under $T^*$ contains the unit ball of $\mathrm{Lip}_0$.

The rest of the proof follows Kislyakov's [32] 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 $\mathrm{Lip}_0$ to $\ell_2(X)$, and to prove two properties of the composed map: Using the fact that $\|T^*\| \leq 2L$ we shall show that $\mathscr{F} \circ T^*$ is *order bounded* with good bound, that is,

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

for some $x \in \ell_2(X)$ such that $\|x\|_2 \leq 4Ln$. 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 $\mathscr{F}(T^*(\phi_i))) \leq x$ for all $i \in I$ then necessarily $\|x\|_2 \geq cn\sqrt{\log n}$, 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

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

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

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

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 [41] (with non-optimal constant), which can be proved as follows. First of all, since $f(0) = 0$, for every $(u,v) \in X$,

$$\begin{aligned} |f(u,v)| &= \left|\sum_{k=0}^{u-1}[f(k+1,v) - f(k,v)] + \sum_{\ell=0}^{v-1}[f(0,\ell+1) - f(0,\ell)]\right| \\ &\le \sum_{k=0}^{n-2}|f(k+1,v) - f(k,v)| + \sum_{\ell=0}^{n-2}|f(0,\ell+1) - f(0,\ell)| := A(v). \end{aligned} \tag{9}$$

Analogously,

$$|f(u,v)| \le \sum_{\ell=0}^{n-2}|f(u,\ell+1) - f(u,\ell)| + \sum_{k=0}^{n-2}|f(k+1,0) - f(k,0)| := B(u). \tag{10}$$

Multiplying (9) and (10), and summing over $X$, we see that

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

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 [51, 54] for more information on this topic). Given two Banach spaces $Y$ and $Z$, the $\pi_1$ 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

$$\sum_{j=1}^{m} \|Ay_j\|_Z \leq K \sum_{j=1}^{m} |y^*(y_j)|. \tag{11}$$

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 $\pi_1(QAP) \leq \|Q\| \cdot \pi_1(A) \cdot \|P\|$. Observe that it is always the case that $\pi_1(A) \geq \|A\|$.

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

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

*Proof.* Fix $f_1, \ldots, f_m : X \to \mathbb{R}$ with $f_1(0) = \cdots = f_m(0) = 0$. Then

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

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 $0 \leq s_0 \leq n-1$ and $0 \leq t_0 \leq n-2$. 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 $\pi_1$ norm is the Pietsch Factorization Theorem (see [51]), 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 $\sigma$ on $\{1, \ldots, N\}$ and a linear operator $R : L_1(\sigma) \to Y$ such that $A = R \circ I$, where $I$ is the formal identity from $\ell_\infty^N$ to $L_1(\sigma)$, and $\|R\| = \pi_1(A)$.*

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

$$\sum_{j=1}^{m} \|Ax_j\| \leq \pi_1(A) \cdot \sup_{\substack{x^* \in (\ell_\infty^N)^* \\ \|x^*\|=1}} \sum_{j=1}^{m} |x^*(x_j)| = \pi_1(A) \cdot \max_{1 \leq k \leq N} \sum_{j=1}^{m} |x_j(k)|, \tag{12}$$

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 $e_1, \ldots, e_N$ the standard basis of $\mathbb{R}^N$ we deduce from (12) that $\pi_1(A) \geq \sum_{j=1}^N \|Ae_j\|$. Define a probability measure $\sigma$ on $\{1, \ldots, N\}$ by $\sigma(k) = \frac{\|Ae_k\|}{\sum_{j=1}^N \|Ae_j\|}$. Then for every $x \in \ell_\infty^N$ we see that

$$\|Ax\| = \left\| \sum_{k=1}^N x(k) Ae_k \right\| \leq \sum_{k=1}^N |x(k)| \cdot \|Ae_k\| = \left( \sum_{j=1}^N \|Ae_j\| \right) \int_{\{1,\ldots,N\}} |x(k)| d\sigma(k) \leq \pi_1(A) \int_{\{1,\ldots,N\}} |x(k)| d\sigma(k).$$

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

From now on let $R$ and $\sigma$ 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 $\|R\| \leq 4nL$. Schematically, we have the following commuting diagram:

$$\begin{array}{ccccccccc} \ell_\infty^N & \xrightarrow{T^*} & \mathrm{Lip}_0 & \xrightarrow{\mathrm{Id}} & W & \xrightarrow{S} & \ell_2(X) & \xrightarrow{\mathscr{F}} & \ell_2(X) \\ & \searrow^{I} & & & & & \nearrow_{R} & & \\ & & L_1(\sigma) & & & & & & \end{array}$$

We need only one more simple result from classical Banach space theory. This result can be generalized to the case in which the target space $\ell_2$ is replaced by a more general Banach lattice. For the sake of simplicity 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 $f : \{0, \ldots, N\} \to [0, \infty)$. Then there is $x \in \ell_2$ with non-negative coordinates such that*

$$R\left(\{g : \{0, \ldots, N\} \to \mathbb{R} : \ \forall j, \ |g(j)| \leq f(j)\}\right) \subseteq \{y \in \ell_2 : \ \forall j, \ |y_j| \leq x_j\},$$

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

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

$$\|R\| = \max_{1 \leq i \leq N} \left( \frac{1}{\sigma(i)^2} \sum_{j=1}^\infty R_{ij}^2 \right)^{1/2}. \tag{13}$$

