Nonembeddability theorems via Fourier analysis

Subhash Khot
Georgia Institute of Technology
khot@cc.gatech.edu

Assaf Naor
Microsoft Research
anaor@microsoft.com

Abstract

Various new nonembeddability results (mainly into L_1) are proved via Fourier analysis. In particular, it is shown that the Edit Distance on $\{0,1\}^d$ has L_1 distortion $\Omega(\sqrt{\log d/\log\log d})$. We also give new lower bounds on the L_1 distortion of flat tori, quotients of the discrete hypercube under group actions, and the transportation cost (Earthmover) metric.

1 Introduction

The bi-Lipschitz theory of metric spaces has witnessed a surge of activity in the past four decades. While the original motivation for this type of investigation came from metric geometry and Banach space theory, since the mid-1990s it has become increasingly clear that understanding metric spaces in the bi-Lipschitz category is intimately related to fundamental algorithmic questions arising in theoretical computer science. Despite the remarkable list of achievements of this field, which includes the best known approximation algorithms for a a wide range of NP hard problems, the bi-Lipschitz theory is still in its infancy. In particular, there are very few known methods for proving nonembeddability results. The purpose of this paper is to the develop a Fourier-analytic approach to proving nonembeddability theorems. In doing so, we resolve several problems, and shed new light on existing results. Additionally, our work points toward several interesting directions for future research, with emphasis on the study of the bi-Lipschitz structure of quotients of metric spaces.

Let (X,d_X) and (Y,d_Y) be metric spaces. The Lipschitz constant of a function $f:X\to Y$ is

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

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

$$\mathrm{dist}(f) := \|f\|_{\mathrm{Lip}}\cdot\|f^{-1}\|_{\mathrm{Lip}}.$$

If f is not one-to-one then we set $\mathrm{dist}(f)=\infty$. The least distortion with which X can be embedded into Y is denoted $c_Y(X)$, namely

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

We are particularly interested in embeddings into L_p spaces. In this case we write $c_p(X) = c_{L_p}(X)$. The most studied type of embeddings are into Hilbert space, in which case the parameter $c_2(X)$ is known as the Euclidean distortion of X. The parameter $c_1(X)$, i.e. the least distortion required to embed X into L_1, is of great algorithmic significance, especially in the study of cut problems in graphs. The Euclidean distortion of a metric space X is relatively well understood: it is enough to understand the distortion of finite subsets of X, and for finite metrics there is a simple semidefinite program which computes their Euclidean distortion [37]. Embeddings into L_1 are much more mysterious (see [36]), and there are very few known methods to bound $c_1(X)$ from below.

The present paper contains several new nonembeddability results, which we now describe. The common theme is that our proofs are based on analytic methods, most notably Fourier analysis on $\{0,1\}^d$ and $\mathbb{R}^n$. We stress that this is not the first time that nonembeddability results have drawn on techniques from harmonic analysis. Indeed, the proofs of results in [14, 54, 50, 32, 43] all have a Fourier analytic component.

Our results.

1) Quotients of the discrete hypercube and transportation cost. A classical theorem of Banach states that every separable Banach space is a quotient of ℓ_1. More precisely, for every separable Banach space X, there is a linear subspace $Y \subseteq \ell_1$ such that ℓ_1/Y is linearly isometric to X. This suggests that interesting "bad examples" of metric spaces can be obtained as *metric* quotients of the Hamming cube. Roughly speaking, this says that we can obtain interesting metrics (i.e. metrics that do not embed into nice spaces, say L_1) by identifying points of the hypercube. Quotients of metric spaces are a well studied concept (see [25, 15] for an introduction, and [42] for a discussion of quotients of finite metric spaces)- we refer the reader to Section 3 for a precise definition of this notion.

Motivated by this analogy, in Section 3 we exhibit classes of quotients of the Hamming cube which do not embed into L_1. A fundamental theorem of Bourgain [10] states that for every finite metric space X, $c_1(X) \le c_2(X) = O(\log |X|)$. In [10] Bourgain used a counting argument to show that there exist arbitrarily large metric spaces X with $c_2(X) = \Omega(\log |X|/\log\log |X|)$. In [37, 3] it was shown that there exit arbitrarily large metric spaces X with $c_1(X) = \Omega(\log |X|)$ (namely X can be taken to be a constant degree expander). In Section 3 we show that there exist simple n-point quotients of the Hamming cube $\{0,1\}^d$ which incur distortion $\Omega(\log n)$ in any L_1 embedding. This can be viewed a non-linear quantitative analog of Banach's theorem stated above. We also show that certain quotients of the Hamming cube obtained from the action of a transitive permutation group of the coordinates do not well-embed into L_1. These results are proved via a flexible Fourier analytic approach.

As an application of the results stated above we settle the problem of the L_1 embeddability of the transportation cost metric (also known as the Earthmover metric in the computer vision/graphics literature) on the set of all probability measures on $\{0,1\}^d$. Denoting by $\mathcal{P}(\{0,1\}^d)$ the space of all probability measures on the Hamming cube $\{0,1\}^d$, let $\mathcal{T}_\rho(\sigma,\tau)$ denote the transportation cost distance between $\sigma,\tau \in \mathcal{P}(\{0,1\}^d)$, with respect to the cost function induced by the Hamming metric ρ (see Section 3.2 for the definition). Such metrics occur in various contexts in computer science: they are a popular distance measure in graphics and vision [27, 30], and they are used as LP relaxations for classification problems such as 0-extension and metric labelling [20, 18, 2]. Transportation cost metrics are also prevalent in several areas of analysis and PDEs (see the book [58] and the references therein).

Motivated by applications to nearest neighbor search (a.k.a. similarity search in the vision literature), the problem of embedding transportation cost metrics into L_1 attracted a lot of attention in recent years (see [18, 30, 41]). In [18, 30] it is shown that $c_1(\mathcal{P}(\{0,1\}^d), \mathcal{T}_\rho) = O(d)$. In Section 3.2 we show that this bound is optimal, i.e. $c_1(\mathcal{P}(\{0,1\}^d), \mathcal{T}_\rho) = \Omega(d)$. From an analytic perspective, Kantorovich duality (see [58]) implies that $(\mathcal{P}(\{0,1\}^d), \mathcal{T}_\rho)$ embeds isometrically into $\mathrm{Lip}(\{0,1\}^d)^*$- the dual of the Banach space of all real valued Lipschitz functions on the hypercube. A result of Bourgain [11] implies that $\sup_{d\in\mathbb{N}} c_1(\mathrm{Lip}(\{0,1\}^d)^*) = \infty$. Our result shows that in fact $c_1(\mathrm{Lip}(\{0,1\}^d)^*) = \Omega(d)$, improving upon the lower bound obtained in [11].

2) Edit Distance does not embed into L_1. Edit Distance (also known as Levenstein distance [35]) is a metric defined on the set of all finite-length binary strings, which we denote $\{0,1\}^*$. This metric is best viewed as the shortest path metric on the following infinite graph : Let G be a graph with set of vertices $\{0,1\}^*$, and $\{x, y\}$ is an edge of the graph if the string y can be obtained from string x by either deleting one character from x or by inserting one character into x. For strings x, y, denote the shortest path distance in G (i.e. the Edit Distance) between x, y as $\mathrm{ED}(x, y)$. In words, $\mathrm{ED}(x, y)$ is the minimum number of edit operations needed to transform x into y. Here we assume that only insertion/deletion operations are allowed. Character substitution can be simulated by a deletion followed by an insertion. Similarly, one can *shift* a string by deleting its first character and inserting it at the end.

Edit Distance is a very useful metric arising in several applications, most notably in string and text comparison problems, which are prevalent in computer science (e.g. compression and pattern matching), computational biology, and web searching (see the papers [47, 21, 1, 29, 5, 52, 17] and the references therein, and the book [28] for a discussion of applications to computational biology).

Let $(\{0,1\}^d, \mathrm{ED})$ denote the space $\{0,1\}^d$ with the Edit Distance metric (inherited from the metric ED on $\{0,1\}^*$). A well known problem, stated e.g. in [41], is whether the space $(\{0,1\}^d, \mathrm{ED})$ embeds into L_1 with uniformly bounded distortion. Had this been true, it would have had significant applications in computer science (see [41]). Most notably it would lead to approximate nearest neighbor search algorithms under Edit Distance, and to efficient algorithms for computing the Edit Distance between two strings (both of these problems are being solved, by computational biologists, every day, hundreds of times. Getting a substantially faster algorithm for any of them would be of great practical importance). In Section 4 we show that the L_1 embedding approach fails, by proving via Fourier analytic methods that

$$c_1(\{0,1\}^d, \mathrm{ED}) = \Omega\left(\sqrt{\frac{\log d}{\log\log d}}\right). \tag{1}$$

The previous best known lower bound is due to [1], where it is shown that $c_1(\{0,1\}^d, \mathrm{ED}) \geq 3/2$.

The best known upper bound on $c_1(\{0,1\}^d, \mathrm{ED})$ is due to [52], where it is proved that

$$c_1(\{0,1\}^d, \mathrm{ED}) \le 2^{O(\sqrt{\log d \log\log d})}.$$

3) Flat tori can be highly non-Euclidean. The Nash embedding theorem [51] states that any n-dimensional Riemannian manifold is isometric to a Riemannian sub-manifold of $\mathbb{R}^{2n}$. In the bi-Lipschitz category this is no longer the case- it is easy to construct Riemannian manifolds (indeed, even Riemannian surfaces) which do not embed bi-Lipschitzly even into infinite dimensional Hilbert space. However, all the known constructions were highly curved, and the possibility remained that any manifold with zero curvature embeds bi-Lipschitzly into L_2, with a uniform bound on the distortion. In Section 5 we show that this isn't the case: there is an n-dimensional flat torus, i.e. $\mathbb{R}^n/\Lambda$ for some lattice $\Lambda \subseteq \mathbb{R}^n$, equipped with the natural Riemannian metric (whose sectional curvature is identically 0), such that $c_1(\mathbb{R}^n/\Lambda) = \Omega(\sqrt{n})$. This result answers the question, posed by W. B. Johnson, whether a Lipschitz quotient (in the sense of [7]) of Hilbert space embeds bi-Lipschitzly into Hilbert space. In [7] it is shown that a Banach space which is a Lipschitz quotient of a Hilbert space is isomorphic to a Hilbert space. Johnson's question is whether the condition that the quotient is a Banach space is necessary. Since the natural quotient map $\pi : \mathbb{R}^n \to \mathbb{R}^n/\Lambda$ is a Lipschitz quotient (see Section 3), the above example shows that Lipschitz quotients of Hilbert space need not embed into Hilbert space (indeed, they may not embed even into L_1). Our approach is a variant of our study of quotient metrics in Section 3, and the proof is based on Fourier analysis over $\mathbb{R}^n$, instead of discrete Fourier analysis over $\{0,1\}^n$.

