ON LIPSCHITZ EXTENSION FROM FINITE SUBSETS

ASSAF NAOR AND YUVAL RABANI

Abstract. We prove that for every $n \in \mathbb{N}$ there exists a metric space (X, d_X), an n-point subset $S \subseteq X$, a Banach space $(Z, \|\cdot\|_Z)$ and a 1-Lipschitz function $f : S \to Z$ such that the Lipschitz constant of every function $F : X \to Z$ that extends f is at least a constant multiple of $\sqrt{\log n}$. This improves a bound of Johnson and Lindenstrauss [JL84]. We also obtain the following quantitative counterpart to a classical extension theorem of Minty [Min70]. For every $\alpha \in (1/2, 1]$ and $n \in \mathbb{N}$ there exists a metric space (X, d_X), an n-point subset $S \subseteq X$ and a function $f : S \to \ell_2$ that is α-Hölder with constant 1, yet the α-Hölder constant of any $F : X \to \ell_2$ that extends f satisfies

$$\|F\|_{\mathrm{Lip}(\alpha)} \gtrsim (\log n)^{\frac{2\alpha-1}{4\alpha}} + \left(\frac{\log n}{\log\log n}\right)^{\alpha^2-\frac{1}{2}}.$$

We formulate a conjecture whose positive solution would strengthen Ball's nonlinear Maurey extension theorem [Bal92], serving as a far-reaching nonlinear version of a theorem of König, Retherford and Tomczak-Jaegermann [KRTJ80]. We explain how this conjecture would imply as special cases answers to longstanding open questions of Johnson and Lindenstrauss [JL84] and Kalton [Kal04].

1. Introduction

Given two metric spaces (X, d_X) and (Z, d_Z), the Lipschitz constant of a mapping $f : X \to Z$ will be denoted below by $\|f\|_{\mathrm{Lip}}$. For every subset $S \subseteq X$, let $e(X, S, Z)$ denote the infimum over those $K \in (0, \infty]$ with the property that for every $f : S \to Z$ there exists $F : X \to Z$ with $\|F\|_{\mathrm{Lip}} \leqslant K\|f\|_{\mathrm{Lip}}$ and whose restriction to S satisfies $F|_S = f$. In its most general form, the *Lipschitz extension problem* asks for estimates on the quantity $e(X, S, Z)$.

It is of great interest to obtain geometric conditions on the metric spaces (X, d_X) and (Z, d_Z) ensuring that $e(X, S, Z) < \infty$ for every $S \subseteq X$. Formally, define $e(X, Z)$ to be the supremum of $e(X, S, Y)$ over all $S \subseteq X$. When $e(X, Z) = \infty$, it is natural to refine the Lipschitz extension problem by taking $n \in \mathbb{N}$ and defining $e_n(X, Z)$ to be the supremum of $e(X, S, Z)$ over all $S \subseteq X$ of cardinality at most n, and asking for the asymptotic behavior of $e_n(X, Z)$ as $n \to \infty$. Another natural quantitative refinement of the parameter $e(X, Z)$ is to fix $\varepsilon \in (0, 1]$ and define $e_\varepsilon(X, Z)$ to be the supremum of $e(X, S, Z)$ over all $S \subseteq X$ that are ε-discrete in the sense that $d_X(x, y) \geqslant \varepsilon\cdot\mathrm{diam}(S)$ for every distinct $x, y \in S$, and asking for the asymptotic behavior of $e_\varepsilon(X, Z)$ as $\varepsilon \to 0$.

Denote the supremum of $e_n(X, Z)$ over all metric spaces (X, d_X) and all Banach spaces $(Z, \|\cdot\|_Z)$ by $\mathfrak{ae}(n)$. Here "$\mathfrak{ae}$" stands for "absolute extendability," where we are following the notation and terminology that was introduced in [LN04]. Thus, if $\mathfrak{ae}(n) < K$ then for *every* n-point metric space (M, d_M), *every* Banach-space valued 1-Lipschitz function defined on M can be extended to *any* metric space that (isometrically) contains (M, d_M) so that the Lipschitz constant of the extended function is at most K. One can similarly define for $\varepsilon > 0$ the quantity $\mathfrak{ae}(\varepsilon)$ by considering the supremum of $e_\varepsilon(X, Z)$ over all metric spaces (X, d_X) and all Banach spaces $(Z, \|\cdot\|_Z)$.

A. N. was supported in part by the NSF, the BSF, the Packard Foundation and the Simons Foundation.
Y. R. was supported in part by the ISF, the BSF, and by the Israeli Center of Excellence on Algorithms.

The finiteness of $\mathfrak{ae}(n)$ and $\mathfrak{ae}(\varepsilon)$ for every $n \in \mathbb{N}$ and $\varepsilon \in (0,1]$ is well-known. Johnson, Lindenstrauss and Schechtman proved in [JLS86] that $\mathfrak{ae}(n) \lesssim \log n$, where here, and in what follows, the notation $A \lesssim B$ (respectively $A \gtrsim B$) stands for $A \leqslant CB$ (respectively $A \geqslant CB$) for some universal constant $C \in (0,\infty)$. We shall also use the notation $A \asymp B$ when $A \lesssim B$ and $A \gtrsim B$. The above upper bound of [JLS86] on $\mathfrak{ae}(n)$ was more recently improved [LN04, LN05] to $\mathfrak{ae}(n) \lesssim (\log n)/\log\log n$, which is the best-known upper bound on $\mathfrak{ae}(n)$ to date. Determining the asymptotic behavior of $\mathfrak{ae}(n)$ as $n \to \infty$ remains a major open problem. The previously best-known lower bound on $\mathfrak{ae}(n)$ is $\mathfrak{ae}(n) \gtrsim \sqrt{(\log n)/\log\log n}$, obtained over thirty years ago by Johnson and Lindenstrauss [JL84] (see also [MM10] for a variant of the construction of Johnson and Lindenstrauss that also yields the same lower bound). Here we obtain the following improvement.

Theorem 1. *For every $n \in \mathbb{N}$ we have $\mathfrak{ae}(n) \gtrsim \sqrt{\log n}$.*

Thus, the best-known bounds for $\mathfrak{ae}(n)$ are now

$$\sqrt{\log n} \lesssim \mathfrak{ae}(n) \lesssim \frac{\log n}{\log\log n}. \tag{1}$$

It would be very interesting to improve any of the bounds in (1), the ultimate goal being to determine the rate at which $\mathfrak{ae}(n)$ tends to ∞ as $n \to \infty$.

In [JL84] it was shown that $\mathfrak{ae}(\varepsilon) \lesssim 1/\varepsilon$, and a different and very simple proof of this fact was given in [JLS86]. In [MN13a] it was proved that $\mathfrak{ae}(\varepsilon) \gtrsim 1/\varepsilon$, via a quantitative refinement of a beautiful argument of Kalton [Kal12]. Thus $\mathfrak{ae}(\varepsilon) \asymp 1/\varepsilon$. The example that leads to Theorem 1 also yields a new proof that $\mathfrak{ae}(\varepsilon) \gtrsim 1/\varepsilon$ which some may find to be somewhat simpler than the proof in [MN13a]; see Remark 14 below. We note, however, that the example of [MN13a] has important properties (see the discussion in Section 1.6 of [MN13a]) that Theorem 1 does not imply.

The value of Theorem 1 does not stem only from the fact that it yields an asymptotic improvement over the best known bound, but rather because this improvement relies on a conceptually different approach than that of [JL84]. In fact, as emphasized explicitly in [JL84], the difficulty to improve the lower bound on $\mathfrak{ae}(n)$ to a constant multiple of $\sqrt{\log n}$ is not purely technical, and the approach of [JL84] inherently cannot yield a lower bound that is better than $o(\sqrt{\log n})$.

To explain the above assertions, we need to briefly sketch the argument of [JL84]. Fix $k \in \mathbb{N}$, $L > 0$ and $\varepsilon \in (0,1/2)$. Suppose that $(X, \|\cdot\|_X)$ is a Banach space and $Y \subseteq X$ is a k-dimensional linear subspace of X. Let $\mathcal{N}_\varepsilon$ be an ε-net in the unit sphere of Y and denote $S_\varepsilon = \mathcal{N}_\varepsilon \cup \{0\}$. Taking $f : S_\varepsilon \to S_\varepsilon \subseteq Y$ to be the identity mapping, it is proved in [JL84] that if $F : X \to Y$ is an L-Lipschitz extension of f and ε is sufficiently small then there exists a linear projection $P : X \to Y$ with $\|P\| \lesssim L$. Hence, if for some $\Lambda > 0$ one knows that $\|P\| \geqslant \Lambda$ for *every* linear projection $P : X \to Y$ then one concludes that $L \gtrsim \Lambda$. Suppose that $\Lambda \gtrsim \sqrt{k}$; since the classical Kadec′–Snobar theorem [KS71] (see also [Woj91, §III.B]) asserts that we always have $\Lambda \leqslant \sqrt{k}$, this lower bound on Λ is the best one could hope for. By standard bounds on the size of ε-nets (e.g. [MS86]), if we set $n = |S_\varepsilon|$ then $\log n \asymp k\log(1/\varepsilon)$, so we have $L \gtrsim \sqrt{(\log n)/\log(1/\varepsilon)}$. For this strategy to yield a lower bound that is a constant multiple of $\sqrt{\log n}$, one needs to take ε to be a universal constant, but in this case we would have $L \lesssim \mathfrak{ae}(2\varepsilon) \lesssim 1/\varepsilon = O(1)$. So, in order to get a lower bound on L that tends to ∞ with n one *must have* $\varepsilon = o(1)$ as $k \to \infty$ (indeed, in [JL84] the choice of ε is $\log(1/\varepsilon) \asymp \log k \asymp \log\log n$). This explains why one cannot prove Theorem 1 via the above strategy of reduction to the nonexistence of linear projections of small norm. Note also that due to the Kadec′–Snobar theorem [KS71], this "linearization" approach cannot yield a lower bound on $\mathfrak{ae}(n)$ that tends to ∞ at a rate that is faster than a constant multiple of $\sqrt{\log n}$.