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

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

Now,

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

where we have used (13). □

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

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

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

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 $\ell_\infty^N$ onto the unit ball of $\mathrm{Lip}_0$, 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

$$\begin{aligned}
R \circ I(\phi_{u,v})(u,v) &= \mathscr{F} \circ S \circ \mathrm{Id} \circ T^*(\phi_{u,v})(u,v) \\
&= \mathscr{F}(\varphi_{u,v})(u,v) \\
&= \frac{1}{n^2} \sum_{(k,\ell)\in X} \frac{1}{u+v} \cdot \sin^2\left(\frac{2\pi(uk+v\ell)}{n}\right) \\
&= \frac{1}{n^2(u+v)} \sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1} \left(\frac{1}{2} - \frac{1}{4}\cdot e^{2\pi i\cdot\frac{2uk}{n}} \cdot e^{2\pi i\cdot\frac{2v\ell}{n}} - \frac{1}{4}\cdot e^{-2\pi i\cdot\frac{2uk}{n}} \cdot e^{-2\pi i\cdot\frac{2v\ell}{n}}\right) \\
&= \begin{cases} \frac{1}{2(u+v)} & (u,v) \neq \left(\frac{n}{2},\frac{n}{2}\right), \\ 0 & (u,v) = \left(\frac{n}{2},\frac{n}{2}\right). \end{cases}
\end{aligned}$$

But

$$(16\pi L)^2 \ge \|x\|_2^2 \ge \sum_{u,v=1}^{n-1} x_{u,v}^2 \ge \sum_{u,v=1}^{n-1} \left[R \circ I(\phi_{u,v})(u,v)\right]^2 \ge \frac{1}{8}\sum_{u,v=1}^{n-1} \frac{1}{(u+v)^2} \ge \frac{\log n}{16},$$

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

### 3.1 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 [44].

**Theorem 3.6** (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 $\varepsilon$-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

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

On the other hand, as we have seen in the proof of Lemma 3.1, for every $\mu \in \mathscr{M}$, $\|\mu\|_\infty \leq \|\mu\|_\tau$. It follows from these consideration, and Theorems 1.1 and 3.6, that for every integer $N \geq e^{e^{C'n^2 \log\log n}}$, 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 $S_\mu \subseteq \{0, \ldots, nM\}^2$ as follows. For every $(u, v) \in X = \{0, \ldots, n-1\}^2$, if $\mu(u, v) = k/N$, where $k \in \{0, \ldots, N\}$, then $S_\mu$ will contain arbitrary $k$ distinct points from the set $(uM, vM) + \left\{0, \ldots, \left\lceil \sqrt{N} \right\rceil\right\}^2$. Provided $M \geq 4\sqrt{N}$, the sets $\{S_\mu\}_{\mu \in \mathscr{Y}}$ thus obtained are $N$ point subsets of $\{0, \ldots, nM\}^2$, 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. □

## 3.2 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, 37] 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

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

and

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

Clearly $\Omega_f$ and $\omega_f$ are non-decreasing, and for every $x, y \in X$,

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

With these definitions, $f$ is uniformly continuous if $\lim_{t\to 0} \Omega_f(t) = 0$, and $f$ is said to be a uniform embedding if $f$ is injective 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 $\bar{z} = (z_j)_{j=1}^\infty \in \prod_{j=1}^\infty Z_j$ such that $\|\bar{z}\| := \sum_{j=1}^\infty \|z_j\|_{Z_j} < \infty$. 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 3.7.** *The spaces $\left\{\mathscr{M}^0_{\{0,\ldots,n\}^2,\tau}\right\}_{n=1}^\infty$ 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 $\lim_{t\to 0} \omega(t) = \lim_{t\to 0} \Omega(t) = 0$, or $\lim_{t\to\infty} \omega(t) = \infty$, and mappings $f_n : \mathscr{M}^0_{\{0,\ldots,n\}^2} \to L_2$, such that $\omega(\|\mu - \nu\|_\tau) \leq \|f_n(\mu) - f_n(\nu)\|_2 \leq \Omega(\|\mu - \nu\|_\tau)$ for all $\mu, \nu \in \mathscr{M}^0_{\{0,\ldots,n\}^2}$ 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 [16, 17]). By a theorem of Aharoni, Maurey and Mityagin [1] in the case of uniform embeddings, and a result of Randrianarivony [43] 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 [39] 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 [28, 4]) whether a Banach space which linearly embeds 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

$$d(S_j, S_k) = \min_{a\in S_j,\ b\in S_k} \|a-b\|_2 > \max\left\{\mathrm{diam} S_j, \mathrm{diam} S_k\right\}. \tag{14}$$

Consider the linear subspace $Y$ of $\mathscr{M}^0_{[0,1]^2}$ consisting of all measures $\mu$ 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

$$\|\mu\|_{[0,1]^2,\tau} = \sum_{j=1}^{\infty} \|\mu_j\|_{S_j,\tau}. \tag{15}$$

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 $j = 1,2,\ldots$ define a measure $\sigma_j$ on $S_j$ as follows: For $A \subseteq S_j$ set $\sigma_j(A) := \pi\left(A \times \bigcup_{k\neq j} S_k\right)$. Thus, in particular, by our assumption (14) for every $y \in S_j$,

$$\int_{S_j}\|x-y\|_2 d\sigma_j(x) = \int_{S_j\times\bigcup_{k\neq j}S_k}\|x-y\|_2 d\pi(x,z) \le \int_{S_j\times\bigcup_{k\neq j}S_k}\|x-z\|_2 d\pi(x,z). \tag{16}$$