This paper is organized as follows. In Section 2 we present some background and preliminary results on Fourier analysis on the Hamming cube. In section 3 we investigate quotients of the hypercube under group actions. In Section 4 we prove our lower bound on the L_1 distortion of Edit Distance, and in Section 5 we discuss the L_1 and L_2 embeddability of flat tori. We end with Section 6, which contains a brief discussion which relates the notion of *length of metric spaces* (first introduced by Schechtman [56] in the context of the concentration of measure phenomenon) to nonembeddability results. This gives, in particular, new lower bounds on the Euclidean distortion of various groups equipped with a group invariant metric.

2 Preliminaries on Fourier analysis on the hypercube

We start by introducing some notation concerning Fourier analysis on the group $\mathbb{F}_2^d = \{0,1\}^d$. For $\varepsilon \in (0,1)$ we denote by μ_ε the product ε-biased measure on $\mathbb{F}_2^d$, i.e. the measure given by

$$\forall\ x \in \mathbb{F}_2^d, \quad \mu_\varepsilon(\{x\}) = \varepsilon^{\sum_{j=1}^d x_j}(1-\varepsilon)^{d-\sum_{j=1}^d x_j}.$$

For the sake of simplicity we write $\mu = \mu_{1/2}$. Given $A \subseteq \{1,\ldots,d\}$ we define the Walsh function $W_A : \mathbb{F}_2^d \to \mathbb{R}$ by

$$W_A(x) = (-1)^{\sum_{j\in A} x_j}.$$

Then $\{W_A : A \subseteq \{1, \ldots, d\}\}$ is an orthonormal basis of $L_2(\mathbb{F}_2^d, \mu)$. In particular any $f : \mathbb{F}_2^d \to L_2$ has a unique Fourier expansion

$$f = \sum_{A \subseteq \{1,\ldots,d\}} \widehat{f}(A) W_A,$$

where

$$\widehat{f}(A) = \int_{\mathbb{F}_2^d} f(x) W_A(x) d\mu(x) \in L_2,$$

and Parseval's identity reads as

$$\int_{\mathbb{F}_2^d} \|f(x)\|_2^2 d\mu(x) = \sum_{A \subseteq \{1,\ldots,d\}} \|\widehat{f}(A)\|_2^2.$$

Let $e_j \in \mathbb{F}_2^d$ be the vector whose only non-zero coordinate is the jth coordinate. We also write $e := e_1 + \ldots + e_d$ for the all 1s vector. The partial differentiation operator on $L_2(\mathbb{F}_2^d)$ is defined by

$$\partial_j f(x) := \frac{f(x + e_j) - f(x)}{2}.$$

Since for every $A \subseteq \{1, \ldots, d\}$ we have that

$$\partial_j W_A = \begin{cases} -W_A & j \in A \\ 0 & j \notin A, \end{cases}$$

we see that for every $f : \mathbb{F}_2^d \to \mathbb{R}$

$$\sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \partial_j f(x)^2 d\mu(x) = \sum_{A \subseteq \{1,\ldots,d\}} |A| \widehat{f}(A)^2. \tag{2}$$

In what follows we denote by ρ the Hamming metric on $\mathbb{F}_2^d$, namely for $x, y \in \mathbb{F}_2^d$,

$$\rho(x, y) := |\{j \in \{1, \ldots, d\} : x_j \neq y_j\}|.$$

Observe that for every $f : \mathbb{F}_2^d \to \mathbb{R}$,

$$\int_{\mathbb{F}_2^d} |f(x) - f(x+e)|^2 d\mu(x) = \sum_{\substack{A \subseteq \{1,\ldots,d\} \\ |A| \equiv 1 \mod 2}} 4\widehat{f}(A)^2 \le 4 \sum_{A \subseteq \{1,\ldots,d\}} |A| \widehat{f}(A)^2 = 4 \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} [\partial_j f(x)]^2 d\mu(x).$$

This famous inequality, first proved by Enflo in [23] via a geometric argument, implies that $c_2(\mathbb{F}_2^d) \ge \sqrt{d}$. Indeed, by integration we see that for every $f : \mathbb{F}_2^d \to L_2$,

$$\int_{\mathbb{F}_2^d} \|f(x) - f(x+e)\|_2^2 d\mu(x) \le 4 \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \|\partial_j f(x)\|_2^2 d\mu(x).$$

Thus, assuming that f is invertible we see that

$$\frac{d^2}{\|f^{-1}\|_{\mathrm{Lip}}^2} \le 4d \cdot \left(\frac{\|f\|_{\mathrm{Lip}}}{2} \right)^2,$$

i.e.

$$\|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}} \geq \sqrt{d}.$$

This Fourier-analytic approach to Enflo's theorem motivates the ensuing arguments in this paper, since it turns out to be remarkably flexible. For future reference we record here the basic Poincaré inequality implied by the above reasoning:

Lemma 2.1. *For every* $f : \mathbb{F}_2^d \to L_2$,

$$\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \|f(x) - f(y)\|_2^2 d\mu(x) d\mu(y) \leq \frac{2}{\min\{|A| :\ A \neq \emptyset,\ \widehat{f}(A) \neq 0\}} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \|\partial_j f(x)\|_2^2 d\mu(x).$$

Proof. We simply observe that

$$\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \|f(x) - f(y)\|_2^2 d\mu(x) d\mu(y) = 2 \int_{\mathbb{F}_2^d} \|f(x) - \widehat{f}(\emptyset)\|^2 d\mu(x) = 2 \sum_{\emptyset \neq A \subseteq \{1,\ldots,d\}} \|\widehat{f}(A)\|_2^2,$$

and the required inequality follows from (2). □

3 Quotients of the hypercube

Let (X, d_X) be a metric space. For $A, B \subseteq X$ the Hausdorff distance between A, B is defined as

$$\mathcal{H}_X(A, B) = \sup\left\{\max\{d_X(a, B), d_X(b, A)\} :\ a \in A,\ b \in B\right\}. \tag{3}$$

Following [25, 15, 42], given a partition $\mathcal{U} = \{U_1, \ldots, U_k\}$ of X, we define the *quotient metric* induced by X on $\mathcal{U}$, denoted $X/\mathcal{U}$, as follows: assign to each $i, j \in \{1, \ldots, k\}$ the weight $w_{ij} = d_X(U_i, U_j) = \min_{x \in U_i,\ y \in U_j} d_X(x, y)$, and let $d_{X/\mathcal{U}}(U_i, U_j)$ be the shortest path distance between i and j in the weighted complete graph on $\{1, \ldots, k\}$ in which the edge $\{i, j\}$ has weight w_{ij}.

In the following lemma the right-hand inequality is an immediate consequence of (3), and the left-hand inequality follows from the fact that the Hausdorff distance is a metric on subsets of X.

Lemma 3.1. *Assume that* $\mathcal{U} = \{U_1, \ldots, U_k\}$ *is a partition of a metric space* X *such that for every* $i, j \in \{1, \ldots, k\}$, *for every* $x \in U_i$ *there exists* $y \in U_j$ *such that* $d_X(x, y) = d_X(U_i, U_j)$. *Then for every* $i, j \in \{1, \ldots, k\}$,

$$d_{X/\mathcal{U}}(U_i, U_j) = \mathcal{H}_X(U_i, U_j) = d_X(U_i, U_j).$$

A particular case of interest is when a group G acts on X by isometries. In this case the orbit partition induced by G on X clearly satisfies the conditions of Lemma 3.1, implying that for all $x, y \in X$,

$$d_{X/G}(Gx, Gy) = d_X(Gx, Gy),$$

where we slightly abuse notation by letting X/G be the quotient of X induced by the orbits of G. This is the only type of quotients that we study in this paper. In particular, Lemma 3.1 implies that the quotients we study here are also *Lipschitz quotients* in the sense of [7] (see Section 6 in [42] for an explanation).

We will require the following lower bound on the average distance in quotients of the hypercube.

Lemma 3.2. *Let G be a group of isometries acting on $\mathbb{F}_2^d$ with $2 < |G| < 2^d$. Then*

$$\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d} \rho_{\mathbb{F}_2^d/G}(Gx,Gy)d\mu(x)d\mu(y) = \Omega\left(\frac{\log_2|G|}{\log_2\left(\frac{d}{d-\log_2|G|}\right)}\right).$$

Proof. For every $t>0$,

$$\mu\times\mu\{x,y\in\mathbb{F}_2^d:\ \rho(Gx,Gy)\ge t\} \ge 1-\sum_{g\in G}\mu\times\mu\{x,y\in\mathbb{F}_2^d:\ \rho(x,gy)\le t\} = 1-\frac{|G|}{2^d}\cdot\sum_{k\le t}\binom{d}{k}.$$

We shall use the following (rough) bounds, which are a simple consequence of Stirling's formula: For every $1/n\le\delta\le 1/2$,

$$\frac{[\delta^\delta(1-\delta)^{1-\delta}]^{-d}}{6\sqrt{\delta d}} \le \sum_{k\le\delta d}\binom{d}{k} \le 2\sqrt{\delta d}\cdot[\delta^\delta(1-\delta)^{1-\delta}]^{-d}. \tag{4}$$

Thus, using Lemma 3.1 we get that

$$\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}\left[\rho_{\mathbb{F}_2^d/G}(Gx,Gy)\right]^2 d\mu(x)d\mu(y) \ge \delta d\left(1-\frac{|G|}{2^d}2\sqrt{\delta d}\cdot[\delta^\delta(1-\delta)^{1-\delta}]^{-d}\right)$$

Choosing $\delta=\Theta\left(\frac{\log_2|G|}{d\log_2\left(\frac{d}{d-\log_2|G|}\right)}\right)$ yields the required result. $\square$

3.1 A simple construction of n-point spaces with $c_1=\Omega(\log n)$

In what follows we refer to [38, 9] for the necessary background on coding theory. Let $C\subseteq\{0,1\}^d$ be a code, i.e. a linear subspace of $\mathbb{F}_2^d$. Denote by $w(C)$ the minimum Hamming weight of nonzero elements of C, i.e.

$$w(C)=\min_{x\in C\setminus\{0\}}\|x\|_1.$$

We also use the standard notation

$$C^\perp := \left\{x\in\mathbb{F}_2^d:\ \forall\ y\in C,\ \langle x,y\rangle\equiv 0 \mod 2\right\},$$

where $\langle x,y\rangle := \sum_{j=1}^n x_jy_j$.