Here we overcome the above obstacle by abandoning entirely the desire to reduce the problem to the nonexistence of linear projections of small norm. We consider a family of metric spaces that arise from (a modification of) expander graphs, and our poorly extendable functions take values in their associated Wasserstein 1 spaces. The proof relies on direct geometric considerations, in particular the use of Banach space-valued Poincaré inequalities (see Lemma 8 below), rather than a reduction to linear mappings. Our approach is inspired by a certain algorithmic clustering problem for graphs, specifically by leveraging the difference between two relaxations of the corresponding combinatorial optimization problem. This motivation is explained in Section 2 below, though the proof of Theorem 1 (in Section 3) is direct and does not rely on any algorithmic background.

1.1. **Hilbert space-valued Hölder functions.** Suppose that (X, d_X) and (Z, d_Z) are metric spaces and $\alpha \in (0,1]$. We shall denote below the α-Hölder constant of a mapping $f : X \to Z$ by $\|f\|_{\mathrm{Lip}(\alpha)}$. Thus $\|f\|_{\mathrm{Lip}(\alpha)}$ is the infimum over those $L \in (0,\infty]$ for which every $x, y \in X$ satisfy $d_Z(f(x), f(y)) \leqslant L d_X(x,y)^\alpha$. Equivalently, $\|f\|_{\mathrm{Lip}(\alpha)}$ is the Lipschitz constant of f when it is viewed as a mapping between from the metric space (X, d_X^α) to the metric space (Z, d_Z). Denote

$$e^\alpha(X,Z) \stackrel{\mathrm{def}}{=} e\big((X,d_X^\alpha),(Z,d_Z)\big) \qquad \text{and} \qquad \forall\, n\in\mathbb{N}, \qquad e_n^\alpha(X,Z) \stackrel{\mathrm{def}}{=} e_n\big((X,d_X^\alpha),(Z,d_Z)\big).$$

Thus, $e^\alpha(X,Z)$ is the infimum over those $K \in [1,\infty]$ such that for every $S \subseteq X$ and every $f : S \to Z$ there exists $F : X \to Z$ with $F|_S = f$ and $\|F\|_{\mathrm{Lip}(\alpha)} \leqslant K\|f\|_{\mathrm{Lip}(\alpha)}$, and analogously for $e_n^\alpha(X,Z)$.

A classical theorem of Minty [Min70] asserts that if H is a Hilbert space then $e^\alpha(X,H) = 1$ for every metric space (X, d_X) and $\alpha \in (0,1/2]$ (equivalently, $e_n^\alpha(\ell_\infty,\ell_2) = 1$ for all $n \in \mathbb{N}$). Minty's theorem fails when $\alpha \in (1/2,1]$ (see [HW71]), but understanding what happens when $\alpha \in (1/2,1]$ remains a mystery. Specifically, Kalton conjectured in [Kal04] that for every $\alpha \in [1/2,1]$ and every $m \in \mathbb{N}$ we have

$$e^\alpha(\ell_\infty,\ell_2^m) \lesssim m^{\alpha-\frac12}. \tag{2}$$

This conjecture appears as Problem 11.3 in [Kal04][1]. See the discussion immediately following Problem 11.3 in [Kal04] for an interesting geometric application that (2) would imply. The estimate (2) holds true when $\alpha = 1/2$ due to Minty's theorem, and also when $\alpha = 1$ by the fact that ℓ_2^m is $\sqrt{m}$-isomorphic to ℓ_∞^m, combined with an application of the nonlinear Hahn–Banach theorem (see e.g. [BL00, Lemma 1.1]). Thus (2) is a natural conjectural interpolation between Minty's theorem and the nonlinear Hahn-Banach theorem.

Kalton's conjecture (2) remains an interesting open problem, and in Section 1.2 we formulate a conjecture whose validity would imply the validity of (2). In the reverse direction, to the best of our knowledge its isn't known whether (2) would be sharp. We therefore ask the following question.

Question 2. Is it true that $e^\alpha(\ell_\infty,\ell_2^m) \gtrsim m^{\alpha-\frac12}$ for every $\alpha \in (1/2,1)$?

We also ask the following variant of Question 2 for the parameter $e_n^\alpha(\ell_\infty,\ell_2)$.

Question 3. Is it true that $e_n^\alpha(\ell_\infty,\ell_2) \gtrsim (\log n)^{\alpha-\frac12}$ for every $\alpha \in (1/2,1)$?

By mimicking an argument of [JL84] one sees that a positive answer to Question 3 would imply a positive answer to Question 2. Indeed, for every metric space (X, d_X) and $\alpha \in (0,1]$ we have

$$e_{2^{7m}}^\alpha(X,\ell_2) \leqslant 3e^\alpha(\ell_\infty,\ell_2^m). \tag{3}$$

(The choice of constants in (3) is somewhat arbitrary; see below.) Since (3) is not stated explicitly in [JL84], we shall now briefly explain how it is proved. Fix $S \subseteq X$ with $|S| \leqslant 2^{7m}$ and $f : S \to \ell_2$. Since $f(S)$ is a subset of a Hilbert space of cardinality at most 2^{7m}, by the Johnson–Lindenstrauss dimensionality reduction lemma [JL84] there exists a mapping $g : f(S) \to \ell_2^m$ such that

$$\forall\, a,b \in f(S), \qquad \|a-b\|_2 \leqslant \|g(a)-g(b)\|_2 \leqslant 3\|a-b\|_2. \tag{4}$$

[1] We note in passing that there is a misprint in [Kal04, Problem 11.3]: in the two displayed equations that appear there the exponent α erroneously appears in the left hand side of the inequality rather than in its right hand side.

The specific parameters in the Johnson–Lindenstrauss lemma that we take here are valid due to the bounds in [Ach03, Theorem 1.1]. By the Kirszbraun extension theorem [Kir34],

$$\exists\, \widetilde{g^{-1}} : \ell_2^m \to \ell_2 \ \text{ such that } \ \widetilde{g^{-1}}\Big|_{g\circ f(S)} = g^{-1} \ \text{ and } \ \left\|\widetilde{g^{-1}}\right\|_{\mathrm{Lip}} \leqslant \left\|g^{-1}\right\|_{\mathrm{Lip}} \leqslant 1, \tag{5}$$

where the last step of (5) uses the left hand inequality in (4). Also, by the definition of $e^{\alpha}(\ell_\infty, \ell_2^m)$,

$$\exists\, \widetilde{g\circ f} : X \to \ell_2^m \ \text{ such that } \ \widetilde{g\circ f}\Big|_S = g\circ f \ \text{ and } \ \left\|\widetilde{g\circ f}\right\|_{\mathrm{Lip}(\alpha)} \leqslant e^{\alpha}(\ell_\infty, \ell_2^m)\, \|g\circ f\|_{\mathrm{Lip}(\alpha)}. \tag{6}$$

Thus, the following diagram commutes.

$$\begin{tikzcd}
X \arrow[rr, "\widetilde{g^{-1}}\circ\widetilde{g\circ f}"] \arrow[rrrr, bend left, "\widetilde{g\circ f}"] & & \ell_2 & & \ell_2^m \arrow[ll, "\widetilde{g^{-1}}"'] \\
S \arrow[u, hook, "\subseteq"] \arrow[rr, "f"] & & f(S) \arrow[u, hook, "\subseteq"] \arrow[rr, bend left, "g"] & & g\circ f(S) \arrow[u, hook, "\subseteq"] \arrow[ll, bend left, "g^{-1}"]
\end{tikzcd}$$

The mapping $\widetilde{g^{-1}}\circ\widetilde{g\circ f} : X \to \ell_2$ therefore extends f and satisfies

$$\begin{aligned}\left\|\widetilde{g^{-1}}\circ\widetilde{g\circ f}\right\|_{\mathrm{Lip}(\alpha)} \leqslant \left\|\widetilde{g^{-1}}\right\|_{\mathrm{Lip}} \cdot \left\|\widetilde{g\circ f}\right\|_{\mathrm{Lip}(\alpha)} &\overset{(5)\wedge(6)}{\leqslant} e^{\alpha}(\ell_\infty, \ell_2^m)\, \|g\circ f\|_{\mathrm{Lip}(\alpha)} \\ &\leqslant e^{\alpha}(\ell_\infty, \ell_2^m)\|g\|_{\mathrm{Lip}}\|f\|_{\mathrm{Lip}(\alpha)} \leqslant 3e^{\alpha}(\ell_\infty, \ell_2^m)\|f\|_{\mathrm{Lip}(\alpha)},\end{aligned} \tag{7}$$

where the last step of (7) uses the right hand inequality in (4). Clearly (7) implies (3).

The link between Question 2 and Theorem 1 is twofold. Firstly, Question 2 presents another situation in which the linearization strategy of [JL84] for proving extension lower bounds seems to be insufficient, because bounded linear maps are Lipschitz rather than α-Hölder for $\alpha \in (0,1)$. Nevertheless, in Section 4.2 we do show how to use a linearization procedure in the spirit of [JL84] to bound $e^{\alpha}(\ell_\infty, \ell_2^m)$ from below. However, our approach requires a substantial modification of the argument of [JL84] and at present we do not see how it could yield a nontrivial result for the entire range $\alpha \in (1/2, 1]$ (the argument of Section 4.2 yields nontrivial bounds only when $\alpha \geqslant 1/\sqrt{2}$). The second link between Question 2 and Theorem 1 is that the bounds that we obtain in Theorem 4 below are based on a modification of a construction of [JLS86] which is part of a general family of constructions that were used in related contexts also in [Lan99, CKR05], and to which the example that underlies Theorem 1 belongs as well. We elaborate further on this in Remark 19 below.