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

$$\begin{aligned}
\sum_{j=1}^{\infty}\|\mu_j\|_{S_j,\tau} &\le \sum_{j=1}^{\infty}\int_{S_j\times S_j}\|x-y\|_2 d\widetilde{\pi}_j(x,y)\\
&= \int_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}\|x-y\|_2 d\pi(x,y) + \sum_{j=1}^{\infty}\frac{1}{\pi\left(S_j\times\bigcup_{k\neq j}S_k\right)}\cdot\int_{S_j\times S_j}\|x-y\|_2 d\sigma_j(x)d\sigma_j(y)\\
&\overset{(16)}{\le} \int_{\bigcup_{j=1}^{\infty}(S_j\times S_j)}\|x-y\|_2 d\pi(x,y) + \sum_{j=1}^{\infty}\int_{S_j\times\bigcup_{k\neq j}S_k}\|x-z\|_2 d\pi(x,z)\\
&= \int_{\left(\bigcup_{j=1}^{\infty}S_j\right)\times\left(\bigcup_{j=1}^{\infty}S_j\right)}\|x-y\|_2 d\pi(x,y).
\end{aligned}$$

This concludes the proof of (15). It follows that $Y$ is isometric to $\left(\bigoplus_{n=1}^{\infty}\mathscr{M}^0_{S_n,\tau}\right)_1$, which in turn is isometric to $\ell_1\left(\mathscr{M}^0_{[0,1]^2,\tau}\right)$. Now, Kalton proved in [28] 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 $n \in \mathbb{N}$, $c_2\left(\mathscr{P}_{\{0,\ldots,n\}^2},\sqrt{\tau}\right) < C$. By the proof of Lemma 3.1 we know that the unit ball of $\mathscr{M}_{\{0,\ldots,n\}^2,\tau}$ is isometric to

a subset of $(\mathscr{P}_{\{0,\ldots,n\}^2}, \tau)$. Thus by our assumption there exist mappings $f_n : \mathscr{M}_{\{0,\ldots,n\}^2} \to L_2$ such that for every $\mu, \nu \in \mathscr{M}_{\{0,\ldots,n\}^2}$ with $\|\mu\|_\tau, \|\nu\|_\tau \leq 1$,

$$\sqrt{\|\mu - \nu\|_\tau} \leq \|f_n(\mu) - f_n(\nu)\|_2 \leq C \cdot \sqrt{\|\mu - \nu\|_\tau} \ . \tag{17}$$

Let $\mathscr{U}$ be a free ultrafilter on $\mathbb{N}$. Define $\widetilde{f_n} : \mathscr{M}_{\{0,\ldots,n\}^2} \to (L_2)_{\mathscr{U}}$ by $\widetilde{f_n}(\mu) = \left(\sqrt{j} \cdot f_n(\mu/j)\right)_{j=1}^{\infty} /\mathscr{U}$. Inequalities (17) imply that all $\mu, \nu \in \mathscr{M}_{\{0,\ldots,n\}^2}$ 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 [16]), we arrive at a contradiction with Theorem 3.7. □

**Remark 3.2.** We believe that Theorem 1.3 can be made quantitative, i.e. one can give explicit quantitative estimates on the rate with which $c_2\left(\mathscr{P}_{\{0,\ldots,n\}^2}, \sqrt{\tau}\right)$ tends to infinity. This would involve obtaining quantitative versions of the proofs in [1, 28, 43], which seems easy but somewhat tedious. We did not attempt to obtain such bounds.

**Remark 3.3.** 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 $\alpha \in (0, 1]$, $\left(\mathscr{P}_{[0,1]^2}, \tau^\alpha\right)$ 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.

# 4 Upper bounds via Fourier analysis

In this section we prove Theorem 1.4, and discuss some related upper bounds. Given a measure $\mu$ on $\mathbb{Z}_n^2$ we decompose it as in (2), and we consider the linear operators $A$ and $B$, from $\mathscr{M}_{\mathbb{Z}_n^2}$ to $L_1\left(\mathbb{Z}_n^2\right)$, defined in (3) and (4), respectively. One checks that the duals of these operators, $A^*, B^* : L_\infty\left(\mathbb{Z}_n^2\right) \to \mathscr{M}^*_{\mathbb{Z}_n^2} = \mathrm{Lip}_0\left(\mathbb{Z}_n^2\right)$, are given by

$$A^* f = \sum_{(u,v) \in \mathbb{Z}_n^2 \setminus \{(0,0)\}} \frac{e^{-\frac{2\pi i u}{n}} - 1}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u, v) \cdot (e_{uv} - 1), \tag{18}$$

and

$$B^* f = \sum_{(u,v) \in \mathbb{Z}_n^2 \setminus \{(0,0)\}} \frac{e^{-\frac{2\pi i v}{n}} - 1}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u, v) \cdot (e_{uv} - 1). \tag{19}$$

To check these identities the reader should verify that for all $\mu \in \mathscr{M}_{\mathbb{Z}_n^2}$, $\int_{\mathbb{Z}_n^2} f d(A\mu) = \int_{\mathbb{Z}_n^2} (A^* f) d\mu$, and similarly for $B$ (to this end, recall that $\mu\left(\mathbb{Z}_n^2\right) = 0$, so that $\widehat{\mu}(0, 0) = 0$. This explains the subtraction of 1 in the identities (18) and (19)).

We claim that for every $\mu \in \mathscr{M}_{Z_n^2}$,

$$\|\mu\|_\tau \leq \|A\mu\|_{L_1(\mathbb{Z}_n^2)} + \|B\mu\|_{L_1(\mathbb{Z}_n^2)} \leq C \log n \cdot \|\mu\|_\tau, \tag{20}$$

where $C$ is a universal constant. This will imply Theorem 1.4 since the mapping $\mu \mapsto \mu - U$, where $U$ is the uniform probability measure on $Z_n^2$, is an isometric embedding of $\mathscr{P}_{\mathbb{Z}_n^2}$ into $\mathscr{M}_{\mathbb{Z}_2^n}$.