Lemma 3.3. *Assume that $f:\mathbb{F}_2^d\to L_2$ satisfies for every $x\in\mathbb{F}_2^d$ and $y\in C^\perp$, $f(x+y)=f(x)$. Then for every nonempty $A\subseteq\{1,\dots,d\}$ with $|A|<w(C)$, $\widehat{f}(A)=0$.*

Proof. Since $(C^\perp)^\perp=C$ (see [9]), $\mathbf{1}_A\notin(C^\perp)^\perp$, implying that there exists $v\in C^\perp$ such that

$\langle \mathbf{1}_A, v\rangle \equiv 1 \mod 2$. Now,

$$\begin{aligned}
\widehat{f}(A) &= \int_{\mathbb{F}_2^n} f(x)W_A(x)d\mu(x) \\
&= \int_{\mathbb{F}_2^n} f(x+v)W_A(x)d\mu(x) \\
&= \int_{\mathbb{F}_2^n} f(x)W_A(x-v)d\mu(x) \\
&= (-1)^{\langle \mathbf{1}_A, v\rangle} \int_{\mathbb{F}_2^n} f(x)W_A(x)d\mu(x) \\
&= -\widehat{f}(A).
\end{aligned}$$

So $\widehat{f}(A) = 0$. $\square$

Theorem 3.4. *Let $C \subseteq \mathbb{F}_2^d$ be a code. Then*

$$c_1(\mathbb{F}_2^d/C^\perp) = \Omega\left(w(C) \cdot \frac{1 - \frac{\dim(C)}{d}}{\log\left(\frac{d}{\dim(C)}\right)} \right).$$

Proof. Let $f : \mathbb{F}_2^d/C^\perp \to L_1$ be a bijection. Define $\widetilde{f} : \mathbb{F}_2^d \to L_1$ by $\widetilde{f}(x) = f(x + C^\perp)$. It is well known [22] that there exists a mapping $T : L_1 \to L_2$ such that for all $x, y \in L_1$,

$$\|T(x) - T(y)\|_2 = \sqrt{\|x - y\|_1}.$$

Define $h : \mathbb{F}_2^d \to L_2$ by $h = T \circ \widetilde{f}$. By Lemma 3.3 and Lemma 2.1 we get that

$$\begin{aligned}
\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \|\widetilde{f}(x) - \widetilde{f}(y)\|_1 d\mu(x)d\mu(y) &= \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \|h(x) - h(y)\|_2^2 d\mu(x)d\mu(y) \\
&\le \frac{2}{w(C)} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \|\partial_j h(x)\|_2^2 d\mu(x) \\
&= \frac{2}{w(C)} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \|\partial_j \widetilde{f}(x)\|_1 d\mu(x) \\
&\le \frac{d}{w(C)} \|f\|_{\mathrm{Lip}}. \qquad (5)
\end{aligned}$$

On the other hand, by Lemma 3.2 we see that

$$\begin{aligned}
\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}\|\widetilde{f}(x)-\widetilde{f}(y)\|_1 d\mu(x)d\mu(y) &= \int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}\|f(x+C^\perp)-f(y+C^\perp)\|_1 d\mu(x)d\mu(y)\\
&\geq \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}\rho_{\mathbb{F}_2^d/C^\perp}(x+C^\perp,y+C^\perp)d\mu(x)d\mu(y)\\
&= \Omega\left(\frac{\log_2|C^\perp|}{\log_2\left(\frac{d}{d-\log_2|C^\perp|}\right)}\right)\cdot\frac{1}{\|f^{-1}\|_{\mathrm{Lip}}}\\
&= \Omega\left(\frac{d-\dim(C)}{\log\left(\frac{d}{\dim(C)}\right)}\right)\cdot\frac{1}{\|f^{-1}\|_{\mathrm{Lip}}}, \qquad (6)
\end{aligned}$$

where we used the fact that $|C^\perp| = 2^{d-\dim(C)}$.

Combining (5) and (6) yields the required result. $\square$

Corollary 3.5. *There exists arbitrarily large finite metric spaces X for which $c_1(X) = \Omega(\log|X|)$.*

Proof. Let $C \subseteq \{0,1\}^d$ be a code with $\dim(C) \geq \frac{d}{4}$ and $w(C) = \Omega(d)$. Such codes are well known to exist (see [38]), and are easy to obtain via the following greedy construction: fix $k \leq d/4$ and let V be a k dimensional subspace of $\mathbb{F}_2^d$ with $w(V) > \delta d$. Then V contains 2^k points. The number vectors $x \in \mathbb{F}_2^d$ with $\|x+v\|_1 \leq \delta d$ for some $v \in V$ is at most $2^k\sum_{\ell\leq\delta d}\binom{d}{\ell} \leq 2^{k+1}\sqrt{\delta d}\cdot[\delta^\delta(1-\delta)^{1-\delta}]^{-d}$. It follows that there exists $\delta = \Omega(1)$ such that for every $k \leq d/4$ there exists $x \in \mathbb{F}_2^d$ such that $w(\mathrm{span}(V\cup\{x\})) > \delta d$, as required. Now, for C as above, Theorem 3.4 implies that

$$c_1(\mathbb{F}_2^d/C^\perp) = \Omega(d) = \Omega(\log|\mathbb{F}_2^d/C^\perp|).$$

$\square$

Remark 3.1. *Using the Matoušek's extrapolation lemma for Poincaré inequalities [40] (see also Lemma 5.5 in [6]), it is possible to prove that for a code C as in Corollary 3.5, for every $p \geq 1$, $c_p(\mathbb{F}_2^d/C^\perp) \geq c(p)d$.*

3.2 The relation to transportation cost

Given a finite metric space (X,d) we denote by $\mathcal{P}(X)$ the set of all probability measures on X. For $\sigma,\tau \in \mathcal{P}(X)$ we define

$$\Pi(\sigma,\tau) = \left\{\pi\in\mathcal{P}(X\times X):\ \forall\, x\in X,\ \int_X d\pi(x,y) = \sigma(x),\quad \text{and}\quad \int_X d\pi(y,x) = \tau(x)\right\}$$

The optimal transportation cost (with respect to the metric d) between σ and τ is defined as

$$\mathcal{T}_d(\sigma,\tau) = \inf_{\pi\in\Pi(\sigma,\tau)}\int_{X\times X} d(x,y)d\pi(x,y).$$

Given $A \subseteq X$ we denote by $\mu_A \in \mathcal{P}(X)$ the uniform probability measure on A. If $A, B \subseteq X$ have the same cardinality then a straightforward extreme point argument (see [58]) shows that

$$\mathcal{T}_d(\mu_A, \mu_B) = \inf\left\{\int_A d(a, f(a))d\mu_A : \ f : A \to B \ \text{ is } 1-1 \text{ and onto}\right\}.$$

Lemma 3.6. *Let G be a finite group, equipped with a group invariant metric d (i.e. $d(xg, yg) = d(x, y)$ for all $g, x, y \in G$). Then for every subgroup $H \subseteq G$ and $x, y \in G$,*

$$d_{G/H}(xH, yH) = \mathcal{T}_d(\mu_{xH}, \mu_{yH}).$$

Proof. For every bijection $f : xH \to yH$,

$$\int_{xH} d(g, f(g))d\mu_{xH}(g) \geq d(xH, yH) = d_{G/H}(xH, yH).$$

On the other hand, fix $h_1, h_2 \in H$ such that $d(xh_1, yh_2) = d(xH, yH)$. Then the mapping $f : xH \to yH$ given by $f(g) = yh_2h_1^{-1}x^{-1}g$ satisfies for all $g \in xH$, $d(g, f(g)) = d(xh_1, yh_2) = d(xH, yH)$, implying the required result. □

Corollary 3.7. *It follows from Corollary 3.5 and Lemma 3.6 that $c_1(\mathcal{P}(\mathbb{F}_2^d), \mathcal{T}_\rho) = \Omega(d)$. This matches the upper bound proved in [18, 30]. In fact, from Remark 3.1 we see that for all $p \geq 1$, $c_p(\mathcal{P}(\mathbb{F}_2^d), \mathcal{T}_\rho) \geq c(p)d$.*

Remark 3.2. *Let $\mathrm{Lip}(\mathbb{F}_2^d)$ be the Banach space of all functions $f : \mathbb{F}_2^d \to \mathbb{R}$ satisfying $f(0) = 0$, equipped with the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$. By Kantorovich duality (see [58]), $(\mathcal{P}(\mathbb{F}_2^d), \mathcal{T}_\rho)$ is isometric to a subset of the dual space $\mathrm{Lip}(\mathbb{F}_2^d)^*$. It follows that $c_1(\mathrm{Lip}(\mathbb{F}_2^d)^*) = \Omega(d)$. As remarked in the introduction, the fact that $\sup_{d\in\mathbb{N}} c_1(\mathrm{Lip}(\mathbb{F}_2^d)^*) = \infty$ was first proved by Bourgain [11] using a different argument (which yields a worse lower bound on the distortion).*

3.3 Actions of transitive permutation groups

Let $G \leq S_d$ be a subgroup of the symmetric group. Clearly G acts by isometries on $\mathbb{F}_2^d$ via permutations of the coordinates.

Theorem 3.8. *Let $f : \mathbb{F}_2^d \to L_1$ be a G-invariant function, where G is transitive. Then*

$$\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \|f(x) - f(y)\|_1 d\mu(x)d\mu(y) \leq \frac{20}{\log d} \sum_{j=1}^d \int_{\mathbb{F}_2^d} \|\partial_j f(x)\|_1 d\mu(x).$$

Proof. Let $A \subseteq \mathbb{F}_2^d$ be a G invariant subset of the hypercube and write $\mu(A) = p$. For $f = \mathbf{1}_A$ the required inequality becomes:

$$2p(1-p) \leq \frac{10}{\log d} \sum_{j=1}^d I_j(A), \tag{7}$$

where $I_j(A) = \mu\{x \in \mathbb{F}_2^d : \ |\{x, x+e_j\} \cap A| = 1\}|$ is the *influence* of the jth variable on A. By [31], $\max_{1\le j\le d} I_j(A) \ge \frac{\log d}{5d} \cdot p(1-p)$. But, since A is invariant under the action of a transitive permutation group, $I_j(A)$ is independent of j, so (7) does indeed hold true.