Here we obtain the best known lower bounds in the context of Question 2 and Question 3.

Theorem 4. *For every $\alpha \in (1/2, 1]$ and every $m, n \in \mathbb{N}$ we have*

$$e^{\alpha}(\ell_\infty, \ell_2^m) \gtrsim m^{\frac{2\alpha-1}{4\alpha}} + m^{\alpha^2-\frac12} \qquad \text{and} \qquad e_n^{\alpha}(\ell_\infty, \ell_2) \gtrsim (\log n)^{\frac{2\alpha-1}{4\alpha}} + \left(\frac{\log n}{\log\log n}\right)^{\alpha^2-\frac12}. \tag{8}$$

Importantly, the lower bounds in (8) tend to ∞ with n for every $\alpha \in (1/2, 1]$. A positive answer to Question 2 (respectively, Question 3) for $\alpha = \frac12 + \varepsilon$ would yield the lower bound $e^{\alpha}(\ell_\infty, \ell_2^m) \gtrsim m^{\varepsilon}$ (respectively, $e_n^{\alpha}(\ell_\infty, \ell_2) \gtrsim (\log n)^{\varepsilon}$), while Theorem 4 implies that

$$e^{\frac12+\varepsilon}(\ell_\infty, \ell_2^m) \gtrsim m^{\varepsilon-2\varepsilon^2} \qquad \text{and} \qquad e_n^{\frac12+\varepsilon}(\ell_\infty, \ell_2) \gtrsim (\log n)^{\varepsilon-2\varepsilon^2},$$

i.e., the exponents of Theorem 4 match what Kalton's conjecture predicts up to lower order terms as $\alpha \to 1/2^+$. The exponent $\alpha^2 - 1/2$ in the second summands in (8) becomes positive only when $\alpha \geqslant 1/\sqrt{2}$, and it becomes greater than the exponent of the first summands in (8), namely $(2\alpha-1)/(4\alpha)$, only when $\alpha \in (\alpha_0, 1]$, where $\alpha_0 = 0.837...$ is the largest root of the polynomial $4x^3 - 4x + 1$. As α tends to 1 the exponent $\alpha^2 - 1/2$ tends to $1/2$, so the bounds in (8) yield an alternative possible interpolation between Minty's theorem and the nonlinear Hahn–Banach theorem. We do not believe that this interpolation is sharp and it seems more likely that Kalton's conjecture, and correspondingly Question 2 and Question 3, have positive answers. It may very well be the case that the example that we analyse in Section 4.2 is itself an example that yields a positive answer to Question 2, but at present we do not know how to prove this.

1.2. **A conjectural strengthening of Ball's extension theorem.** A longstanding open problem posed by Johnson and Lindenstrauss in [JL84, Problem 1] asks whether or not for every $p \in (1,2)$ and every $m \in \mathbb{N}$ we have

$$e(\ell_p, \ell_2^m) \lesssim_p m^{\frac{1}{p}-\frac{1}{2}}. \tag{9}$$

Here, and in what follows, the notation $A \lesssim_p B$ stands for $A \leqslant C(p)B$, where $C(p) \in (0,\infty)$ is allowed to depend only on p. The notations $A \gtrsim_p B$ and $A \asymp_p B$ are defined analogously.

Johnson and Lindenstrauss were motivated to ask whether (9) holds true by a Lipschitz extension theorem of Marcus and Pisier [MP84], which states that for every $p \in (1,2)$ and $n \in \mathbb{N}$ we have

$$e_n(\ell_p, \ell_2) \lesssim_p (\log n)^{\frac{1}{p}-\frac{1}{2}}. \tag{10}$$

Specifically, the validity of (9) would yield a new proof of the Marcus–Pisier theorem (10) through an application of the Johnson–Lindenstrauss dimensionality reduction lemma (we recalled how such arguments are carried out in the proof of (3) above). Thus, the validity of (9) would yield an illuminating new perspective on the work of Marcus and Pisier [MP84], who proved that (10) holds true via an entirely different argument.

Our goal in this section is to formulate a conjecture that implies both Kalton's conjecture on the validity of (2) and the Johnson–Lindenstrauss conjecture on the validity of (9), in addition to a wealth of yet unknown Lipschitz extension results. In order to do so, we need to quickly recall Ball's work [Bal92] (itself addressing another open question that was posed by Johnson and Lindenstrauss in [JL84]) on the nonlinear version of Maurey's extension theorem [Mau74].

For $p, q \in [1,\infty]$, the Rademacher type p constant and Rademacher cotype q constant of a Banach space $(X, \|\cdot\|_X)$, denoted $T_p(X)$ and $C_q(X)$, respectively, are defined to be the infima of those $T, C \in [1,\infty]$ such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ we have

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \Big\| \sum_{j=1}^n \varepsilon_j x_j \Big\|_X^p \leqslant T^p \sum_{j=1}^n \|x_j\|_X^p \qquad \text{and} \qquad \sum_{j=1}^n \|x_j\|_X^q \leqslant \frac{C^q}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \Big\| \sum_{j=1}^n \varepsilon_j x_j \Big\|_X^q.$$

Maurey's extension theorem [Mau74] asserts that if $(X, \|\cdot\|_X)$ and $(Z, \|\cdot\|_Z)$ are Banach spaces and $E \subseteq X$ is a linear subspace then for every linear operator $U : E \to Z$ there exists a linear operator $V : X \to Z$ that extends U and satisfies the operator norm bound $\|V\|_{X\to Z} \lesssim T_2(X) C_2(Z) \|U\|_{E\to Z}$.

In [Bal92], Ball discovered a powerful nonlinear version of the Maurey extension theorem. In order to do so, he introduced notions of type and cotype for metric spaces that have since proven to be useful in several contexts beyond their original use for the Lipschitz extension problem. For $p \in (0,\infty)$, the Markov type p constant of a metric space (X, d_X), denoted $M_p(X)$, is defined to be the infimum over those $M \in [1,\infty]$ such that for every $n, t \in \mathbb{N}$, every $n \times n$ symmetric stochastic matrix $A=(a_{ij})$, and every $x_1,\ldots,x_n\in X$ we have

$$\sum_{i=1}^n\sum_{j=1}^n (A^t)_{ij} d_X(x_i,x_j)^p \leqslant tM^p \sum_{i=1}^n\sum_{j=1}^n a_{ij} d_X(x_i,x_j)^p.$$

For $q\in(0,\infty)$, the metric Markov cotype q constant of a metric space (X,d_X), denoted $N_p(X)$, is defined to be the infimum over those $N\in(0,\infty]$ such that for every $n,t\in\mathbb{N}$, every $n\times n$ symmetric stochastic matrix $A=(a_{ij})$, and every $x_1,\ldots,x_n\in X$, there exist $y_1,\ldots,y_n\in X$ such that

$$\sum_{i=1}^n d_X(x_i,y_i)^p + t\sum_{i=1}^n\sum_{j=1}^n a_{ij} d_X(y_i,y_j)^p \leqslant N^p \sum_{i=1}^n\sum_{j=1}^n \frac{1}{t}\sum_{s=1}^t (A^s)_{ij} d_X(x_i,x_j)^p.$$

Ball's extension theorem [Bal92] asserts that for every metric space (X,d_X) and every Banach space $(Z,\|\cdot\|_Z)$ we have

$$e(X,Z^{**}) \lesssim M_2(X)N_2(Z).$$

This formulation is not the strongest known version of Ball's theorem, but it suffices for our purposes; see [MN13a] for a more complete discussion.

Remark 5. In the setting of Ball's extension theorem one actually needs to know the Markov type of the *complement* of the subset from which one wishes to extend rather than the Markov type of the entire ambient space. Namely, we actually have

$$e(X,S,Z^{**}) \lesssim M_2(X\smallsetminus S)N_2(Z).$$

This assertion holds true also for the more general version of Ball's extension theorem that was obtained in [MN13a]. The above stronger statement follows effortlessly from an inspection of the proofs in [Bal92, MN13a], though it has never been stated in the literature; we believe that it is worthwhile to record it here for potential future applications.

A beautiful strengthening of Maurey's extension theorem for finite-dimensional targets was discovered by König, Retherford and Tomczak-Jaegermann [KRTJ80], who proved that for every $p\in[1,2]$, $q\in[2,\infty)$ and $m\in\mathbb{N}$, if $(X,\|\cdot\|_X)$ and $(Z,\|\cdot\|_Z)$ are Banach spaces with $\dim(Z)=m$ and $E\subseteq X$ is a linear subspace, then for every linear operator $U:E\to Z$ there exists a linear operator $V:X\to Z$ that extends U and satisfies

$$\|V\|_{X\to Z} \lesssim_{p,q} T_p(X)C_q(Z)m^{\frac{1}{p}-\frac{1}{q}}\|U\|_{E\to Z}. \tag{11}$$

See also the expository article of Pisier [Pis79] for an elegant proof of (11). Important earlier special cases of the above result (other than Maurey's extension theorem itself) can be found (via different proofs) in the works of Lewis [Lew78] and Figiel and Tomczak-Jaegermann [FTJ79].

In light of (11), we ask the following question.