By duality, (20) is equivalent to the fact that the mapping $(f,g) \mapsto A^*f + B^*g$ from $L_\infty\left(\mathbb{Z}_n^2\right) \oplus L_\infty\left(\mathbb{Z}_n^2\right)$ to $\mathrm{Lip}_0\left(\mathbb{Z}_n^2\right)$ is a $C \log n$ quotient map, i.e. for every $(f,g) \in L_\infty\left(\mathbb{Z}_n^2\right) \oplus L_\infty\left(\mathbb{Z}_n^2\right)$

$$\left\|A^*f + B^*g\right\|_{\mathrm{Lip}} \leq C \log n \cdot \max\left\{\|f\|_\infty, \|g\|_\infty\right\}, \tag{21}$$

and for every $h \in \mathrm{Lip}_0\left(\mathbb{Z}_n^2\right)$ there is some $(f,g) \in L_\infty\left(\mathbb{Z}_n^2\right) \oplus L_\infty\left(\mathbb{Z}_n^2\right)$ satisfying $A^*f + B^*g = h$ and $\max\{\|f\|_\infty, \|g\|_\infty\} \leq \|h\|_{\mathrm{Lip}}$. The second assertion is proved as follows: Take $f = \partial_1 h$ and $g = \partial_2 h$, where for $j = 1, 2$, $\partial_j h(x) = h(x + e_j) - h(x)$ (here $e_1 = (1,0)$ and $e_2 = (0,1)$). Clearly $\|f\|_\infty, \|g\|_\infty \leq \|h\|_{\mathrm{Lip}}$, and

$$\begin{aligned}
A^*f + B^*g &= \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \left( \frac{\left(e^{-\frac{2\pi i u}{n}} - 1\right) \cdot \widehat{\partial_1 h}(u,v) + \left(e^{-\frac{2\pi i v}{n}} - 1\right) \cdot \widehat{\partial_2 h}(u,v)}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \right) (e_{uv} - 1) \\
&= \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \left( \frac{\left(e^{-\frac{2\pi i u}{n}} - 1\right) \cdot \left(e^{\frac{2\pi i u}{n}} - 1\right) + \left(e^{-\frac{2\pi i v}{n}} - 1\right) \cdot \left(e^{\frac{2\pi i v}{n}} - 1\right)}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \right) \cdot \widehat{h}(u,v)(e_{uv} - 1) \\
&= \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \widehat{h}(u,v) e_{uv} - \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \widehat{h}(u,v) \\
&= \sum_{(u,v)\in\mathbb{Z}_n^2} \widehat{h}(u,v) e_{uv} = h,
\end{aligned}$$

where we used the fact that $h(0) = 0$.

It remains to prove (21). To this end, it is enough to show that $\|A^*f\|_{\mathrm{Lip}} \leq O(\log n) \cdot \|f\|_\infty$ and $\|B^*g\|_{\mathrm{Lip}} \leq O(\log n) \cdot \|g\|_\infty$. We will establish this for $A^*$- the case of $B^*$ is entirely analogous. Observe that

$$\|A^*f\|_{\mathrm{Lip}} \leq \|\partial_1 A^*f\|_\infty + \|\partial_2 A^*f\|_\infty,$$

so it is enough to establish the following two inequalities:

$$\left\| \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u,v) e_{uv} \right\|_\infty \leq O(\log n) \cdot \|f\|_\infty, \tag{22}$$

and

$$\left\| \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{\left(e^{-\frac{2\pi i u}{n}} - 1\right) \cdot \left(e^{\frac{2\pi i v}{n}} - 1\right)}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u,v) e_{uv} \right\|_\infty \leq O(\log n) \cdot \|f\|_\infty. \tag{23}$$

Since for $p > 0$ the norms on $L_\infty\left(\mathbb{Z}_n^2\right)$ and $L_p\left(\mathbb{Z}_n^2\right)$ are equivalent with constant $n^{2/p}$ (by Hölder's inequality), it is enough to show that for $p \geq 2$,

$$\left\| \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u,v) e_{uv} \right\|_p \leq O(p) \cdot \|f\|_p, \tag{24}$$

and

$$\left\| \sum_{(u,v)\in\mathbb{Z}_n^2\setminus\{(0,0)\}} \frac{\left(e^{-\frac{2\pi i u}{n}} - 1\right) \cdot \left(e^{\frac{2\pi i v}{n}} - 1\right)}{\left|e^{\frac{2\pi i u}{n}} - 1\right|^2 + \left|e^{\frac{2\pi i v}{n}} - 1\right|^2} \cdot \widehat{f}(u,v) e_{uv} \right\|_p \leq O(p) \cdot \|f\|_p. \tag{25}$$

To prove inequalities (24) and (25) we will assume that $n$ is odd (all of our results are valid for even $n$ as well, and the proofs in this case require minor modifications). We think of $\mathbb{Z}_n^2$ as $[-(n-1)/2, (n-1)/2]^2 \cap \mathbb{Z}^2$. As before, given $m : \mathbb{Z}_n^2 \to \mathbb{C}$ we denote

$$\partial_1 m(x,y) = m(x+1,y) - m(x,y), \quad \text{and} \quad \partial_2 m(x,y) = m(x,y+1) - m(x,y).$$

Thus

$$\partial_1^2 m(x,y) = m(x+2,y) - 2m(x+1,y) + m(x,y) \quad \text{and} \quad \partial_2^2 m(x,y) = m(x,y+2) - 2m(x,y+1) + m(x,y),$$

and

$$\partial_1\partial_2 m(x,y) = \partial_2\partial_1 m(x,y) = m(x+1,y+1) - m(x+1,y) - m(x,y+1) + m(x,y).$$