In the general case let $f : \mathbb{F}_2^d \to L_1$ be a G invariant function. Denote by $\pi : \mathbb{F}_2^d \to \mathbb{F}_2^d/G$ the natural quotient map, i.e. $\pi(x) = Gx$. Since f is G-invariant, there is a function $h : \mathbb{F}_2^d/G \to L_1$ such that $f = h \circ \pi$. By the cut-cone representation of L_1 metrics (see [22]), there are nonnegative weights $\{\lambda_A\}_{A\subseteq \mathbb{F}_2^d/G}$ such that for every $x, y \in \mathbb{F}_2^d$,

$$\begin{aligned}
\|f(x) - f(y)\|_1 &= \|h(\pi(x)) - h(\pi(y))\|_1 \\
&= \sum_{A\subseteq \mathbb{F}_2^d/G} \lambda_A |\mathbf{1}_A(\pi(x)) - \mathbf{1}_A(\pi(y))| \\
&= \sum_{A\subseteq \mathbb{F}_2^d/G} \lambda_A |\mathbf{1}_{\pi^{-1}(A)}(x) - \mathbf{1}_{\pi^{-1}(A)}(y)|.
\end{aligned}$$

Observe that for every $A \subseteq \mathbb{F}_2^d/G$, $\pi^{-1}(A) \subseteq \mathbb{F}_2^d$ is G-invariant. Thus by the above reasoning

$$\begin{aligned}
\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d} \|f(x) - f(y)\|_1 d\mu(x)d\mu(y) &= \sum_{A\subseteq \mathbb{F}_2^d/G} \lambda_A \int_{\mathbb{F}_2^d\times\mathbb{F}_2^d} |\mathbf{1}_{\pi^{-1}(A)}(x) - \mathbf{1}_{\pi^{-1}(A)}(y)| d\mu(x)d\mu(y) \\
&\le \sum_{A\subseteq \mathbb{F}_2^d/G} \lambda_A \cdot \frac{20}{\log d} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} |\partial_j \mathbf{1}_{\pi^{-1}(A)}(x)| d\mu(x) \\
&= \frac{20}{\log d} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \sum_{A\subseteq \mathbb{F}_2^d/G} \lambda_A \left| \frac{\mathbf{1}_{\pi^{-1}(A)}(x) - \mathbf{1}_{\pi^{-1}(A)}(x+e_j)}{2} \right| d\mu(x) \\
&= \frac{20}{\log d} \sum_{j=1}^{d} \int_{\mathbb{F}_2^d} \|\partial_j f(x)\|_1 d\mu(x).
\end{aligned}$$

□

We thus get many examples of spaces which do not well-embed into L_1:

Corollary 3.9. *Let G be a transitive permutation group with $|G| < 2^{\varepsilon d}$, for some $\varepsilon \in (0,1)$. Then*

$$c_1(\mathbb{F}_2^d/G) \ge \Omega\left(\frac{(1-\varepsilon)}{\log(1/\varepsilon)} \cdot \log d \right).$$

Proof. This is a direct consequence of Theorem 3.8 and lemma 3.2. □

Remark 3.3. *It is possible to obtain slightly stronger results analogous to Corollary 3.9 when we have additional information on the structure of the group G. Indeed, in this case, in the proof of Theorem 3.8, one can use the results of Bourgain and Kalai [13] on the influence of variables on group invariant Boolean functions, instead of using [31].*

Remark 3.4. *We do not know if $(\mathbb{F}_2^d, \|\cdot\|_2)/G$ embeds bi-Lipschitzly in Hilbert space with uniformly bounded distortion. This seems to be unknown even in the case when G is generated by the cyclic shift of the coordinates. This problem is interesting since if this space does embed into Hilbert space, then the results of this section will yield an alternative approach to the recent solution of the Goemans-Linial conjecture in [32].*

4 Edit Distance does not embed into L_1

In this section we settle the L_1 embeddability problem of Edit Distance negatively, by proving the following theorem:

Theorem 4.1. *The following lower bound holds true:*

$$c_1(\mathbb{F}_2^d, \mathrm{ED}) = \Omega\left(\sqrt{\frac{\log d}{\log\log d}}\right).$$

The following lemma is a useful way to prove L_1 nonembeddability results. The case $\delta = 0$ of this lemma is due to [37]. Variants of the case $\delta > 0$, which is the case used in our proof of Theorem 4.1, seem to be folklore. We include here the formulation we need for the sake of completeness (the main part of the proof below is a variant of the proof of Lemma 3.6 in [49]).

Lemma 4.2. *Fix $\alpha > 0$ and $0 < \delta < \frac{1}{3}$. Let (X, d) be a finite metric space, σ a probability measure on X, and τ a probability measure on $X \times X$. Assume that for every $A \subseteq X$ with $\delta \le \sigma(A) \le \frac{2}{3}$ we have that $\tau(\{(x, y) \in X \times X : \ |\{x, y\} \cap A| = 1\}) \ge \alpha\sigma(A)$. Then,*

$$c_1(X) \ge \frac{\alpha}{2} \cdot \frac{\int_{X\times X} d(x, y)\, d\sigma(x)d\sigma(y) - 2\delta\,\mathrm{diam}(X)}{\int_{X\times X} d(x, y)\, d\tau(x, y)}.$$

Proof. We claim that there exists a subset $Y \subseteq X$ with $\sigma(Y) \ge 1-\delta$ such that for every $f : Y \to L_1$,

$$\int_{Y\times Y} \|f(x) - f(y)\|_1 d\sigma(x)d\sigma(y) \le \frac{2}{\alpha} \int_{Y\times Y} \|f(x) - f(y)\|_1 d\tau(x, y). \tag{8}$$

This will imply the required lower bound on $c_1(X)$ since if $f : X \to L_1$ is a bijection then

$$\int_{Y\times Y} \|f(x) - f(y)\|_1 d\tau(x, y) \le \|f\|_{\mathrm{Lip}} \int_{X\times X} d(x, y)\, d\tau(x, y),$$

while

$$\begin{aligned}\int_{Y\times Y} \|f(x) - f(y)\|_1 d\sigma(x)d\sigma(y) \ &\ge\ \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}}\left(\int_{X\times X} d(x, y)\, d\sigma(x)d\sigma(y) - \right.\\ &\qquad \left. 2\int_{X\times(X\setminus Y)} d(x, y)\, d\sigma(x)d\sigma(y)\right)\\ &\ge\ \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}}\left(\int_{X\times X} d(x, y)\, d\sigma(x)d\sigma(y) - 2\delta\,\mathrm{diam}(X)\right).\end{aligned}$$

It remains to prove the existence of the required subset Y. For simplicity we denote for every $A, B \subseteq X$,

$$\{A, B\} = \Big\{(x, y) \in X \times X : \ \{x, y\} \cap A \neq \emptyset \wedge \{x, y\} \cap B \neq \emptyset\Big\}.$$

Define inductively disjoint subsets $\emptyset = W_0, W_1, \ldots, W_k \subseteq X$ as follows. Having defined $W_1, \ldots, W_i$, write $Y_i = \cup_{\ell=1}^{i} W_\ell$ and let $W_{i+1} \subseteq X \setminus Y_i$ be an arbitrary nonempty subset for which

$$\tau(\{W_{i+1}, X \setminus (Y_i \cup W_{i+1})\}) < \alpha\sigma(W_{i+1}) \leq \frac{\alpha}{2}\sigma(X \setminus Y_i).$$

If no such W_j exists then this process terminates. We claim that $\sigma(Y_k) < \delta$. Indeed, otherwise let j be the first time at which $\sigma(Y_j) \geq \delta$. Observe that

$$\sigma(Y_j) = \sigma(Y_{j-1}) + \sigma(W_j) < \sigma(Y_{j-1}) + \frac{1}{2}\sigma(X \setminus Y_{j-1}) \leq \frac{1+\delta}{2} \leq \frac{2}{3}.$$

By our assumptions it follows that $\tau(\{Y_j, X \setminus Y_j\}) \geq \alpha\sigma(Y_j)$. But from the following simple inclusion

$$\{Y_j, X \setminus Y_j\} = \left\{\bigcup_{i=1}^{j} W_i, X \setminus \bigcup_{i=1}^{j} W_i\right\} \subseteq \bigcup_{i=1}^{j}\{W_i, X \setminus (Y_{i-1} \cup W_i)\}$$

we deduce that

$$0 < \alpha\sigma(Y_j) \leq \tau(\{Y_j, X \setminus Y_j\}) \leq \sum_{i=1}^{j} \tau(\{W_i, X \setminus (Y_{i-1} \cup W_i)\}) < \sum_{i=1}^{j} \alpha\sigma(W_i) = \alpha\sigma(Y_j),$$

a contradiction. Thus, taking $Y = Y_k$ we see that for every $A \subseteq Y$ with $\sigma(A) \leq \frac{1}{2}$ we have $\tau(\{A, Y \setminus A\}) \geq \alpha\sigma(A)$. In other words,

$$\begin{aligned}
\int_{Y \times Y} \|\mathbf{1}_A(x) - \mathbf{1}_A(y)\|_1 d\tau(x, y) &= \tau(\{A, Y \setminus A\}) \\
&\geq \alpha\sigma(A) \\
&\geq \alpha\sigma(A)[\sigma(Y) - \sigma(A)] \\
&= \frac{\alpha}{2}\int_{Y \times Y} \|f(x) - f(y)\|_1 d\sigma(x) d\sigma(y),
\end{aligned}$$

which implies (8) by the cut cone representation of L_1 metrics (as in the proof of Theorem 3.8). $\square$

In what follows we let S denote the cyclic shift operator on $\mathbb{F}_2^d$, namely

$$S(x_1, \ldots, x_d) = (x_d, x_1, x_2, \ldots, x_{d-1}).$$

Lemma 4.3. *There exists a universal constant $C > 0$ such that for every $\varepsilon \in (0, 1/2)$, every integer $k \geq (1/\varepsilon)^{C/\varepsilon}$, and every $f : \mathbb{F}_2^d \to \{-1, 1\}$,*

$$\frac{1}{2}\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |f(x) - f(y)| d\mu(x) d\mu(y) - 3\varepsilon \leq \frac{8}{k\sqrt{\varepsilon}} \sum_{j=1}^{k} \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |f(x) - f(S^j(x) + y)| d\mu(x) d\mu_\varepsilon(y). \quad (9)$$

Proof. Observe that for every $x, y \in \mathbb{F}_2^d$, $|f(x) - f(y)| = 1 - f(x)f(y)$. Thus for every $j \in \{1, \ldots, d\}$

$$
\begin{aligned}
&\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |f(x) - f(S^j(x) + y)| d\mu(x) d\mu_\varepsilon(y) = 1 - \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} f(x) f(S^j(x) + y) d\mu(x) d\mu_\varepsilon(y) \\
&= 1 - \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \left(\sum_{A,B \subseteq \{1,\ldots,d\}} \widehat{f}(A)\widehat{f}(B) W_A(x) W_B(S^j(x) + y) \right) d\mu(x) d\mu_\varepsilon(y) \\
&= 1 - \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \left(\sum_{A,B \subseteq \{1,\ldots,d\}} \widehat{f}(A)\widehat{f}(B) W_A(x) W_{S^{-j}(B)}(x) W_B(y) \right) d\mu(x) d\mu_\varepsilon(y) \\
&= 1 - \sum_{A \subseteq \{1,\ldots,d\}} (1 - 2\varepsilon)^{|A|} \widehat{f}(A) \widehat{f}(S^j(A)) \\
&\geq 1 - \sum_{A \subseteq \{1,\ldots,d\}} (1 - 2\varepsilon)^{|A|} \widehat{f}(A)^2, \qquad (10)
\end{aligned}
$$

where we used the Cauchy-Schwartz inequality and the facts that for all $B \subseteq \{1, \ldots, d\}$ we have $\int_{\mathbb{F}_2^d} W_B(y) d\mu_\varepsilon(y) = (1 - 2\varepsilon)^{|B|}$ and $\int_{\mathbb{F}_2^d} W_A W_{S^{-j}(B)} d\mu = 0$ when $B \neq S^j(A)$. Averaging (10) over $j = 1, \ldots, k$ we see that

$$
\frac{1}{k} \sum_{j=1}^{k} \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |f(x) - f(S^j(x) + y)| d\mu(x) d\mu_\varepsilon(y) \geq 1 - \sum_{A \subseteq \{1,\ldots,d\}} (1 - 2\varepsilon)^{|A|} \widehat{f}(A)^2.
$$