Question 6. Suppose that $1\leqslant p\leqslant q<\infty$ and that (X,d_X) is a metric space with Markov type p, i.e., $M_p(X)<\infty$. Suppose also that $m\in\mathbb{N}$ and that $(Z,\|\cdot\|_Z)$ is an m-dimensional normed space. Is it true that there exists a constant $K=K(M_p(X),N_q(Z),p,q)\in(0,\infty)$, which may depend only on the parameters $M_p(X),N_q(Z),p,q$, such that

$$e(X,Z)\leqslant Km^{\frac{1}{p}-\frac{1}{q}} \ ?$$

Since, by a straightforward application of the triangle inequality, every metric space (X,d_X) has Markov type 1 with $M_1(X)=1$, for every $\alpha\in(0,1]$ the Markov type $1/\alpha$ constant of (X,d_X^α)

equals 1. Also, by [Bal13, MN13a], for every $q \in (1,\infty)$ we have $N_{\max\{q,2\}}(\ell_q) \lesssim \sqrt{q/(q-1)}$. Hence, a positive answer to Question 6 would imply that for every $k \in \mathbb{N}$ we have

$$\frac{1}{\max\{q,2\}} \leqslant \alpha \leqslant 1 \implies e^{\alpha}(\ell_\infty, \ell_q^m) \lesssim_q m^{\alpha - \frac{1}{\max\{q,2\}}}. \tag{12}$$

The case $q = 2$ of (12) is the same as (2), i.e., Kalton's conjecture is a special case of Question 6. By [Bal92, NPSS06], for every $p \in [1,\infty)$ we have $M_{\min\{p,2\}}(\ell_p) \lesssim \sqrt{p}$. This implies formally (directly from the definition of Markov type) that for $\alpha \in (0,1]$ the Markov type $\min\{p,2\}/\alpha$ constant of the metric space $(\ell_p, \|x-y\|_p^{\alpha})$ is at most a constant multiple of $p^{\alpha/2}$. Hence, a positive answer to Question 6 would also imply that for every $p, q \in (1,\infty)$ and every $k \in \mathbb{N}$ we have

$$\frac{\min\{p,2\}}{\max\{q,2\}} \leqslant \alpha \leqslant 1 \implies e^{\alpha}(\ell_p, \ell_q^m) \lesssim_{p,q} m^{\frac{\alpha}{\min\{p,2\}} - \frac{1}{\max\{q,2\}}}. \tag{13}$$

The case $p \in (1,2)$, $q = 2$ and $\alpha = 1$ of (13) is the same as (9), i.e., the above conjecture of Johnson and Lindenstrauss is a special case of Question 6. The validity of (13) would complement the fact that $e^{\alpha}(\ell_p, \ell_q) < \infty$ for every $\alpha \in (0, \min\{p,2\}/\max\{q,2\}]$, as shown in [Nao01, NPSS06].

At present we see several obstacles to adapting the proofs in [Bal92, MN13a] so as to incorporate the finite dimensionality of the target in order to answer Question 6. We therefore leave Question 6 as an intriguing direction for future research, itself part of the Ribe program (see [Bal13, Nao12]).

1.3. **Lipschitz extension between ℓ_p spaces.** Theorem 1 yields the best known lower bound on $\mathfrak{ae}(n)$ as $n \to \infty$, but there are several cases of special interest where the best known lower bound on $e_n(X,Z)$ is $o(\sqrt{\log n})$. Here we recall the best known bounds when $X = \ell_p$ and $Y = \ell_q$ for $p, q \in [1,\infty]$, as a survey of very basic questions on Lipschitz extension that remain open.

The case of Hilbert space-valued functions was famously studied by Johnson and Lindenstrauss in [JL84], answering a question posed by Marcus and Pisier in [MP84]. The bounds on $e_n(\ell_\infty, \ell_2)$ that were obtained in [JL84] are as follows, and they remain the best known bounds to date.

$$\frac{\sqrt{\log n}}{\sqrt{\log\log n}} \lesssim e_n(\ell_\infty, \ell_2) \lesssim \sqrt{\log n}. \tag{14}$$

In the special case of Hilbert space-valued functions defined on finite subsets of ℓ_p for $p \in (1,2)$, the best known bounds are

$$\left(\frac{\log n}{\log\log n}\right)^{\frac{1}{p}-\frac{1}{2}} \lesssim e_n(\ell_p, \ell_2) \lesssim_p (\log n)^{\frac{1}{p}-\frac{1}{2}}. \tag{15}$$

The right hand inequality in (15) is due to [MP84] and the left hand inequality in (15) is due to [JL84], where it is shown to hold for every $p \in [1,2]$.

As shown by Makarychev and Makarychev in [MM10], the parameters $e_n(\ell_\infty, \ell_1)$ and $e_n(\ell_1, \ell_1)$ have a special algorithmic significance. The best known bounds for these quantities are

$$\frac{\sqrt{\log n}}{\sqrt{\log\log n}} \lesssim e_n(\ell_\infty, \ell_1) \lesssim \frac{\log n}{\log\log n}, \tag{16}$$

and

$$\frac{\sqrt{\log n}}{\log\log n} \lesssim e_n(\ell_1, \ell_1) \lesssim \frac{\log n}{\log\log n}. \tag{17}$$

The right hand inequalities in (16) and (17) are a special case of the general upper bound of [LN05]. The left hand inequalities in (16) and (17) appear in [MM10], with the left hand inequality of (17) being based on the work of Figiel, Johnson and Schechtman [FJS88] (an asymptotically weaker lower bound in (17) follows from earlier work of Bourgain [Bou81]).

By Ball's extension theorem [Bal92], combined with the fact [NPSS06] that ℓ_p has Markov type 2 when $p \in [2,\infty)$, if $1 < q \leqslant 2 \leqslant p < \infty$ then $e(\ell_p,\ell_q) \leqslant C(p,q)$ for some $C(p,q) \in (0,\infty)$. Here the asymptotic dependence of $C(p,q)$ as $p \to \infty$ or $q \to 1$ remains unknown. In particular, a famous and longstanding open question of Ball [Bal92] asks whether $e(\ell_2,\ell_1)$ is finite or infinite. The best known bounds on $e_n(\ell_p,\ell_q)$ for the remaining values of $p,q \in [1,\infty]$ are as follows.

$$p,q \in [1,2] \text{ and } q \neq 1 \implies \left(\frac{\log n}{\log\log n}\right)^{\frac{1}{p}-\frac{1}{2}} \lesssim_q e_n(\ell_p,\ell_q) \lesssim_p (\log n)^{\frac{1}{p}}, \tag{18}$$

$$1 \leqslant p \leqslant 2 \implies \frac{(\log n)^{\frac{1}{p}-\frac{1}{2}}}{(\log\log n)^{\frac{1}{p}}} \lesssim e_n(\ell_p,\ell_1) \lesssim_p (\log n)^{\frac{1}{p}}. \tag{19}$$

$$p,q \in (2,\infty) \implies \left(\frac{\log n}{\log\log n}\right)^{\frac{q-2}{q^2}} \lesssim e_n(\ell_p,\ell_q) \lesssim \frac{\log n}{\log\log n}, \tag{20}$$

$$1 \leqslant p \leqslant 2 \leqslant q < \infty \implies \left(\frac{\log n}{\log\log n}\right)^{\max\left\{\frac{1}{p}-\frac{1}{2},\frac{q-p}{q^2}\right\}} \lesssim_q e_n(\ell_p,\ell_q) \lesssim_p (\log n)^{\frac{1}{p}}. \tag{21}$$

The right hand inequalities in (18), (19), (21) are due to [LN05]. More generally, it was shown in [LN05] that for $p \in (1,2]$ we have $e_n(\ell_p,Z) \lesssim_p (\log n)^{1/p}$ for every Banach space Z. It would be interesting to determine the best exponent of $\log n$ in this context of general Banach space targets, as well as when the target Z is allowed to range over some special classes of Banach spaces (e.g., for Banach spaces of cotype 2 it is known [MN13a] that this exponent cannot be smaller than $1/4$).

No upper bound on $e_n(\ell_p,\ell_q)$ that is asymptotically smaller than the general upper bound [LN05] of $O((\log n)/\log\log n)$ is known when $p,q \in (2,\infty)$, and similarly for $e_n(\ell_1,\ell_q)$ for $q \in (2,\infty)$.

Let $(Z,\|\cdot\|_Z)$ be a Banach space that has Rademacher type p for some $p > 1$. By a theorem of Figiel and Tomczak-Jaegermann [FTJ79] there exists $K(Z) \in (0,\infty)$ such that for every $n \in \mathbb{N}$ one can find an n-dimensional subspace Z_n of Z that is 2-isomorphic to ℓ_2^n and there exists a projection from Z onto Z_n of norm at most $K(Z)$. This implies that for any metric space X we have $e_n(X,Z) \gtrsim_Z e_n(X,\ell_2)$. In particular, $e_n(Z,\ell_q) \gtrsim_q e_n(X,\ell_2)$ for every $q > 1$. This implies the validity of the left hand inequality in (18), as a consequence of the left hand inequality of (15). However, this reasoning does not apply when the target space is ℓ_1, i.e., in order to prove the left hand inequality in (19). For this purpose, one argues by adapting the proof of the left hand inequality in (17) that appears in [MM10] (the key tool being [FJS88]); the adaptation of this argument is simple and we omit it (the result itself, however, is far from trivial).