In what follows we think of $m$ as a *Fourier multiplier* in the sense that it corresponds to a translation invariant operator $T_m$ on $L_2\left(\mathbb{Z}_n^2\right)$ given by

$$T_m(f) \coloneqq \sum_{(u,v)\in\mathbb{Z}_n^2} m(u,v)\cdot \widehat{f}(u,v)\cdot e_{uv}. \tag{26}$$

Recall that an operator $T : L_1\left(\mathbb{Z}_n^2\right) \to L_1\left(\mathbb{Z}_n^2\right)$ is said to be weak $(1,1)$ with constant $K$ if for every $f : \mathbb{Z}_n^2 \to \mathbb{C}$ and every $a > 0$,

$$\left|\left\{(u,v) \in \mathbb{Z}_n^2 : \ |Tf(u,v)| \geq a\right\}\right| \leq \frac{K}{a}\cdot \|f\|_1 = \frac{K}{a}\cdot \sum_{(u,v)\in\mathbb{Z}_n^2} |f(u,v)|.$$

We will use the following discrete version of the Hörmander-Mihlin multiplier theorem [38, 18].

**Theorem 4.1** (Hörmander-Mihlin multiplier criterion on $\mathbb{Z}_n^2$)**.** *For $j \in \mathbb{N}$ denote $Q_j = [-2^j, 2^j] \times [-2^j, 2^j]$. Fix $B > 0$ and $m : \mathbb{Z}_n^2 \to \mathbb{C}$ with $m(0,0) = 0$, and assume that for all $j = 0, 1, \ldots, \lfloor \log_2(n-1)\rfloor - 1$,*

$$\sum_{(u,v)\in(Q_j\setminus Q_{j-1})\cap\mathbb{Z}_n^2} \Big[2^{-2j}|m(u,v)|^2 + |\partial_1 m(u,v)|^2 + |\partial_2 m(u,v)|^2 + 2^{2j}|\partial_1^2 m(u,v)|^2 + 2^{2j}|\partial_2^2 m(u,v)|^2 + 2^{2j}|\partial_1\partial_2 m(u,v)|^2\Big] \leq B^2.$$

*Then the translation invariant operator $T_m$ corresponding to $m$ is weak $(1,1)$ with constant $O(B)$.*

While the continuous version of the Hörmander-Mihlin multiplier theorem is a powerful tool which appears in several texts (e.g. in the books [12, 48, 52]), we could not locate a statement of the above discrete version in the literature. It is, however, possible to prove it using several minor modifications of the existing proofs. The standard proof of the Hörmander-Mihlin criterion is usually split into two parts. The first part, which is based on the Calderón-Zygmund decomposition, transfers virtually verbatim to the discrete setting- see Theorem 3 in Chapter 1 of [48], and Remark 8.1 there which explains how this part of the proof transfers from $\mathbb{R}^n$ to the setting of finitely generated groups of polynomial growth (in fact, the Calderón-Zygmund decomposition itself, as presented in Theorem 2 in Chapter 1 of [48], is valid in the setting of general metric spaces equipped with a doubling measure). The second part of the proof of the Hörmander-Mihlin theorem, as presented in Theorem 2.5 of [18], requires several straightforward modifications in order to pass to the discrete setting. We leave the simple details to the reader. For the sake of readers that are not familiar with these aspects of Fourier analysis, we will later present a complete reduction to a continuous problem whose proof appears in print, which yields slightly worse bounds on the distortion guarantee.

In order to apply Theorem 4.1 we consider the following two multipliers,

$$m_1(u,v) \coloneqq \frac{\left|e^{\frac{2\pi iu}{n}}-1\right|^2}{\left|e^{\frac{2\pi iu}{n}}-1\right|^2+\left|e^{\frac{2\pi iv}{n}}-1\right|^2}, \quad \text{and} \quad m_2(u,v) \coloneqq \frac{\left(e^{-\frac{2\pi iu}{n}}-1\right)\cdot\left(e^{\frac{2\pi iv}{n}}-1\right)}{\left|e^{\frac{2\pi iu}{n}}-1\right|^2+\left|e^{\frac{2\pi iv}{n}}-1\right|^2}, \tag{27}$$

where we set $m_1(0,0) = m_2(0,0) = 0$. A direct (albeit tedious!) computation shows that $m_1$ and $m_2$ satisfy the conditions of Theorem 4.1 with $B = O(1)$. Thus, the operators $T_{m_1}$ and $T_{m_2}$ are weak $(1,1)$ with constant $O(1)$. Since $m_1$ and $m_2$ are bounded functions, the operator norms $\|T_{m_1}\|_{L_2(\mathbb{Z}_n^2)\to L_2(\mathbb{Z}_n^2)}$ and $\|T_{m_2}\|_{L_2(\mathbb{Z}_n^2)\to L_2(\mathbb{Z}_n^2)}$ are $O(1)$. Since these operators are self adjoint, by the Marcinkiewicz interpolation theorem (see [56]) it follows that for $p \geq 2$, the operator norms $\|T_{m_1}\|_{L_p(\mathbb{Z}_n^2)\to L_p(\mathbb{Z}_n^2)}$ and $\|T_{m_2}\|_{L_p(\mathbb{Z}_n^2)\to L_p(\mathbb{Z}_n^2)}$ are $O(p)$. This is precisely (24) and (25).

The above argument is based on Theorem 4.1, which does not appear exactly as stated in the literature, but its proof is a straightforward adaptation of existing proofs (which is too simple to justify rewriting the lengthy argument here). However, making the necessary changes easily does require some familiarity with Calderón-Zygmund theory. We therefore present now another argument which gives a polylog$(n)$ bound on the distortion, but uses only statements which appear in the literature. This alternative approach appears to be quite versatile, and might be useful elsewhere.