Thus, in order to prove (9) we may assume that

$$
\sum_{A \subseteq \{1,\ldots,d\}} (1 - 2\varepsilon)^{|A|} \widehat{f}(A)^2 \geq 1 - \frac{\sqrt{\varepsilon}}{4}. \qquad (11)
$$

By a recent theorem of Mossel, O'Donnell and Oleszkiewicz [46], together with a theorem of Friedgut [24] (weaker estimates can be obtained using Bourgain's noise sensitivity theorem [12]), inequality (11) implies that there exists a constant $c > 0$, an integer $t \leq (1/\varepsilon)^{c/\varepsilon}$, a function $g : \mathbb{F}_2^t \to \{-1, 1\}$, and indices $1 \leq i_1 < i_2 < \cdots < i_t \leq d$ such that if we extend g to a function $\widetilde{g} : \mathbb{F}_2^d \to \{-1, 1\}$ by setting

$$
\widetilde{g}(x_1, \ldots, x_d) = g(x_{i_1}, x_{i_2}, \ldots, x_{i_d}),
$$

then

$$
\int_{\mathbb{F}_2^d} |f(x) - \widetilde{g}(x)| d\mu(x) \leq \varepsilon.
$$

Write $I = \{i_1, \ldots, i_t\}$ and for $j \in \{1, \ldots, d\}$ define $I + j = \{i_1 + j \mod d, \ldots, i_d + j \mod d\}$. If $I \cap (I + j) = \emptyset$ then we have the identity

$$
\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |\widetilde{g}(x) - \widetilde{g}(S^j(x) + y)| d\mu(x) d\mu_\varepsilon(y) = \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} |\widetilde{g}(x) - \widetilde{g}(y)| d\mu(x) d\mu(y).
$$

Assume that $k \geq 2t^2$. In this case

$$
|\{j \in \{1, \ldots, k\} : \ I \cap (I + j) = \emptyset\}| \geq k - t^2 \geq \frac{k}{2}.
$$

Thus

$$\frac{1}{k}\sum_{j=1}^{k}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}|\widetilde{g}(x)-\widetilde{g}(S^j(x)+y)|d\mu(x)d\mu_\varepsilon(y)\geq\frac{1}{2}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}|\widetilde{g}(x)-\widetilde{g}(y)|d\mu(x)d\mu(y).$$

It follows that

$$\begin{aligned}\frac{1}{k}\sum_{j=1}^{k}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}|f(x)-f(S^j(x)+y)|d\mu(x)d\mu_\varepsilon(y) &\geq \frac{1}{2}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}|f(x)-f(y)|d\mu(x)d\mu(y)-\\ &\quad 3\int_{\mathbb{F}_2^d}|f(x)-\widetilde{g}(x)|d\mu(x)\\ &\geq \frac{1}{2}\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}|f(x)-f(y)|d\mu(x)d\mu(y)-3\varepsilon.\end{aligned}$$

This completes the proof of (9). □

We require the following rough bound on the average edit distance on $\mathbb{F}_2^d$.

Lemma 4.4. *We have the following lower bound on the average Edit Distance on $\mathbb{F}_2^d$:*

$$\int_{\mathbb{F}_2^d\times\mathbb{F}_2^d}\mathrm{ED}(x,y)d\mu(x)d\mu(y)\geq\frac{d}{160}.$$

Proof. For every $x\in\mathbb{F}_2^d$ and every integer $r<d/2$,

$$|\{y\in\mathbb{F}_2^d:\ \mathrm{ED}(x,y)=r\}|\leq 2^r\binom{2d}{r}.$$

This is best seen by observing that any sequence of r insertions or deletions can be executed in a sorted order, that is, the indices of positions on which the operation is performed increases. There are at most $\binom{2d}{r}$ ways to choose the r locations of these edit operations, and 2^r possible insertion/deletion operations on these r locations.

Now,

$$\begin{aligned}\mu\times\mu(\{(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d:\ \mathrm{ED}(x,y)>d/16\}) &\geq 1-\frac{1}{2^d}\sum_{r\leq d/16}2^r\binom{2d}{r}\\ &\geq 1-\frac{1}{2^d}\cdot 2^{d/8}\cdot 2\sqrt{d/8}[(1/16)^{1/16}(15/16)^{15/16}]^{-2d}\\ &\geq \frac{1}{10}.\end{aligned}$$

□

Proof of Theorem 4.1. Let C be the constant in Lemma 4.3. Fix $\varepsilon\in(0,1/2)$ such that $\varepsilon d>(1/\varepsilon)^{C/\varepsilon}-1$, and an integer $\varepsilon d\geq k\geq(1/\varepsilon)^{C/\varepsilon}$. Define a distribution τ on $\mathbb{F}_2^d\times\mathbb{F}_2^d$ as follows: pick a pair $(x,y)\in\mathbb{F}_2^d\times\mathbb{F}_2^d$ according to the measure $\mu\times\mu_\varepsilon$, pick $j\in\{1,\ldots,k\}$ uniformly at random, and consider the random pair $(x, S^j(x)+y)$. This induces a probability distribution τ on $\mathbb{F}_2^d \times \mathbb{F}_2^d$. Observe that

$$\mathrm{ED}(x, S^j(x)+y) \le 2\rho(0,y) + 2j \le 2\rho(0,y) + 2\varepsilon d.$$

Thus

$$\begin{aligned}
\int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \mathrm{ED}(x,y) d\tau(x,y) &= \frac{1}{k}\sum_{j=1}^{k} \int_{\mathbb{F}_2^d \times \mathbb{F}_2^d} \mathrm{ED}(x, S^j(x)+y) d\tau(x,y) \\
&\le 2\varepsilon d + 2\sum_{r=0}^{d} \binom{d}{r} r\varepsilon^r (1-\varepsilon)^{d-r} = 4\varepsilon d. \qquad (12)
\end{aligned}$$

Lemma 4.3 implies that for every $A \subseteq \mathbb{F}_2^d$,

$$\frac{16}{\sqrt{\varepsilon}} \cdot \tau(\{(x,y) \in \mathbb{F}_2^d \times \mathbb{F}_2^d : \ |\{x,y\} \cap A| = 1\}) \ge 2\mu(A)[1-\mu(A)] - 3\varepsilon.$$

Thus, the conditions of Lemma 4.2 hold true with $\delta = 6\varepsilon$ and $\alpha = \frac{\sqrt{\varepsilon}}{32}$. Hence by (12) and Lemma 4.4

$$c_1(\mathbb{F}_2^d, \mathrm{ED}) \ge \frac{\sqrt{\varepsilon}}{64} \cdot \frac{\frac{d}{80} - 6\varepsilon \cdot 2d}{4\varepsilon d}.$$

This implies the required result when we choose $10^{-3} > \varepsilon \approx \frac{\log\log d}{\log d}$. $\square$

5 Flat tori which do not embed into L_1

Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice in $\mathbb{R}^n$ of rank n. The quotient space $\mathbb{R}^n/\Lambda$ is a Riemannian manifold (n-dimensional torus) whose curvature is identically zero. Nevertheless, we show here that it is possible to construct lattices Λ such that $c_1(\mathbb{R}^n/\Lambda) = \Omega(\sqrt{n})$. For a lattice $\Lambda \subseteq \mathbb{R}^n$ we denote its fundamental parallelepiped by P_Λ. The dual lattice of Λ, denoted Λ^*, is defined by

$$\Lambda^* = \{x \in \mathbb{R}^n : \ \forall y \in \Lambda, \ \langle x, y\rangle \in \mathbb{Z}\}.$$

We shall use the following notation

$$N(\Lambda) = \min_{x \in \Lambda \setminus \{0\}} \|x\|_2 \quad \text{and} \quad r(\Lambda) = \max_{x \in \mathbb{R}^n} \min_{y \in \Lambda} \|x-y\|_2.$$

In words, $N(\Lambda)$ is the length of the shortest vector in Λ, and $r(\Lambda)$ is the smallest r such that balls of radius r centered at lattice points cover $\mathbb{R}^n$.

Theorem 5.1. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice. Then*

$$c_1(\mathbb{R}^n/\Lambda) = \Omega\left(\frac{N(\Lambda^*)}{r(\Lambda^*)} \cdot \sqrt{n}\right).$$

Corollary 5.2. *Let Λ be a lattice such that Λ^* is almost perfect, i.e. $N(\Lambda^*) = 1$ and $r(\Lambda^*) \le 4$, say. Such lattices are well known to exists (see [59, 39]). Then Theorem 5.1 implies that $c_1(\mathbb{R}^n/\Lambda) = \Omega(\sqrt{n})$. This is, in particular, an example of a Riemannian manifold whose curvature is identically zero which does not well-embed bi-Lipschitzly into ℓ_2. This fact should be contrasted with the Nash embedding theorem [51], which says that any n-dimensional Riemannian manifold is isometric to a Riemannian submanifold of $\mathbb{R}^{2n}$.*

Remark 5.1. *Some restrictions on the Lattice Λ should be imposed in order to obtain a torus $\mathbb{R}^n/\Lambda$ which does not embed into ℓ_2. Indeed, the mapping $f : \mathbb{R}^n/\mathbb{Z}^n \to \mathbb{C}^n$ defined by $f(x_1, \ldots, x_n) = (e^{2\pi i x_1}, \ldots, e^{2\pi i x_n})$ has distortion $O(1)$. We leave open the interesting problem of determining the value of $c_1(\mathbb{R}^n/\Lambda)$ and $c_2(\mathbb{R}^n/\Lambda)$ as a function of intrinsic geometric parameters of the lattice Λ. In Theorem 5.8 below we show that for every n*

$$L_n := \sup\{c_2(\mathbb{R}^n/\Lambda) : \ \Lambda \subseteq \mathbb{R}^n \text{ is a lattice}\} < \infty.$$

Corollary 5.2 shows that $L_n = \Omega(\sqrt{n})$, while the upper bound obtained in Theorem 5.8 is $L_n = O(n^{3n/2})$. It would be of great interest to close the large gap between these bounds.