If $(X,\|\cdot\|_X)$ is an infinite dimensional Banach space then, by Dvoretzky's theorem [Dvo61], X contains a 2-isomorphic copy of ℓ_2^n for every $n \in \mathbb{N}$. Consequently, $e_n(X,Z) \gtrsim e_n(\ell_2,Z)$ for every metric space Z. In particular, the left hand inequality in (20) follows from the special case $p = 2$ of the left hand inequality in (21). The latter inequality consists of two lower bounds, one with the exponent $1/p - 1/2$ and the other with the exponent $(q-p)/q^2$. Since, as explained in the previous paragraph, $e_n(\ell_p,\ell_q) \gtrsim_q e_n(\ell_p,\ell_2)$, the lower bound with exponent $1/p - 1/2$ follows from the left hand inequality in (15). The lower bound with exponent $(q-p)/q^2$ is due to [Nao01], with the explicit asymptotic dependence being computed when $p = 2$ in [LN05, Lemma 1.13], and when $p \in (1,2)$ the corresponding bound follows mutatis mutandis by the same argument.

It is worthwhile to point out here an interesting feature of the left hand inequality in (21). Suppose that $p \in [1,2]$. For every n-dimensional subspace Y of ℓ_p there exists a projection from ℓ_p onto Y of norm at most $c_p n^{1/p-1/2}$. This assertion follows from the work of Lewis [Lew78], and it is also a consequence (11). Thus, for every Banach space Z, every linear operator $U : Y \to Z$ can be extended to a linear operator $V : \ell_p \to Z$ with $\|V\|_{\ell_p \to Z} \lesssim_p n^{1/p-1/2}\|U\|_{Y\to Z}$. Consequently, the linearization method of Johnson and Lindenstrauss [JL84] cannot yield a lower bound on $e_n(\ell_p,Z)$ that is at least a constant multiple of $(\log n)^{1/p-1/2}$. However, $(q-p)/q^2>1/p-1/2$ if and only if

$$\frac{3}{2}<p\leqslant 2 \qquad\text{and}\qquad 2+\frac{\left(1-\sqrt{2p-3}\right)^2}{1+\sqrt{2p-3}}<q<2+\frac{\left(1-\sqrt{2p-3}\right)^3}{2(2-p)}. \tag{22}$$

So, if p,q satisfy (22) then the left hand inequality in (21) shows that

$$\lim_{n\to\infty}\frac{e_n(\ell_p,\ell_q)}{(\log n)^{\frac{1}{p}-\frac{1}{2}}}=\infty, \tag{23}$$

i.e., we have a lower bound on $e_n(\ell_p,\ell_q)$ that is asymptotically larger than any bound that can be deduced by reducing the problem to the extension problem for linear operators. We conjecture that the restrictions in (22) are not needed here, i.e., (23) holds true whenever $1\leqslant p\leqslant 2<q<\infty$. It even seems to be unknown whether for every $p\in[1,2]$ there exists a Banach space Z for which

$$\lim_{n\to\infty}\frac{e_n(\ell_p,Z)}{(\log n)^{\frac{1}{p}-\frac{1}{2}}}=\infty. \tag{24}$$

The best-known result towards (24) seems to follow from [MN13a], where it is shown that there exists a Banach space Z for which $e_n(\ell_2,Z)\gtrsim\sqrt[4]{(\log n)/\log\log n}$. Hence, by Dvoretzky's theorem [Dvo61], for every infinite dimensional Banach space X we have $e_n(X,Z)\gtrsim\sqrt[4]{(\log n)/\log\log n}$, implying that (24) holds true if $p\in[1,2]$ satisfies $1/p-1/2<1/4$, i.e., for every $p\in(4/3,2]$.

2. Algorithmic clustering

An algorithmic optimization problem called 0-*Extension*, which we describe below, served as inspiration for our proof of Theorem 1. In this section we shall survey this background so as to clarify the context and explain the "twist" over the existing approaches to 0-Extension that we introduce in order to prove Theorem 1. We stress, however, that our present work does not have new algorithmic implications, and the sole purpose of this section is to explain how this context motivated our approach to Theorem 1. Those who are interested only in the proof of Theorem 1 can skip this section on first reading: the proof itself appears in Section 3 below and is entirely self-contained, with none of the algorithmic background that we describe here being used.

In what follows, the vertices of a combinatorial graph G are denoted V_G and its edges are denoted E_G. The 0-Extension problem is a clustering framework for finite graphs that was introduced by Karzanov [Kar98]. The input of the 0-Extension problem is a graph G with edge weights $w:E_G\to[0,\infty)$, and a subset $T\subseteq V_G$ equipped with a metric $d_T:T\times T\to[0,\infty)$. The subset T is called in the literature the set of *terminals*. The output of the 0-Extension problem is a partition of V_G into $|T|$ subsets $\{C_x\subseteq V_G\}_{x\in T}$ with the requirement that $x\in C_x$ for every $x\in T$. The cost of this partition is defined as follows. Every edge $\{u,v\}\in E_G$ with $|\{u,v\}\cap C_x|=|\{u,v\}\cap C_y|=1$ for some distinct $x,y\in T$ contributes its weight $w(u,v)$ times $d_T(x,y)$ to the total cost, and all other edges (i.e., edges that are entirely within one cluster C_x for some $x\in T$) do not contribute to the total cost. The goal is to find efficiently (in polynomial time) such a partition with minimum cost, or with cost that is guaranteed to be close to the minimum possible cost. Formally, define

$$\mathrm{COST}_{(G,w,d_T)}\left(\{C_x\}_{x\in T}\right)\stackrel{\mathrm{def}}{=}\sum_{\substack{x,y\in T\\ x\neq y}}\ \sum_{\substack{\{u,v\}\in E_G\\ |\{u,v\}\cap C_x|=|\{u,v\}\cap C_y|=1}} w(u,v)d_T(x,y),$$

and

$$\mathrm{OPT}(G,w,d_T)\stackrel{\mathrm{def}}{=}\min_{\substack{\{C_x\}_{x\in T}\text{ partition of }V_G\\ \forall\, x\in T,\quad x\in C_x}}\mathrm{COST}_{(G,w,d_T)}\left(\{C_x\}_{x\in T}\right). \tag{25}$$

The goal is to find efficiently a partition $\{C_x\}_{x\in T}$ of V_G with $x \in C_x$ for every $x \in T$ such that

$$\frac{\mathrm{COST}_{(G,w,d_T)}\left(\{C_x\}_{x\in T}\right)}{\mathrm{OPT}(G,w,d_T)}$$

is guaranteed to be at most a (hopefully small) value $\alpha \in [1,\infty)$.

Obtaining $\alpha = 1$ here would imply that $P = NP$, since in the special when $w(u,v) = 1$ and $d_T(x,y) = 1$ for every $\{u,v\} \in E_G$ and every distinct $x,y \in T$, the quantity $\mathrm{COST}_{(G,w,d_T)}\left(\{C_x\}_{x\in T}\right)$ is nothing more than the total number of edges that are incident to distinct elements of the partition $\{C_x\}_{x\in T}$. Computing the minimum cost in this special case is the MULTIWAY CUT problem, and it was shown in [DJP$^+$94] that if $P \neq NP$ then there exists $\alpha_0 > 1$ such that no polynomial-time algorithm outputs a partition whose cost is guaranteed to be less than α_0 times the minimum possible cost. For the 0-Extension problem in its full generality, it was shown in [KKMR09] that for every $\varepsilon \in (0,1)$ there exists $C(\varepsilon) \in (0,\infty)$ such that the existence of a polynomial-time algorithm for 0-Extension whose approximation factor on instances of size n is $\alpha \leqslant (\log n)^{1/4-\varepsilon}$ would imply that any problem in NP of size n could be solved in time $\exp((\log n)^{C(\varepsilon)})$.

Due to the above evidence for the nonexistence of a constant-factor approximation algorithm for 0-Extension, the literature has focused on the design of approximation algorithms with potentially unbounded approximation factor, based on two competing continuous relaxations of the discrete optimization problem. The first is the *metric relaxation*, which was formulated in [Kar98] and studied extensively in [CKR05]. Given an instance of 0-Extension, i.e., an n-vertex graph G, a metric $d_T : T \times T \to [0,\infty)$ on a subset $T \subseteq V_G$, and edge weights $w : E_G \to [0,\infty)$, define

$$\mathrm{MET}(G,w,d_T) \stackrel{\mathrm{def}}{=} \min_{\substack{d:V_G\times V_G\to[0,\infty)\\ d|_T=d_T\\ d\text{ is a semi-metric}}} \sum_{\{u,v\}\in E_G} w(u,v)d(u,v). \tag{26}$$

The minimization in the right hand side of (26) amounts to a linear program because the constraint that $d : V_G \times V_G \to [0,\infty)$ is a semi-metric that extends d_T corresponds to $O(n^3)$ linear inequalities in the variables $\{d(u,v)\}_{u,v\in V_G}$. Therefore $\mathrm{MET}(G,w,d_T)$ can be computed in polynomial time.

A second relaxation for the 0-Extension problem, called the *earthmover relaxation*, was proposed in [CKNZ05] and studied extensively in [AFH$^+$04, KKMR09]. The idea is similar to (26), except that the semi-metric $d : V_G\times V_G \to [0,\infty)$ is further restricted to a special class of semi-metrics, over which the corresponding minimization can still be cast as a linear program. Given two measures μ,ν on T with the same total mass, i.e., $\mu(T) = \nu(T)$, let $\Pi(\mu,\nu)$ be the set of all couplings of μ and ν. Thus, $\Pi(\mu,\nu)$ consists of all the measures π on $T \times T$ such that $\sum_{y\in T}\pi(x,y) = \mu(x)$ and $\sum_{y\in T}\pi(y,x) = \nu(x)$ for every $x \in T$. The assumption that μ and ν have the same total mass ensures that $\Pi(\mu,\nu) \neq \emptyset$; specifically $(\mu\times\nu)/\mu(T) \in \Pi(\mu,\nu)$. The Wasserstein 1 distance between μ and ν (also known as the earthmover distance between μ and ν) is defined to be

$$W_1^{d_T}(\mu,\nu) \stackrel{\mathrm{def}}{=} \min_{\pi\in\Pi(\mu,\nu)} \sum_{x,y\in T} d_T(x,y)\pi(x,y). \tag{27}$$

Let $\mathcal{P}_T$ denote the set of all probability measures on T and define

$$\mathrm{EMD}(G,w,d_T) \stackrel{\mathrm{def}}{=} \min_{\substack{\{\mu_u\}_{u\in V_G}\subseteq\mathcal{P}_T\\ \forall\, x\in T,\quad \mu_x=\delta_x}} \sum_{\{u,v\}\in E_G} w(u,v)W_1^{d_T}(\mu_u,\mu_v). \tag{28}$$

Here for $x \in T$ the point mass at x is denoted δ_x. The minimization in the right hand side of (28) also amounts to a linear program, with variables corresponding to couplings π_{uv} of the probability measures μ_u and μ_v for each $\{u,v\} \in E_G$, and objective $\sum_{\{u,v\}\in E_G}\sum_{x,y\in T} w(u,v)d_T(x,y)\pi_{uv}(x,y)$;

the requirements of being a probability measure or a coupling are clearly linear constraints. The quantity $\mathrm{EMD}(G, w, d_T)$ can therefore be computed in polynomial time.