The following lemma reduces the problem of proving inequalities such as (24) and (25) (with perhaps a different dependence on $p$) to a continuous inequality. The argument is based on the proof of a theorem of Marcinkiewicz from [56] (see Theorem 7.5 in chapter X there). In what follows we denote by $\mathbb{T}$ the Euclidean unit circle in the plane.

**Proposition 4.2** (Transferring multipliers from the torus to $\mathbb{Z}_n^2$)**.** *Fix an odd integer $n$. Let $\{\lambda(u,v)\}_{u,v=0}^{\infty}$ be complex numbers such that $\lambda(u,v) = 0$ for $\max\{u,v\} \geq n$. Consider the operators $M : L_p\left(\mathbb{T}^2\right) \to L_p\left(\mathbb{T}^2\right)$ and $M_n : L_p\left(\mathbb{Z}_n^2\right) \to L_p\left(\mathbb{Z}_n^2\right)$ given by*

$$M\left(\sum_{u,v=-\infty}^{\infty} \widehat{f}(u,v)e^{2\pi i(ux+vy)}\right) = \sum_{u,v=0}^{\infty} \lambda(u,v)\widehat{f}(u,v)e^{2\pi i(ux+vy)},$$

*and*

$$M_n\left(\sum_{u,v=0}^{n-1} \widehat{f}(u,v)e^{\frac{2\pi i}{n}(ua+vb)}\right) = \sum_{u,v=0}^{n-1} \lambda(u,v)\widehat{f}(u,v)e^{\frac{2\pi i}{n}(ua+vb)}.$$

*Then,*

$$\|M_n\|_{L_p(\mathbb{Z}_n^2)\to L_p(\mathbb{Z}_n^2)} \leq 81 \cdot \|M\|_{L_p(\mathbb{T}^2)\to L_p(\mathbb{T}^2)}.$$

*Proof.* The proof is a variant of the first part of the proof of Theorem 7.5 in chapter X in [56], and a small twist on the second part. Since the terminology in [56] is different from ours, we repeat the proof of the first part as well. Recall that the Dirichlet kernels $D_\ell : [0,1] \to \mathbb{C}$ are defined as

$$D_\ell(x) = \sum_{j=-\ell}^{\ell} e^{2\pi ijx},$$

and the Fejér kernels $K_m : [0,1] \to \mathbb{C}$ are

$$K_m(x) = \frac{1}{m+1} \sum_{\ell=0}^{m} D_\ell x = \sum_{j=-m}^{m} \left(1 - \frac{|j|}{m+1}\right) e^{2\pi i j x}.$$

A basic property of $D_\ell$ is that for any trigonometric polynomial $S(x)$ of degree at most $\ell$, namely $S(x) = \sum_{j=-\ell}^{\ell} a_j e^{2\pi i j x}$, we have that $S(x) = S * D_\ell(x) = \int_0^1 S(t) D_\ell(x-t)dt$. The same is true with any other function all of whose $j$th Fourier coefficients for $j$ between $-\ell$ and $\ell$ are 1; in particular for the de la Vallée Poussin kernel $2K_{2\ell-1} - K_{\ell-1}$ (see [29]). The well known advantage of the Fejér kernel over the Dirichlet kernel is that it is everywhere (real and) nonnegative. Note also that $\int_0^1 K_m(t)dt = 1$ for all $m$. Thus, by convexity of the function $t^p$, for any trigonometric polynomial $S$ of degree at most $\ell$, and for all $x \in [0,1]$,

$$\begin{aligned} |S(x)|^p &= |2S * K_{2\ell-1}(x) - S * K_\ell(x)|^p \\ &\le 3^p \left( \frac{2}{3} \int_0^1 |S(t)|^p K_{2\ell-1}(x-t)dt + \frac{1}{3} \int_0^1 |S(t)|^p K_{\ell-1}(x-t)dt \right). \end{aligned} \tag{28}$$

Let now $\omega_{2\ell+1}$ be the measure which assign mass $\frac{1}{2\ell+1}$ to each of $2\ell+1$ equally spaced points on $[0,1]$. Then it is easy to check that

$$\int_0^1 K_m(x-t)d\omega_{2\ell+1}(x) = \int_0^1 K_m(x-t)dx = 1$$

for all $m \le 2\ell$ and for all $t \in [0,1]$. Integrating (28) with respect to $\omega_{2\ell+1}$, we get that for any trigonometric polynomial $S$ of degree at most $\ell$

$$\int_0^1 |S(x)|^p d\omega_{2\ell+1}(x) \le 3^p \int_0^1 |S(x)|^p dx. \tag{29}$$

It follows that if $S(x,y)$ is a two-variable trigonometric polynomial of degree at most $\ell$ in each of the variables, i.e. $S(x,y) = \sum_{u,v=-\ell}^{\ell} a_{uv} e^{2\pi i(ux+vy)}$,

$$\int_{[0,1]^2} |S(x,y)|^p d\omega_{2\ell+1}(x) d\omega_{2\ell+1}(y) \le 9^p \int_{[0,1]^2} |S(x,y)|^p dxdy.$$

It follows from this that, since $n$ is odd, for every $f \in L_p\left(\mathbb{T}^2\right)$,

$$\left\| M_n \left( \sum_{u,v=0}^{n-1} \widehat{f}(u,v) e^{\frac{2\pi i}{n}(ua+vb)} \right) \right\|_{L_p(\mathbb{Z}_n^2)} \le 9 \left\| M \left( \sum_{u,v=-\infty}^{\infty} \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right) \right\|_{L_p(\mathbb{T}^2)}.$$