The proof of Theorem 5.1 will be broken down into a few lemmas. In what follows we fix a lattice $\Lambda \subseteq \mathbb{R}^n$ and denote by m the *normalized* Riemannian volume measure on the torus $\mathbb{R}^n/\Lambda$. Given a function $f : \mathbb{R}^n/\Lambda \to L_1$ we also think of f as an Λ-invariant function defined on $\mathbb{R}^n$. We refer to [57] for the necessary background on Fourier analysis on tori used in the ensuing arguments.

Lemma 5.3. *Let γ denote the standard Gaussian measure on $\mathbb{R}^n$, i.e. $d\gamma(x) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$. Then for every continuous $f : \mathbb{R}^n/\Lambda \to L_1$,*

$$\begin{aligned}
&\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_1 dm(x)dm(y) \\
&\le \frac{1}{1 - e^{-2\pi^2[N(\Lambda^*)]^2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n/\Lambda} \|f(x) - f(x+y)\|_1 dm(x)d\gamma(y).
\end{aligned}$$

Proof. By integration it is clearly enough to deal with the case of real-valued functions, i.e. $f : \mathbb{R}^n/\Lambda \to \mathbb{R}$. Moreover, we claim that it suffices to prove the required inequality when f takes values in $\{0, 1\}$. Indeed, assuming the case of $f : \mathbb{R}^n/\Lambda \to \{0, 1\}$, we pass to the general case as follows:

$$\begin{aligned}
&\int_{(\mathbb{R}/\Lambda)\times(\mathbb{R}^n/\Lambda)} |f(x) - f(y)| dm(x)dm(y) \\
=\ & \int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \left(\int_{-\infty}^{\infty} |\mathbf{1}_{(-\infty,t]}(f(x)) - \mathbf{1}_{(-\infty,t]}(f(y))| dt \right) dm(x)dm(y) \\
\le\ & \frac{1}{1 - e^{-2\pi^2[N(\Lambda^*)]^2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n/\Lambda} \left(\int_{-\infty}^{\infty} |\mathbf{1}_{(-\infty,t]}(f(x)) - \mathbf{1}_{(-\infty,t]}(f(x+y))| dt \right) dm(x)d\gamma(y) \\
=\ & \frac{1}{1 - e^{-2\pi^2[N(\Lambda^*)]^2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n/\Lambda} |f(x) - f(x+y)| dm(x)d\gamma(y).
\end{aligned}$$

So, it remains to prove the required inequality for a measurable function $f : \mathbb{R}^n/\Lambda \to \{0,1\}$. The function f can be decomposed into a Fourier series indexed by the dual lattice Λ^*:

$$f(y) = \sum_{x\in\Lambda^*} \widehat{f}(x)e^{2\pi i\langle x,y\rangle},$$

where

$$\widehat{f}(x) = \int_{\mathbb{R}^n/\Lambda} f(y)e^{-2\pi i\langle x,y\rangle}dm(y).$$

Using the fact that $|f(x) - f(x+y)| = f(x) + f(x+y) - 2f(x)f(x+y)$ we get from Parseval's identity that for every $y \in \mathbb{R}^n$

$$\begin{aligned}\int_{\mathbb{R}^n/\Lambda} |f(x)-f(x+y)|dm(x) &= 2\widehat{f}(0) - 2\int_{\mathbb{R}^n/\Lambda}\left(\sum_{u,v\in\Lambda^*}\widehat{f}(u)\widehat{f}(v)e^{2\pi i(\langle u,x\rangle+\langle v,x+y\rangle)}\right)dm(x)\\ &= 2\widehat{f}(0) - 2\sum_{w\in\Lambda^*} e^{2\pi i\langle w,y\rangle}|\widehat{f}(w)|^2.\end{aligned}$$

Integrating with respect to the Gaussian measure, and using the identity $\int_{\mathbb{R}^n} e^{2\pi i\langle w,y\rangle}d\gamma(y) = e^{-2\pi^2\|w\|_2^2}$, we get that

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n/\Lambda} |f(x)-f(x+y)|dm(x)d\gamma(x) = 2\widehat{f}(0)[1-\widehat{f}(0)] - 2\sum_{w\in\Lambda^*\setminus\{0\}} e^{-2\pi^2\|w\|_2^2}|\widehat{f}(w)|^2. \tag{13}$$

On the other hand, since f is Boolean function we have the identities:

$$\int_{(\mathbb{R}/\Lambda)\times(\mathbb{R}^n/\Lambda)} |f(x)-f(y)|dm(x)dm(y) = 2\widehat{f}(0)[1-\widehat{f}(0)] = 2\sum_{w\in\Lambda^*\setminus\{0\}}|\widehat{f}(w)|^2. \tag{14}$$

Combining (13) and (14) we get

$$\begin{aligned}&\int_{\mathbb{R}^n}\int_{\mathbb{R}^n/\Lambda} |f(x)-f(x+y)|dm(x)d\gamma(x) = 2\sum_{w\in\Lambda^*\setminus\{0\}}\left(1-e^{-2\pi^2\|w\|_2^2}\right)|\widehat{f}(w)|^2\\ &\geq\ 2\left(1-e^{-2\pi^2[N(\Lambda^*)]^2}\right)\sum_{w\in\Lambda^*\setminus\{0\}}|\widehat{f}(w)|^2\\ &=\ \left(1-e^{-2\pi^2[N(\Lambda^*)]^2}\right)\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)}|f(x)-f(y)|dm(x)dm(y).\end{aligned}$$

□

Lemma 5.4. *For every lattice $\Lambda \subseteq \mathbb{R}^n$,*

$$\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} d_{\mathbb{R}^n/\Lambda}(x,y)dm(x)dm(y) \geq \frac{n}{16r(\Lambda^*)}.$$

Proof. Let V_Λ be the Voronoi cell of Λ centered at 0, i.e.

$$V_\Lambda = \{x \in \mathbb{R}^n : \ \|x\|_2 = d(x, \Lambda)\}.$$

Denote by B_2^n the unit Euclidean ball of $\mathbb{R}^n$ centered at 0. Then by the definition of $r(\Lambda^*)$ we have that $V_{\Lambda^*} \subseteq r(\Lambda^*) B_2^n$. Hence $\mathrm{vol}(V_{\Lambda^*}) \le [r(\Lambda^*)]^n \mathrm{vol}(B_2^n)$. It is well known (see [26, 44, 39]) that

$$\mathrm{vol}(V_\Lambda) \cdot \mathrm{vol}(V_{\Lambda^*}) = \mathrm{vol}(P_\Lambda) \cdot \mathrm{vol}(P_{\Lambda^*}) = 1.$$

Thus

$$\mathrm{vol}(V_\Lambda) \ge \frac{1}{[r(\Lambda^*)]^n \mathrm{vol}(B_2^n)}.$$

It follows that

$$\begin{aligned}
\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} d_{\mathbb{R}^n/\Lambda}(x,y)dm(x)dm(y) &= \frac{1}{\mathrm{vol}(V_\Lambda)} \int_{V_\Lambda} \|x\|_2 dx \\
&\ge \frac{n}{8r(\Lambda^*)} \cdot \frac{\mathrm{vol}\left(\left\{x \in V_\Lambda : \ \|x\|_2 \ge \frac{n}{8r(\Lambda^*)}\right\}\right)}{\mathrm{vol}(V_\Lambda)} \\
&\ge \frac{n}{8r(\Lambda^*)} \cdot \left(1 - \left(\frac{n}{8r(\Lambda^*)}\right)^n \mathrm{vol}(B_2^n) \cdot [r(\Lambda^*)]^n \mathrm{vol}(B_2^n)\right) \\
&\ge \frac{n}{16r(\Lambda^*)}.
\end{aligned}$$

□

Proof of Theorem 5.1. If $f : \mathbb{R}^n/\Lambda \to L_1$ is bi-Lipschitz then

$$\begin{aligned}
\int_{\mathbb{R}^n}\int_{\mathbb{R}^n/\Lambda} \|f(x) - f(x+y)\|_1 dm(x)d\gamma(y) &\le \|f\|_{\mathrm{Lip}} \cdot \int_{\mathbb{R}^n}\int_{\mathbb{R}^n/\Lambda} d_{\mathbb{R}^n/\Lambda}(x, x+y)dm(x)d\gamma(y) \\
&\le \|f\|_{\mathrm{Lip}} \cdot \int_{\mathbb{R}^n} \|y\|_2 d\gamma(y) \\
&\le \|f\|_{\mathrm{Lip}} \cdot \sqrt{n}.
\end{aligned}$$

On the other hand, using Lemma 5.4 we see that

$$\begin{aligned}
\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_1 dm(x)dm(y) &\ge \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}} \int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} d_{\mathbb{R}^n/\Lambda}(x,y)dm(x)dm(y) \\
&\ge \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}} \cdot \frac{n}{16r(\Lambda^*)},
\end{aligned}$$

so by Lemma 5.3 we deduce that

$$\|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}} = \Omega\left(\frac{1 - e^{-2\pi^2[N(\Lambda^*)]^2}}{r(\Lambda^*)} \cdot \sqrt{n}\right).$$

It follows that for every $t > 0$,

$$c_1(\mathbb{R}^n/\Lambda) = c_1(\mathbb{R}^n/(t\Lambda)) = \Omega\left(\frac{1 - e^{-2\pi^2[N((t\Lambda)^*)]^2}}{r((t\Lambda)^*)} \cdot \sqrt{n}\right) = \Omega\left(\frac{1 - e^{-2\pi^2[N(\Lambda^*)]^2/t^2}}{r(\Lambda^*)/t} \cdot \sqrt{n}\right).$$

Optimizing over t yields the required result. □

If one is interested only in bounding the Euclidean distortion of $\mathbb{R}^n/\Lambda$, then the following lemma gives an alternative proof of Theorem 5.1 (in the case of embeddings into L_2).