If $\{\mu_u\}_{u\in V_G} \subseteq \mathcal{P}_T$ are such that $\mu_x = \delta_x$ for every $x \in T$ then $d(u, v) = W_1^{d_T}(\mu_u, \mu_v)$ is a semi-metric on V_G that extends d_T. Consequently, the minimization in (28) is over a smaller set than the minimization in (26). At the same time, if $\{C_x\}_{x\in T}$ is a partition of V_G with $x \in C_x$ for every $x \in T$, then by defining $\mu_u = \delta_x$ for every $(u, x) \in V_G \times T$ we see that the minimization in (25) is over a smaller set than the minimization in (28). These observations show that every instance of 0-Extension satisfies

$$\mathrm{MET}(G, w, d_T) \leqslant \mathrm{EMD}(G, w, d_T) \leqslant \mathrm{OPT}(G, w, d_T). \tag{29}$$

Building on ideas of [CKR05], it was shown in [FHRT03] that every instance of 0-Extension satisfies $\mathrm{OPT}(G, w, d_T)/\mathrm{MET}(G, w, d_T) \lesssim (\log |T|)/\log\log |T|$. This yields the best known approximation algorithm for the 0-Extension problem, the algorithm being to compute the quantity $\mathrm{MET}(G, w, d_T)$, i.e., to apply the metric relaxation. It follows from (29) that by using the earthmover relaxation, i.e., by computing $\mathrm{EMD}(G, w, d_T)$, one could potentially obtain an even better algorithm. This was realized in [AFH$^+$04] in special cases (e.g. when (T, d_T) is a planar graph), but in [LN03] it was shown that the same approximation guarantees as in [AFH$^+$04] can be obtained by using the metric relaxation. Nevertheless, in [KKMR09] it was shown that in a certain sense the earthmover relaxation of 0-Extension does perform better than the metric relaxation: the earthmover relaxation behaves better than the metric relaxation if one measures the approximation factor in term of the ratio between the largest distance and the smallest nonzero distance in (T, d_T).

On the negative side, it was shown in [CKR05] that there exist instances of 0-Extension with $|T|$ arbitrarily large and $\mathrm{OPT}(G, w, d_T)/\mathrm{MET}(G, w, d_T) \gtrsim \sqrt{\log |T|}$. More recently, in [KKMR09] it was shown that there also exist such instances with $\mathrm{OPT}(G, w, d_T)/\mathrm{EMD}(G, w, d_T) \gtrsim \sqrt{\log |T|}$.

As described above, the earthmover relaxation of 0-Extension is in principle better than the metric relaxation, but at present the general bounds that are available in the literature are to a large extent the same for both relaxations. Nevertheless, our approach to Theorem 1 is based on exploiting the *differences* between the metric relaxation and the earthmover relaxation. The idea is that if an instance of 0-Extension behaves significantly worse for the metric relaxation than for the earthmover relaxation then the terminal metric (T, d_T) can be embedded isometrically into a larger semi-metric (V_G, d) for which $\sum_{\{u,v\}\in E_G} w(u, v)d(u, v)$ is much smaller than $\mathrm{OPT}(G, w, d_T)$, yet whenever one assigns to every vertex $u \in V_G$ a probability measure $\mu_u \in \mathcal{P}_T$ such that $\mu_x = \delta_x$ for every $x \in T$, then $\sum_{\{u,v\}\in E_G} w(u, v)W_1^{d_T}(\mu_u, \mu_v)$ must be close to $\mathrm{OPT}(G, w, d_T)$. This discrepancy between the two relaxations can be used to show that the mapping that assigns to every $x \in T$ the point mass $\delta_x \in \mathcal{P}_T$ cannot be extended to a function from (V_G, d) to $(\mathcal{P}_T, W_1^{d_T})$ that has a small Lipschitz constant. Here the target space $(\mathcal{P}_T, W_1^{d_T})$ is not a normed space, but this idea can be modified so as to yield a poorly extendable function with values in a related normed space, namely the dual of the space of real-valued Lipschitz functions on (T, d_T) that vanish at a fixed point. In the end, the entire argument as presented in Section 3 uses geometric and combinatorial considerations that are self-contained and do not make any reference to the clustering objective $\mathrm{OPT}(G, w, d_T)$. The examples of metric spaces that we use to prove Theorem 1 are modifications of the examples that were considered in [CKR05], which are themselves in the spirit of an example that was used in [JLS86] (and, we use yet another variant of the example of [JLS86] to prove Theorem 4).

3. Proof of Theorem 1

Before passing to the proof of Theorem 1, which appears in Section 3.4 below, we need to recall some (simple) background and to introduce some basic constructions.

3.1. **The r-magnification of a metric space.** Given a metric space (X, d_X) and $r > 0$, for every subset $S \subseteq X$ we shall define a new metric space $X_r(S)$ that we call the r-magnification of (X, d_X) at S. As a set, $X_r(S)$ equals X. The new metric $d_{X_r(S)}$ on X is defined by setting $d_{X_r(S)}(x, x) = 0$ for every $x \in X$, and by defining for every distinct $x, y \in X$,

$$d_{X_r(S)}(x, y) \stackrel{\mathrm{def}}{=} d_X(x, y) + r|\{x, y\} \cap S|. \tag{30}$$

It is immediate to check that $d_{X_r(S)}$ is indeed a metric on S. Note that this construction increases all the positive pairwise distances within S while keeping the pairwise distances within $X \smallsetminus S$ unchanged. This is the reason for our choice of terminology. If G is a connected graph then d_G will always stand for the shortest-path metric on V_G, and for $S \subseteq V_G$ and $r > 0$, the metric space $G_r(S)$ will always be understood to be the r-magnification of (V_G, d_G) at S.

The relevance of the r-magnification of (X, d_X) at $S \subseteq X$ to potentially proving impossibility results for Lipschitz extension is simple to explain. By making the positive pairwise distances within S larger, we make it easier for functions that are defined on S to be Lipschitz, so there are more "potential counterexamples" on $(S, d_{X_r(S)})$ than there were on (S, d_X). At the same time, by keeping the pairwise distances within $X \smallsetminus S$ unchanged, we do not make the Lipschitz condition on $V \smallsetminus S$ any less stringent. However, there is a limitation to this reasoning because as r becomes larger the minimum positive distance in $(S, d_{X_r(S)})$ becomes closer to the diameter of $(S, d_{X_r(S)})$, in which case bounds on $\mathfrak{ae}(\varepsilon)$ become relevant. Indeed, by (30) any distinct $x, y \in S$ satisfy

$$d_{X_r(S)}(x, y) \geqslant 2r + 1 = \frac{2r + 1}{2r + \mathrm{diam}(S, d_X)} \mathrm{diam}(S, d_{X_r(S)}). \tag{31}$$

Since $\mathfrak{ae}(\varepsilon) \lesssim 1/\varepsilon$, it follows that any 1-Lipschitz function from $(S, d_{X_r(S)})$ to any Banach space can be extended to a function defined on $(X, d_{X_r(S)})$ whose Lipschitz constant is at most a constant multiple of $(2r + \mathrm{diam}(S, d_X))/(2r + 1) \leqslant 1 + \mathrm{diam}(S, d_X)/r$. There is therefore a tradeoff that limits how large r could be if one wishes to use the r-magnification for the purpose of obtaining a lower bound on $\mathfrak{ae}(n)$. Below we shall balance these constraints, for an appropriate choice of an initial metric space (X, d_X), so as to yield Theorem 1.