Note that for each trigonometric polynomial of the form $P(x,y) = \sum_{u,v=-n+1}^{n-1} a_{uv} e^{2\pi i(ux+vy)}$,

$$\int_{[0,1]^2} P(x,y) d\omega_n(x) d\omega_n(y) = a_0 = \int_{[0,1]^2} P(x,y) dxdy.$$

Fix $f \in L_p\left(\mathbb{Z}_n^2\right)$, $1 < p < \infty$. By the first part of the proof and duality, there is $g \in L_{p^*}(\mathbb{T}^2)$ ($p^* = p/(p-1)$) with $\|g\|_{p^*} = 1$ such that

$$\begin{aligned}
\|M_n f\|_{L_p(\mathbb{Z}_n^2)} &\leq 9 \int_{[0,1]^2} \left( \sum_{u,v=0}^{n-1} \lambda_j \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right) \overline{g(x,y)} dxdy \\
&= 9 \int_{[0,1]^2} \left( \sum_{u,v=0}^{n-1} \lambda(u,v) \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right) \left( \sum_{u,v=0}^{n-1} \overline{\widehat{g}(u,v)} e^{-2\pi i(ux+vy)} \right) dxdy \\
&= 9 \int_{[0,1]^2} \left( \sum_{u,v=0}^{n-1} \lambda(u,v) \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right) \left( \sum_{u,v=0}^{n-1} \overline{\widehat{g}(u,v)} e^{-2\pi i(ux+vy)} \right) d\omega_n(x) d\omega_n(y) \\
&= 9 \int_{[0,1]^2} \left( \sum_{u,v=0}^{n-1} \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right) \left( \sum_{u,v=0}^{n-1} \lambda(u,v) \overline{\widehat{g}(u,v)} e^{-2\pi i(ux+vy)} \right) d\omega_n(x) d\omega_n(y) \\
&\leq 9 \left( \int_{[0,1]^2} \left| \sum_{u,v=0}^{n-1} \widehat{f}(u,v) e^{2\pi i(ux+vy)} \right|^p d\omega_n(x) d\omega_n(y) \right)^{1/p} \cdot \\
&\qquad \left( \int_{[0,1]^2} \left| \sum_{u,v=0}^{n-1} \lambda(u,v) \overline{\widehat{g}(u,v)} e^{-2\pi i(ux+vy)} \right|^{p^*} d\omega_n(x) \omega_n(y) \right)^{1/p^*} \\
&\leq 81 \cdot \|f\|_{L_p(\mathbb{Z}_n^2)} \left( \int_{[0,1]^2} \left| \sum_{u,v=0}^{n-1} \overline{\lambda(u,v)} \widehat{g}(u,v) e^{2\pi i(ux+vy)} \right|^{p^*} dx \right)^{1/p^*} \\
&\leq 81 \cdot \|f\|_{L_p(\mathbb{Z}_n^2)} \cdot \|M\|_{L_p(\mathbb{T}^2) \to L_p(\mathbb{T}^2)}.
\end{aligned}$$

where the inequality before last follows from (29) and the last inequality (that is the fact that the norm of a multiplier in $L_p\left(\mathbb{T}^2\right)$ is the same as the norm of the conjugate multiplier in $L_{p^*}\left(\mathbb{T}^2\right)$) follows from duality. The case $p = 1$ (and also a similar inequality for the $\infty$ norm) follows easily from the $L_p$ cases. □

Proposition 4.2 implies that it is enough to obtain $L_p$ to $L_p$ bounds for the operators $T_{m_1}$ and $T_{m_2}$, where $m_1, m_2$ are as in (27), as operators on functions on the torus $\mathbb{T}^2$. By a theorem of de Leeuw [11] it is enough to obtain such bounds when we think of $T_{m_1}$ and $T_{m_2}$ as operators on functions on $\mathbb{R}^2$ (see [55] for the respective result in the case of weak $(1,1)$ bounds). The continuous version of the Hörmander-Mihlin multiplier theorem now applies, but unfortunately its conditions are not satisfied. However, a (once again tedious) computation shows it is possible to apply the Marcinkiewicz multiplier theorem (see [47, 52]), in combination with bounds on the Hilbert transform [47, 52], to obtain bounds similar to (24) and (25) with $O(p)$ replaced by $O(\text{poly}(p))$ (it is quite easy to obtain a bound of $O(p^3)$, and with more work this can be reduced to $O(p^2)$. However we do not see a simple way to obtain $O(p)$ using this approach).

**Remark 4.1.** Consider the mapping $S : \mathscr{P}_{\mathbb{Z}_n^2} \to L_1\left(\mathbb{Z}_n^2\right)$ given by

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

Using considerations similar to the above (see Proposition III.A.3 in [54] for a continuous counterpart) it is possible to show that $S$ has distortion $O(\text{polylog}(n))$. However, we were unable to get this bound down to

$O(\log n)$ as in Theorem 1.4. Nevertheless, this embedding might be of interest since it reduces the dimension of the ambient $L_1$ space by a factor of 2.

# 5 Discussion and open problems

There are several interesting problems that arise from the results presented in this paper- we shall discuss some of them in the list below.

1. The most natural problem is to determine the asymptotic behavior of $c_1\left(\{0,1\ldots,n\}^2,\tau\right)$. It seems hard to use the ideas in Section 4 to obtain an embedding of distortion $O\left(\sqrt{\log n}\right)$, as the known bounds on multipliers usually give a weak $(1,1)$ inequality at best. We do not know what is the actual distortion of the embedding in Theorem 1.4.