Lemma 5.5. *For every continuous $f : \mathbb{R}^n/\Lambda \to L_2$,*

$$\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_2^2 dm(x)dm(y) \le \frac{2}{\left[N(\Lambda^*)\right]^2} \int_{\mathbb{R}^n/\Lambda} \|\nabla f(x)\|_2^2 dm(x).$$

Proof. By Parseval's identity

$$\begin{aligned}
\int_{\mathbb{R}^n/\Lambda} \|\nabla f(x)\|_2^2 dm(x) &= \sum_{j=1}^n \int_{\mathbb{R}^n/\Lambda} \left(\frac{\partial f}{\partial x_j}(x)\right)^2 dm(x) \\
&= \sum_{x\in\Lambda^*} \|\widehat{f}(x)\|_2^2 \cdot \|x\|_2^2 \\
&\ge [N(\Lambda^*)]^2 \sum_{x\in\Lambda^*\setminus\{0\}} \|\widehat{f}(x)\|_2^2 \\
&= [N(\Lambda^*)]^2 \int_{\mathbb{R}^n/\Lambda} \|f(x) - \widehat{f}(0)\|_2^2 dm(x) \\
&= \frac{[N(\Lambda^*)]^2}{2} \int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_2^2 dm(x)dm(y).
\end{aligned}$$

□

Lemma 5.5 yields the lower bound $c_2(\mathbb{R}^n/\Lambda) = \Omega\left(\frac{N(\Lambda^*)}{r(\Lambda^*)} \cdot \sqrt{n}\right)$ as follows. Let $f : \mathbb{R}^n/\Lambda \to L_2$ be a bi-Lipschitz function. Since L_2 has the Radon-Nikodym property, f is differentiable almost everywhere (see [8]). Now, by Lemma 5.5,

$$\begin{aligned}
\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_2^2 dm(x)dm(y) &\le \frac{2}{\left[N(\Lambda^*)\right]^2} \sum_{j=1}^n \int_{\mathbb{R}^n/\Lambda} \left\|\frac{\partial f}{\partial x_j}\right\|_2^2 dm(x) \\
&\le \frac{2}{\left[N(\Lambda^*)\right]^2} \cdot n\|f\|_{\mathrm{Lip}}^2.
\end{aligned}$$

On the other hand, arguing as in the proof of Theorem 5.1, we get

$$\begin{aligned}
\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} \|f(x) - f(y)\|_2^2 dm(x)dm(y) &\ge \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}^2} \int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} d_{\mathbb{R}^n/\Lambda}(x,y)^2 dm(x)dm(y) \\
&= \frac{1}{\|f^{-1}\|_{\mathrm{Lip}}^2} \cdot \Omega\left(\frac{n^2}{\left[r(\Lambda^*)\right]^2}\right).
\end{aligned}$$

It follows that

$$c_2(\mathbb{R}^n/\Lambda) = \Omega\left(\frac{N(\Lambda^*)}{r(\Lambda^*)} \cdot \sqrt{n}\right).$$

The following corollary of Lemma 5.5 will not be used in the sequel, but we record it here for future reference.

Corollary 5.6. *For every continuous $f : \mathbb{R}^n/\Lambda \to \mathbb{R}$,*

$$\int_{(\mathbb{R}^n/\Lambda)\times(\mathbb{R}^n/\Lambda)} |f(x) - f(y)| dm(x) dm(y) \le \frac{2\sqrt{10}}{N(\Lambda^*)} \int_{\mathbb{R}^n/\Lambda} \|\nabla f(x)\|_2 dm(x).$$

Proof. Lemma 5.5 implies that $\lambda_1(\mathbb{R}^n/\Lambda) \ge \left[N(\Lambda^*)\right]^2$, where $\lambda_1(\mathbb{R}^n/\Lambda)$ is the smallest nonzero eigenvalue of the Laplace-Beltrami operator on $\mathbb{R}^n/\Lambda$. Since $\mathbb{R}^n/\Lambda$ has curvature 0, an inequality of Buser [16] implies that $\lambda_1(\mathbb{R}^n/\Lambda) \le 10\left[h(\mathbb{R}^n/\Lambda)\right]^2$, where $h(\mathbb{R}^n/\Lambda)$ is the Cheeger constant of $\mathbb{R}^n/\Lambda$ (Buser's inequality can be viewed as a reverse Cheeger inequality [19] when the Ricci curvature is bounded from below). Thus $h(\mathbb{R}^n/\Lambda) \ge N(\Lambda^*)/\sqrt{10}$, which is precisely the required inequality. $\square$

We end this section by showing that there exists a constant $D_n < \infty$ such that for any rank n lattice $\Lambda \subseteq \mathbb{R}^n$, $c_2(\mathbb{R}^n/\Lambda) \le D_n$.

Lemma 5.7. *Every rank n lattice $\Lambda \subseteq \mathbb{R}^n$ has a basis (over $\mathbb{Z}^n$) $x_1, \ldots, x_n$ such that for every $u_1, u_2, \ldots, u_n \in \mathbb{R}$,*

$$\frac{1}{n^{(3n-1)/2}} \cdot \left(\sum_{j=1}^n u_j^2 \|x_j\|_2^2\right)^{1/2} \le \left\|\sum_{j=1}^n u_j x_j\right\|_2 \le \sqrt{n} \cdot \left(\sum_{j=1}^n u_j^2 \|x_j\|_2^2\right)^{1/2}. \tag{15}$$

Proof. Let $\{x_1, \ldots, x_n\}$ be a basis of Λ, and denote by A the matrix whose columns are the vectors $\frac{x_1}{\|x_1\|_2}, \ldots, \frac{x_n}{\|x_n\|_2}$. If we let $\{x_1, \ldots, x_n\}$ be the Korkin-Zolotarev basis of Λ, we can ensure that (see [33]):

$$|\det(A)| \ge \frac{1}{n^n}.$$

Denote by $s_1(A) \ge s_2(A) \ge \cdots s_n(A) > 0$ the singular values of A. Given a vector $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ we have by the Cauchy-Schwartz inequality that

$$\|Au\|_2 = \left\|\sum_{j=1}^n \frac{u_j}{\|x_j\|_2} \cdot x_j\right\|_2 \le \sum_{j=1}^n |u_j| \le \sqrt{n} \cdot \|u\|_2.$$

This proves the right-hand side of (15), and also shows that $s_1(A) \le \sqrt{n}$. Now

$$\frac{1}{n^n} \le |\det(A)| = \prod_{j=1}^n s_j(A) \le s_1(A) \cdot [s_n(A)]^{n-1} \le s_1(A) \cdot n^{(n-1)/2},$$

i.e. $s_1(A) \ge n^{-(3n-1)/2}$. It follows that for every $u \in R^n$, $\|Au\|_2 \ge s_1(A)\|u\|_2 \ge n^{-(3n-1)/2}\|u\|_2$, which is precisely the left-hand side of (15). $\square$

Theorem 5.8. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice of rank n. Then $\mathbb{R}^n/\Lambda$ embeds into $\mathbb{R}^{2n}$ with distortion $O(n^{3n/2})$.*

Proof. Let $\{x_1, \dots, x_n\}$ be a basis as in Lemma 5.7. Define $f : \mathbb{R}^n \to \mathbb{C}^n$ by

$$f\left(\sum_{j=1}^{n} a_j x_j\right) = \left(\|x_1\|_2 e^{2\pi i a_1}, \dots, \|x_n\|_2 e^{2\pi i a_n}\right).$$

Since f is Λ-invariant, we may think of it a a function defined on the torus $\mathbb{R}^n/\Lambda$. For every $t \in \mathbb{R}$ let $m(t)$ be the unique integer such that $t - m(t) \in [-1/2, 1/2)$. Given $u, v \in \mathbb{R}^n$,

$$\begin{aligned}\left\| f\left(\sum_{j=1}^{n} u_j x_j\right) - f\left(\sum_{j=1}^{n} v_j x_j\right)\right\|_2^2 &= \sum_{j=1}^{n} \left| e^{2\pi i (u_j - v_j)} - 1\right|^2 \cdot \|x_j\|_2^2 \\ &= 2\sum_{j=1}^{n} [1 - \cos(2\pi(u_j - v_j))] \cdot \|x_j\|_2^2.\end{aligned}$$

Since for every $t \in \mathbb{R}$,

$$\frac{[t - m(t)]^2}{12} \le 1 - \cos(2\pi t) \le \frac{[t - m(t)]^2}{2},$$

we get that

$$\left\| f\left(\sum_{j=1}^{n} u_j x_j\right) - f\left(\sum_{j=1}^{n} v_j x_j\right)\right\|_2^2 = \Theta\left(\sum_{j=1}^{n} [u_j - v_j - m(u_j - v_j)]^2 \|x_j\|_2^2\right).$$

On the other hand, by (15),

$$\begin{aligned} d_{\mathbb{R}^n/\Lambda}\left(\sum_{j=1}^{n} u_j x_j, \sum_{j=1}^{n} v_j x_j\right) &= d_{\mathbb{R}^n}(\sum_{j=1}^{n}(u_j - v_j)x_j, \Lambda) \\ &\le \left\|\sum_{j=1}^{n}(u_j - v_j)x_j - \sum_{j=1}^{n} m(u_j - v_j)x_j\right\|_2 \\ &\le \sqrt{n}\left(\sum_{j=1}^{n}[u_j - v_j - m(u_j - v_j)]^2 \|x_j\|_2^2\right)^{1/2}.\end{aligned}$$

In the reverse direction, let $m_1, \dots, m_n \in \mathbb{Z}$ be such that $\sum_{j=1}^{n} m_j x_j \in \Lambda$ is a closest lattice point to $u - v$. Then

$$\begin{aligned} d_{\mathbb{R}^n/\Lambda}\left(\sum_{j=1}^{n} u_j x_j, \sum_{j=1}^{n} v_j x_j\right) &= \left\|\sum_{j=1}^{n}[u_j - v_j - m_j]x_j\right\|_2 \\ &\ge \frac{1}{n^{(3n-1)/2}}\left(\sum_{j=1}^{n}[u_j - v_j - m_j]^2 \cdot \|x_j\||_2^2\right)^{1/2} \\ &\ge \frac{1}{n^{(3n-1)/2}}\left(\sum_{j=1}^{n}[u_j - v_j - m(u_j - v_j)]^2 \cdot \|x_j\|_2^2\right)^{1/2}.\end{aligned}$$

It follows that f has distortion $O(n^{3n/2})$. $\square$

6 Length of metric spaces

The following definition, due to G. Schechtman [56], plays an important role in the study of the concentration of measure phenomenon and Levy families [45, 34].

Definition 6.1. *Let (X, d) be a finite metric space. The length of (X, d), denoted $\ell(X, d)$ is the least constant ℓ such that there exists a sequence of partitions of X, $P^0, P^1, \ldots, P^N$ with the following properties:*

1. *For every $i \geq 1$, P^i is a refinement of P^{i-1}.*
2. *$P^0 = \{X\}$ and $P^N = \{\{x\} : \ x \in X\}$.*
3. *For every $i \geq 1$ there exists $a_i > 0$ such that if $A \in P^{i-1}$ and $B, C \in P^i$ are such that $B, C \subseteq A$, then there exists a one-to-one onto function $\phi = \phi_{B,C} : B \to C$ such that for every $x \in B$, $d(x, \phi(x)) \leq a_i$.*
4. *$\ell = \sqrt{\sum_{i=1}^N a_i^2}$.*

For $p \geq 1$ we can can define an analogous concept if we demand that $\ell = \left(\sum_{i=1}^N a_i^p\right)^{1/p}$. In this case we call the parameter obtained the ℓ_p length of (X, d), and denote it by $\ell_p(X, d)$. Observe that it is always the case that $\ell_p(X, d) \leq \operatorname{diam}(X)$.