3.2. **The Wasserstein 1 norm.** For a finite set X we denote (as usual) by $\mathbb{R}^X$ the $|X|$-dimensional vector space of all $f : X \to \mathbb{R}$. We also denote by $\mathbb{R}^X_0$ the subspace of $\mathbb{R}^X$ consisting of those $f : X \to \mathbb{R}$ that satisfy $\sum_{x \in X} f(x) = 0$. The standard basis of $\mathbb{R}^X$ will be denoted by $\{e_x\}_{x \in X}$, i.e., $e_x(y) = \mathbf{1}_{\{x=y\}}$ for every $x, y \in X$. For $f \in \mathbb{R}^X$ we shall use the (standard) notation

$$\|f\|_{\ell_1(X)} \stackrel{\mathrm{def}}{=} \sum_{x \in X} |f(x)|.$$

If (X, d_X) is a finite metric space then let $K_{(X,d_X)} \subseteq \mathbb{R}^X_0$ be the the following convex hull.

$$K_{(X,d_X)} \stackrel{\mathrm{def}}{=} \mathrm{conv}\left\{ \frac{e_x - e_y}{d_X(x, y)} : \ x, y \in X \text{ and } x \neq y \right\}. \tag{32}$$

$K_{(X,d_X)}$ is clearly an origin-symmetric convex body, so it is a unit ball of a norm on $\mathbb{R}^X_0$, called the Wasserstein 1 norm induced by X, which we denote by $\|\cdot\|_{W_1(X,d_X)}$. By the Kantorovich–Rubinstein duality theorem (see [Vil03, Thm. 1.14]), denoting as usual $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ for every $f \in \mathbb{R}^X$, and recalling (27), we have

$$\forall\, f \in \mathbb{R}^X_0, \qquad \|f\|_{W_1(X,d_X)} = W_1^{d_X}(f^+, f^-) = \inf_{\pi \in \Pi(f^+, f^-)} \sum_{x,y \in X} d_X(x, y)\pi(x, y). \tag{33}$$

Observe that this makes sense because the assumption $f\in\mathbb{R}^X_0$, i.e., that $\sum_{x\in X} f(x)=0$, implies that the nonnegative functions f^+ and f^- satisfy, $\sum_{x\in X} f^+(x)=\sum_{x\in X} f^-(x)$. So, f^+ and f^- are nonnegative measures with the same total mass.

We record for future use the following very simple lemma.

Lemma 7. *Let (X,d_X) be a finite metric space. Then $\|e_x-e_y\|_{W_1(X,d_X)}=d_X(x,y)$ for every $x,y\in X$ and*

$$\forall\, f\in\mathbb{R}^X_0,\qquad \frac12 \min_{\substack{x,y\in X\\ x\neq y}} d_X(x,y)\|f\|_{\ell_1(X)}\leqslant \|f\|_{W_1(X,d_X)}\leqslant \frac12 \mathrm{diam}(X)\|f\|_{\ell_1(X)}. \tag{34}$$

In particular, for every $r>0$ and $S\subseteq X$ we have

$$\forall\, f\in\mathbb{R}^S_0,\qquad r\|f\|_{\ell_1(S)}\leqslant \|f\|_{W_1(S,d_{X_r(S)})}\leqslant \left(r+\frac{\mathrm{diam}(X)}{2}\right)\|f\|_{\ell_1(S)}. \tag{35}$$

Proof. The fact that $\|e_x-e_y\|_{W_1(X,d_X)}=d_X(x,y)$ for every distinct $x,y\in X$ is immediate from (33), and also directly from (32), which says that $(e_x-e_y)/d_X(x,y)$ has unit norm in $W_1(X,d_X)$. To prove (34), denote by m the minimum nonzero value that d_X attains. For distinct $x,y\in X$,

$$\max_{\substack{x,y\in X\\ x\neq y}}\left\|\frac{e_x-e_y}{d_X(x,y)}\right\|_{\ell_1(X)}\leqslant \max_{\substack{x,y\in X\\ x\neq y}}\frac{\|e_x-e_y\|_{\ell_1(X)}}{m}=\frac{2}{m}.$$

This means that for every distinct $x,y\in X$ we have $(e_x-e_y)/d_X(x,y)\in(2/m)B_{\ell_1(X)}$, where $B_{\ell_1(X)}$ is the unit ball of $\ell_1(X)$. By (32) we therefore have $K_{(X,d_X)}\subseteq(2/m)B_{\ell_1(X)}$, which is the same as the first inequality in (34). The second inequality in (34) follows directly from (33), noting that the sum in the right hand side of (33) is at most $\mathrm{diam}(X)\sum_{x\in X}f^+(x)=\mathrm{diam}(X)\|f\|_{\ell_1(X)}/2$. The estimate (35) is a special case of (34) because by (30) the minimum nonzero distance within S of the r-magnification of X at S is at least $2r$, and the diameter of $X_r(S)$ is at most $2r+\mathrm{diam}(X)$. □

3.3. **Properties of expanders.** In the proof of Theorem 1 we shall use several properties of graphs in general, and expander graphs in particular. Here we collect these facts for ease of later reference. Fix two integers $n,d\geqslant 3$ and let G be a connected n-vertex d-regular graph. Recall that the shortest-path metric that G induces on V_G is denoted d_G. The following estimate is standard.

$$\forall\,\emptyset\neq S\subseteq V_G,\qquad \frac{1}{|S|^2}\sum_{x,y\in S} d_G(x,y)\geqslant \frac{\log|S|}{4\log d}. \tag{36}$$

To quickly verify the validity of (36), observe that since the smallest nonzero distance in G is at least 1, the average that appears in the left hand side of (36) is always at least $|S|(|S|-1)/|S|^2=1-1/|S|$. One checks directly that $1-1/a\geqslant(\log a)/(4\log 3)$ for every $a\in\{1,\ldots,15\}$, so in order to establish (36) we may assume that $|S|\geqslant 16$. Denote $k=1+\lfloor\log_d(|S|/4)\rfloor$ and observe that for every $x\in V_G$ the number of $y\in V_G$ with $d_G(x,y)\leqslant k-1$ is at most $1+d+\ldots+d^{k-1}\leqslant 2d^{k-1}\leqslant|S|/2$. The average that appears in the left hand side of (36) is therefore at least $k/2$, and it remains to note that $k\geqslant\log_d(|S|/4)\geqslant(\log|S|)/(2\log d)$, since $|S|\geqslant 16$.

Fixing $S\subseteq V_G$ and $r>0$, we shall need later the following straightforward evaluation of the average length of edges of G in the r-magnification of (G,d_G) at S.

$$\frac{1}{|E_G|}\sum_{\{x,y\}\in E_G} d_{G_r(S)}(x,y)=1+\frac{2r|S|}{n}. \tag{37}$$

Indeed, let E_1 be those edges in E_G that are contained in S and let E_2 be those edges in E_G that contain exactly one element of S. Because G is d-regular, we have $2|E_1|+|E_2|=d|S|$.

Recalling (30), for every $\{x,y\}\in E_1$ we have $d_{G_r(S)}(x,y)=2r+1$, for every $\{x,y\}\in E_2$ we have $d_{G_r(S)}(x,y)=r+1$, and for every $\{x,y\}\in E_G\setminus(E_1\cup E_2)$ we have $d_{G_r(S)}(x,y)=1$. Consequently,

$$\frac{1}{|E_G|}\sum_{(x,y)\in E_G} d_{G_r(S)}(x,y)=\frac{(|E_G|-|E_1|-|E_2|)+(2r+1)|E_1|+(r+1)|E_2|}{|E_G|}$$

$$=1+\frac{r(2|E_1|+|E_2|)}{dn/2}=1+\frac{2r|S|}{n}.$$

Given two disjoint subsets $S,T\subseteq V_G$, denote the number of edges in E_G that intersect both S and T by $E_G(S,T)$. The edge-expansion of G, denoted $\phi(G)$, is the largest $\phi\in[0,\infty)$ such that

$$\forall\, S\subseteq V_G,\qquad E_G(S,V_G\setminus S)\geqslant \phi\frac{|S|(n-|S|)}{n^2}|E_G|. \tag{38}$$

It is well known that (38) is equivalent to the assertion that every $h:V_G\to\ell_1$ satisfies.

$$\frac{\phi}{n^2}\sum_{x,y\in V_G}\|h(x)-h(y)\|_1\leqslant \frac{1}{|E_G|}\sum_{\{x,y\}\in E_G}\|h(x)-h(y)\|_1. \tag{39}$$

The equivalence of (38) and (39) is a standard application of the cut-cone decomposition of subsets of ℓ_1; see e.g. inequality (4) in [Mat97] or [NRS05, Fact 2.1]. The following simple combination of (39) and Lemma 7 will be used later.

Lemma 8. *Fix $n\in\mathbb{N}$ and $\phi\in(0,1]$. Suppose that G be an n-vertex graph with $\phi(G)\geqslant\phi$. For every $\emptyset\neq S\subseteq V_G$ and $r>0$, every $F:V_G\to\mathbb{R}_0^S$ satisfies*

$$\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\leqslant\frac{2r+\mathrm{diam}(S,d_G)}{(2r+1)\phi}\cdot\frac{1}{|E_G|}\sum_{\{x,y\}\in E_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}.$$

Proof. By (39) we have

$$\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{\ell_1(S)}\leqslant\frac{1}{\phi|E_G|}\sum_{\{x,y\}\in E_G}\|F(x)-F(y)\|_{\ell_1(S)} \tag{40}$$

Recalling (30), we have $\mathrm{diam}(S,d_{G_r(S)})=2r+\mathrm{diam}(S,d_G)$ and the smallest positive distance in $(S,d_{G_r(S)})$ equals $2r+1$. So, by Lemma 7, every $x,y\in V_G$ satisfy

$$\frac{2r+1}{2}\|F(x)-F(y)\|_{\ell_1(S)}\leqslant\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\leqslant\frac{2r+\mathrm{diam}(S,d_G)}{2}\|F(x)-F(y)\|_{\ell_1(S)}.$$

Lemma 8 now follows by substituting these estimates for pairwise distances into (40). □

We end by recording for future use the following direct consequence of Menger's theorem [Men27].