2. Remark 4.1 implies that the Banach-Mazur distance between the $n^2-1$ dimensional normed space $\mathscr{M}_{\mathbb{Z}_n^2,\tau}$ and $\ell_1^{n^2-1}$ is $O(\text{polylog}(n))$. It would be interesting to determine the asymptotic behavior of this distance. In particular, it isn't clear whether the $L_1$ (embedding) distortion of $\mathscr{M}_{\mathbb{Z}_n^2,\tau}$ behaves differently from its Banach-Mazur distance from $\ell_1^{n^2-1}$.

3. We did not attempt to study the $L_1$ distortion of $\mathscr{M}_{\{0,1,\ldots,n\}^d,\tau}$ for $d \geq 3$. Observe that this space contains $\mathscr{M}_{\{0,1,\ldots,n\}^2,\tau}$, so the $\Omega\left(\sqrt{\log n}\right)$ lower bound still applies. But the result of [30] shows that the transportation cost metric on the Hamming cube $\{0,1\}^d$ has distortion $\Theta(d)$, so some improvements are still possible. Note that in higher dimensions it becomes interesting to study the transportation cost distance when $\mathbb{R}^d$ is equipped with other norms. The Banach-Mazur distance between $\ell_1^d$ and arbitrary $d$-dimensional norms has been studied in [7, 49, 13]. In particular, the result of [13] states that any $d$-dimensional Banach space is at distance $O\left(d^{5/6}\right)$ from $\ell_1^d$. Combining this fact with the lower bound on the $L_1$ distortion of the transportation cost distance on the Hamming ($\ell_1$) cube cited above, we see that for any norm $\|\cdot\|$ on $\mathbb{R}^d$, $c_1\left(\mathscr{P}_{(\mathbb{R}^d,\|\cdot\|),\tau}\right) = \Omega\left(d^{1/6}\right)$. It would be interesting to study the dependence on $d$ for general norms on $\mathbb{R}^d$.

4. As stated in Remark 3.2, it would be interesting to study the rate with which $c_2\left(\mathscr{P}_{\{0,\ldots,n\}^2},\ \sqrt{\tau}\right)$ tends to infinity.

5. As stated in Remark 3.3, we do not know whether $\left(\mathscr{P}_{[0,1]^2},\tau\right)$ admits a uniform embedding into Hilbert space.

6. The present paper rules out the "low distortion approach" to nearest neighbor search in the Earthmover metric via embeddings into $L_1$. However, it might still be possible to find *nearest neighbor preserving embeddings* into $L_1$ in the sense of [25].

7. On the more "applied side", as stated in the introduction, there is a possibility that the embedding of Theorem 1.4 behaves better than the theoretical distortion guarantee of $O(\log n)$ in "real life" situations, since it is often the case that the bulk of the Fourier spectrum is concentrated on a sparse set of frequencies. Additionally, it might be worthwhile to "thin out" some frequencies of the given set of images before embedding into $L_1$ (and then using the known $L_1$ nearest neighbor search databases). It would be interesting to carry out such "tweaking" of our algorithm in a more experimental setting.

**Acknowledgements.** We are grateful to David Jerison and Terence Tao for several helpful suggestions. The first named author wishes to thank Piotr Indyk for his encouragement to work on the planar Earthmover problem. We also thank Guillaume Aubrun for correcting several mistakes in our original manuscript. This work was carried out while the second named author was a long-term visitor at the Theory Group of Microsoft Research.

# 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] J. Bourgain and S. J. Szarek. The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization. *Israel J. Math.*, 62(2):169–180, 1988.

[8] 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.

[9] 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.

[10] 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.

[11] K. de Leeuw. On $L_p$ multipliers. *Ann. of Math. (2)*, 81:364–379, 1965.

[12] J. García-Cuerva and J. L. Rubio de Francia. *Weighted norm inequalities and related topics*, volume 116 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1985. Notas de Matemática [Mathematical Notes], 104.

[13] A. Giannopoulos. A note on the Banach-Mazur distance to the cube. In *Lindenstrauss, J. (ed.) et al., Geometric aspects of functional analysis. Israel seminar (GAFA) 1992-94. Basel: Birkhauser. Oper. Theory, Adv. Appl. 77, 67-73* . 1995.

[14] 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.

[15] 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.

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

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

[18] L. Hörmander. Estimates for translation invariant operators in $L^p$ spaces. *Acta Math.*, 104:93–140, 1960.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. Manuscript. Available at `http://research.microsoft.com/research/theory/naor/homepage%20files/lowdim-journal.pdf`, 2005.

[26] 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.

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

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

[29] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.

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

[31] 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.

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

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

[34] 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.

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

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

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

[38] S. G. Mihlin. On the multipliers of Fourier integrals. *Dokl. Akad. Nauk SSSR (N.S.)*, 109:701–703, 1956.

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

[40] 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.

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

[42] 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.

[43] 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.

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

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

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

[47] E. Stein. *Singular integrals and differentiability properties of functions*. Princeton University Press. XIV. Princeton, N.J. , 1970.

[48] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.

[49] S. J. Szarek and M. Talagrand. An "isomorphic" version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 105–112. Springer, Berlin, 1989.

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

[51] 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.

[52] A. Torchinsky. *Real-variable methods in harmonic analysis*. Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1986 original [Dover, New York; MR0869816].

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

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

[55] K. Woźniakowski. A new proof of the restriction theorem for weak type $(1,1)$ multipliers on $\mathbf{R}^n$. *Illinois J. Math.*, 40(3):479–483, 1996.

[56] A. Zygmund. *Trigonometric series. Volumes I and II combined. With a foreword by Robert Fefferman. 3rd ed.* Cambridge Mathematical Library. Cambridge: Cambridge University Press. xiii, 2002.