Recall that for $p \in [1, 2]$, a Banach space Y is called p-smooth with constant S if for every $x, y \in Y$,

$$\|x + y\|_Y^p + \|x - y\|_Y^p \leq 2\|x\|_Y^p + 2S^p\|y\|_Y^p.$$

The least constant S for which this inequality holds is called the p-smoothness constant of Y, and is denoted $S_p(Y)$. It is known [4] that for $q \geq 2$, $S_2(L_q) \leq \sqrt{q-1}$, and for $q \in [1, 2]$, $S_q(L_q) \leq 1$.

The following theorem relates the notion of length to nonembeddability results.

Theorem 6.2. *Let (X, d) be a metric space and Y a p-smooth Banach space. Then*

$$c_Y(X, d) \geq \frac{1}{2^{1-1/p} \cdot S_p(Y)\ell_p(X, d)} \left(\frac{1}{|X|^2} \sum_{x,y \in X} d(x, y)^p\right)^{1/p}.$$

In particular for $2 \leq p < \infty$,

$$c_p(X, d) \geq \frac{1}{\ell(X, d)\sqrt{2p-2}} \left(\frac{1}{|X|^2} \sum_{x,y \in X} d(x, y)^2\right)^{1/2}.$$

Proof. Let $\{P^i\}_{i=0}^N$, $\{a_i\}_{i=1}^N$ be as above, and denote by $\mathcal{F}_i$ the σ-algebra generated by the partition P^i. In what follows all expectations are taken with respect to the uniform probability measure on X. Given a bijection $f : X \to Y$ we let $f_i = \mathbb{E}(f|\mathcal{F}_i)$. In other words, if $A \in P^i$ and $x \in A$ then

$$f_i(x) = \frac{1}{|A|} \sum_{y \in A} f(y).$$

Now $\{f_i\}_{i=0}^N$ is a martingale, so by Pisier's inequality [53] (see Theorem 4.2 in [48] for the constant we use below), we see that

$$\mathbb{E}\|f_N - f_0\|_Y^p \le \frac{S_p(Y)^p}{2^{p-1}-1}\sum_{j=0}^{N-1}\mathbb{E}\|f_{j+1}-f_j\|_Y^p.$$

Now $f_0 = \mathbb{E}f$ and $f_N = f$. Thus

$$\begin{aligned}
\mathbb{E}\|f_N - f_0\|_Y^p &= \frac{1}{|X|}\sum_{x\in X}\left\|f(x) - \frac{1}{|X|}\sum_{y\in X} f(y)\right\|_Y^p \\
&\ge \frac{1}{2^{p-1}|X|^2}\sum_{x,y\in X}\|f(x)-f(y)\|_Y^p \\
&\ge \frac{1}{2^{p-1}\|f^{-1}\|_{\mathrm{Lip}}^p}\cdot\frac{1}{|X|^2}\sum_{x,y\in X} d(x,y)^p.
\end{aligned}$$

On the other hand fix $j\in\{0,\dots,N-1\}$, and $A\in P^j$, $B\in P^{j+1}$ such that $x\in B\subseteq A$. Then

$$f_j(x) - f_{j+1}(x) = \frac{1}{|A|}\sum_{y\in A} f(y) - \frac{1}{|B|}\sum_{y\in B} f(y) = \frac{1}{|A|}\sum_{A\supseteq C\in P^{j+1}}\left(\sum_{y\in C}[f(\phi_{C,B}(y)) - f(y)]\right).$$

So by convexity

$$\|f_j(x) - f_{j+1}(x)\|_Y \le \|f||_{\mathrm{Lip}}\cdot a_{j+1}.$$

It follows that

$$\begin{aligned}
c_Y(X,d) &\ge \frac{1}{2^{1-1/p}\cdot S_p(Y)}\cdot\left(\frac{\frac{1}{|X|^2}\sum_{x,y\in X} d(x,y)^p}{\sum_{j=1}^N a_j^p}\right)^{1/p} \\
&= \frac{1}{2^{1-1/p}\cdot S_p(Y)\ell_p(X,d)}\left(\frac{1}{|X|^2}\sum_{x,y\in X} d(x,y)^p\right)^{1/p}.
\end{aligned}$$

□

As shown in [45, 34], if we consider the group of permutations of $\{1,\dots,n\}$, S_n, equipped with the metric $d(\sigma,\pi) = |\{i:\ \sigma(i)\ne\tau(i)\}|$, then $\ell(S_n,d)\le 2\sqrt{n}$, while $\mathrm{diam}(S_n) = \Theta(n)$. It follows from Theorem 6.2 that $c_2(S_n) = \Omega(\sqrt{n})$. On the other hand, by mapping each permutation $\pi\in S_n$ to the matrix $(\mathbf{1}_{\pi(i)=j})$ we see that $c_2(S_n) = O(\sqrt{n})$. Thus

$$c_2(S_n) = \Theta\left(\sqrt{\frac{\log|S_n|}{\log\log|S_n|}}\right).$$

Similar optimal bounds can be deduced for $c_p(S_n)$, $p\ge 1$.

The metric d on S_n is the shortest path metric induced by the Cayley graph on S_n obtained by taking the set of all transpositions as generators. It is of interest to study the Euclidean distortion of metrics on S_n induced by Cayley graphs coming from other generating sets. In particular, it is a long standing conjecture (see [55]) that there exists a bounded set of generators of S_n with respect to which the Cayley graph is an expander. It is thus natural to ask whether there exists a set of generators of S_n with respect to which the metric induced by the Cayley graph has Euclidean distortion $\Omega(\log |S_n|) = \Omega(\log n \log\log n)$.

Another example discussed in [45, 34] is the case of the Hamming cube. In this case $\ell(\mathbb{F}_2^d, \rho) = O(\sqrt{d})$, and so Theorem 6.2 implies that for $p \geq 2$, $c_p(\mathbb{F}_2^d, \rho) \geq c(p)\sqrt{d}$. This result was first proved in [50].

More generally, let G be a finite group equipped with a translation invariant metric d. Let $G = G_0 \supseteq G_1 \supseteq \cdot \supseteq G_n = \{e\}$ be a decreasing sequence of subgroups. Then it is shown in [45, 34] that

$$\ell(G, d) \leq \sqrt{\sum_{j=1}^{n} \left[\operatorname{diam}(G_{i-1}/G_i)\right]^2}.$$

This estimate implies a wide range of additional nonembeddability results.

7 Acknowledgments

We are grateful to Keith Ball, Henry Cohn, Piotr Indyk, László Lovász and Gideon Schechtman for helpful suggestions.

References

[1] A. Andoni, M. Deza, A. Gupta, P. Indyk, and S. Raskhodnikova. Lower bounds for embedding edit distance into normed spaces. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 523–526. Society for Industrial and Applied Mathematics, 2003.

[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] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

[4] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.

[5] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. Approximating edit distance efficiently. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 550–559. IEEE, Oct. 2004.

[6] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. Math.*, 2005. To appear.

[7] J. Bates, W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Affine approximation of Lipschitz functions and nonlinear quotients. *Geom. Funct. Anal.*, 9(6):1092–1127, 1999.

[8] 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.

[9] A. Beutelspacher and U. Rosenbaum. *Projective geometry: from foundations to applications*. Cambridge University Press, Cambridge, 1998.

[10] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

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

[12] J. Bourgain. On the distributions of the Fourier spectrum of Boolean functions. *Israel J. Math.*, 131:269–276, 2002.

[13] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geom. Funct. Anal.*, 7(3):438–461, 1997.

[14] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.

[15] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[16] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982.

[17] M. Charikar and R. Krauthgamer. Embedding edit distance submetrics into ℓ_1. Manuscript, 2005.

[18] 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.

[19] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

[20] 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.

[21] G. Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. In *SODA '02: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 667–676. Society for Industrial and Applied Mathematics, 2002.

[22] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[23] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105 (1969), 1969.

[24] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[25] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[26] P. M. Gruber and C. G. Lekkerkerker. *Geometry of numbers*, volume 37 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1987.

[27] 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.

[28] D. Gusfield. *Algorithms on strings, trees, and sequences.* Cambridge University Press, Cambridge, 1997. Computer science and computational biology.

[29] P. Indyk. Approximate nearest neighbor under edit distance via product metrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 646–650. Society for Industrial and Applied Mathematics, 2004.

[30] 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.

[31] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *29th Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[32] S. Khot and N. Vishnoi. On embeddability of negative type metrics into ℓ_1. Manuscript, 2005.

[33] J. C. Lagarias, H. W. Lenstra JR., and C. P. Schnorr. Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice. *Combinatorica*, 10(4):333–348, 1990.

[34] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[35] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Dokl.*, 10:707–710, 1965.

[36] N. Linial. Finite metric spaces - combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians III*, pages 573–586, 2002.

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

[38] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I.* North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16.

[39] J. Martinet. *Perfect lattices in Euclidean spaces*, volume 327 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.

[40] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[41] J. Matoušek. Open problems on embeddings of finite metric spaces. *Available at http://kam.mff.cuni.cz/∼matousek/metrop.ps.gz*, 2002.

[42] M. Mendel and A. Naor. Euclidean quotients of finite metric spaces. *Adv. Math.*, 189(2):451–494, 2004.

[43] M. Mendel and A. Naor. Metric cotype. Manuscript, 2005.

[44] D. Micciancio and S. Goldwasser. *Complexity of lattice problems.* The Kluwer International Series in Engineering and Computer Science, 671. Kluwer Academic Publishers, Boston, MA, 2002. A cryptographic perspective.

[45] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[46] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. Manuscript, 2005.

[47] S. Muthukrishnan and S. C. Sahinalp. Approximate nearest neighbors and sequence comparison with block operations. In *STOC 2000: Proceedings of the 32th Annual ACM Symposium on Theory of Computing*, pages 416–424, 2000.

[48] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov hyperbolic metric spaces. Preprint, 2004.

[49] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. *J. Funct. Anal.* To appear.

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

[51] J. Nash. C^1 isometric imbeddings. *Ann. of Math. (2)*, 60:383–396, 1954.

[52] R. Ostrovsky and Y. Rabani. Low distortion embeddings for edit distance. In *STOC 2005: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 2005. To appear.

[53] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.

[54] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[55] E. Rozenman, A. Shalev, and A. Wigderson. A new family of Cayley expanders? In *STOC 2004: Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 445–454, 2004.

[56] G. Schechtman. Lévy type inequality for a class of finite metric spaces. In *Martingale theory in harmonic analysis and Banach spaces (Cleveland, Ohio, 1981)*, volume 939 of *Lecture Notes in Math.*, pages 211–215. Springer, Berlin, 1982.

[57] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces.* Princeton University Press, Princeton, N.J., 1971. Princeton Mathematical Series, No. 32.

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

[59] C. Zong. From deep holes to free planes. *Bull. Amer. Math. Soc. (N.S.)*, 39(4):533–555 (electronic), 2002.