Lemma 9. *Let G be an n-vertex graph and $A,B\subseteq V_G$ satisfy $A\cap B=\emptyset$. Fix $\phi\in(0,\infty)$ and suppose that $\phi(G)\geqslant\phi$. Then the number of edge-disjoint paths joining A and B is at least $\phi\min\{|A|,|B|\}|E_G|/(2n)$. In other words, there exists an integer $m\geqslant\phi\min\{|A|,|B|\}|E_G|/(2n)$, $k_1,\ldots,k_m\in\mathbb{N}$ and $\{u_{i,1},u_{i,2},\ldots,u_{i,k_i}\}_{i=1}^m\subseteq V_G$ such that $\{u_{i,1}\}_{i=1}^m\subseteq A$, $\{u_{i,k_i}\}_{i=1}^m\subseteq B$. Moreover, $\{u_{i,s},u_{i,s+1}\}\in E_G$ for every $i\in\{1,\ldots,m\}$ and $s\in\{1,\ldots,k_i-1\}$, and if for some $i,j\in\{1,\ldots,m\}$, $s\in\{1,\ldots,k_i-1\}$ and $t\in\{1,\ldots,k_j-1\}$ we have $\{u_{i,s},u_{i,s+1}\}=\{u_{j,t},u_{j,t+1}\}$ then necessarily $i=j$ and $s=t$.*

Proof. Let m be the maximal number of edge-disjoint paths joining A and B. By the classical Menger theorem [Men27] (see also e.g. [Die10, Chapter 3]) there exists a subset of edges $E^*\subseteq E_G$ with $|E^*|=m$ such that every path in G that joins a vertex in A with a vertex in B contains an edge from E^*. Since in the graph $G^*=(V_G,E_G\setminus E^*)$ there is no path that joins an element of A

with an element of B, if we let $C \subseteq V_G$ be the union of all the connected components of G^* that contain an element of A then $C \supseteq A$ and $C \cap B = \emptyset$. Since C is a union of connected components of G^*, all the edges of E_G joining C and $V_G \smallsetminus C$ belong to E^*. Hence, $E_G(C, V_G \smallsetminus C) \leqslant |E^*| = m$. It remains to note that by (38) we have

$$m \geqslant E_G(C, V_G \smallsetminus C) \geqslant \phi \frac{\max\{|C|, n-|C|\} \cdot \min\{|C|, n-|C|\}}{n^2} |E_G| \geqslant \frac{\phi \min\{|A|, |B|\} |E_G|}{2n}, \quad (41)$$

where the last step of (41) holds true because clearly $\max\{|C|, n-|C|\} \geqslant n/2$, and since $C \supseteq A$ and $V_G \smallsetminus C \supseteq B$ we have $\min\{|C|, n-|C|\} \geqslant \min\{|A|, |B|\}$. $\square$

3.4. **A Wasserstein-valued poorly extendable function.** Fix $d, n \in \mathbb{N}$ and $\phi \in (0,1)$. Throughout this section, G will be fixed to be an n-vertex d-regular graph with $\phi(G) \geqslant \phi$. We shall also fix $\emptyset \neq S \subseteq V_G$ and $r > 0$. Define a mapping

$$f : \big(S, d_{G_r(S)}\big) \to \Big(\mathbb{R}_0^S, \|\cdot\|_{W_1(S, d_{G_r(S)})}\Big)$$

by

$$\forall\, x \in S, \qquad f(x) \stackrel{\mathrm{def}}{=} e_x - \frac{1}{|S|} \sum_{z \in S} e_z. \quad (42)$$

Thus f is an isometry (recall Lemma 7). Suppose that $F : V_G \to \mathbb{R}_0^S$ extends f and for some $L \in (0, \infty)$ we have

$$\forall\, x, y \in V_G, \qquad \|F(x) - F(y)\|_{W_1(S, d_{G_r(S)})} \leqslant L d_{G_r(S)}(x, y). \quad (43)$$

Our goal is to bound L from below.

For every $x \in V_G$ and $s \in (0, \infty)$ define $\mathcal{B}_s(x) \subseteq V_G$ to be the inverse image under F of the $W_1(S, d_{G_r(S)})$-ball of radius s centered at $F(x)$, i.e.,

$$\mathcal{B}_s(x) \stackrel{\mathrm{def}}{=} \Big\{ y \in V_G : \ \|F(x) - F(y)\|_{W_1(S, d_{G_r(S)})} \leqslant s \Big\}.$$

By Lemma 9 there exists an integer

$$m \geqslant \frac{\phi d}{4} \min\left\{ |S \smallsetminus \mathcal{B}_s(x)|, |\mathcal{B}_s(x)| \right\}, \quad (44)$$

and m edge-disjoint paths joining $S \smallsetminus \mathcal{B}_s(x)$ and $\mathcal{B}_s(x)$. This means that we can find $k_1, \ldots, k_m \in \mathbb{N}$ and $\{z_{j,1}, z_{j,2}, \ldots, z_{j,k_j}\}_{j=1}^m \subseteq V_G$ such that $\{z_{j,1}\}_{j=1}^m \subseteq S \smallsetminus \mathcal{B}_s(x)$, $\{z_{j,k_j}\}_{j=1}^m \subseteq \mathcal{B}_s(x)$, and such that $\{\{z_{j,i}, z_{j,i+1}\} : j \in \{1, \ldots, m\} \ \wedge \ i \in \{1, \ldots, k_j - 1\}\}$ are distinct edges in E_G.

Let $J \subseteq \{1, \ldots, m\}$ be such that $\{z_{j,1}\}_{j \in J}$ are distinct and $\{z_{j,1}\}_{j \in J} = \{z_{i,1}\}_{i=1}^m$. For every $j \in J$ denote the number of those $i \in \{1, \ldots, m\}$ for which $z_{j,1} = z_{i,1}$ by d_j. Since $\{\{z_{i,1}, z_{i,2}\}\}_{i=1}^m$ are distinct edges in E_G, and G is d-regular, $\max_{j \in J} d_j \leqslant d$. Because $\sum_{j \in J} d_j = m$, it follows that

$$|J| \geqslant \frac{m}{d} \stackrel{(44)}{\geqslant} \frac{\phi}{4} \min\left\{ |S \smallsetminus \mathcal{B}_s(x)|, |\mathcal{B}_s(x)| \right\}. \quad (45)$$

The following lemma provides an upper bound on $|J|$ that we will later contrast with (45).

Lemma 10. *Under the above notation and assumptions we have*

$$|J| \leqslant \max\left\{ d^{16(s-r)}, \frac{16 n L d \log d}{\log n} \left(1 + \frac{2r|S|}{n}\right) \right\}. \quad (46)$$

Proof. If $|J| \leqslant d^{16(s-r)}$ then we are done, so we may assume that $|J| > d^{16(s-r)}$, or equivalently

$$s - r < \frac{\log |J|}{16 \log d}. \tag{47}$$

Observe that because $\{z_{j,1}\}_{j \in J} \subseteq S$ and $F|_S = f$ is an isometry on $(S, d_{G_r(S)})$,

$$\forall\, i, j \in J, \qquad \|F(z_{i,1}) - F(z_{j,1})\|_{W_1(S, d_{G_r(S)})} = d_{G_r(S)}(z_{i,1}, z_{j,1}) \overset{(30)}{=} 2r + d_G(z_{i,1}, z_{j,1}). \tag{48}$$

Hence,

$$\begin{aligned}
&\sum_{j \in J} \left\|F(z_{j,1}) - F(z_{j,k_j})\right\|_{W_1(S, d_{G_r(S)})} && (49)\\
&= \frac{1}{2|J|} \sum_{i,j \in J} \Big(\|F(z_{i,1}) - F(z_{i,k_i})\|_{W_1(S, d_{G_r(S)})} + \left\|F(z_{j,1}) - F(z_{j,k_j})\right\|_{W_1(S, d_{G_r(S)})} \Big)\\
&\geqslant \frac{1}{2|J|} \sum_{i,j \in J} \Big(\|F(z_{i,1}) - F(z_{j,1})\|_{W_1(S, d_{G_r(S)})} - \left\|F(z_{i,k_i}) - F(z_{j,k_j})\right\|_{W_1(S, d_{G_r(S)})} \Big)\\
&\geqslant \frac{1}{2|J|} \sum_{i,j \in J} d_G(z_{i,1}, z_{j,1}) - (s-r)|J| && (50)\\
&\geqslant \frac{|J| \log |J|}{8 \log d} - (s-r)|J| > \frac{|J| \log |J|}{16 \log d}, && (51)
\end{aligned}$$

where in (50) we used (48) and the fact that, because $\{z_{j,k_j}\}_{j \in J} \subseteq \mathcal{B}_s(x)$, it follows from the definition of $\mathcal{B}_s(x)$ that for every $i, j \in J$ we have

$$\|F(z_{i,k_i}) - F(z_{j,k_j})\|_{W_1(S, d_{G_r(S)})} \leqslant \|F(z_{i,k_i}) - F(x)\|_{W_1(S, d_{G_r(S)})} + \|F(x) - F(z_{j,k_j})\|_{W_1(S, d_{G_r(S)})} \leqslant 2s.$$

The penultimate inequality in (51) is an application of (36), and the final inequality in (51) uses (47).

The quantity in (49) can be bounded from above using the Lipschitz condition (43) and the triangle inequality as follows.

$$\sum_{j \in J} \left\|F(z_{j,1}) - F(z_{j,k_j})\right\|_{W_1(S, d_{G_r(S)})} \leqslant L \sum_{j \in J} d_{G_r(S)}(z_{j,1}, z_{j,k_j}) \leqslant L \sum_{j \in J} \sum_{i=1}^{k_j - 1} d_{G_r(S)}(z_{j,i}, z_{j,i+1}). \tag{52}$$

Since $\{\{z_{j,i}, z_{j,i+1}\} : j \in J \ \wedge\ i \in \{1, \ldots, k_j - 1\}\}$ are distinct edges in E_G,

$$\sum_{j \in J} \sum_{i=1}^{k_j - 1} d_{G_r(S)}(z_{j,i}, z_{j,i+1}) \leqslant \sum_{\{u,v\} \in E_G} d_{G_r(S)}(u, v) \overset{(37)}{=} \frac{nd}{2} \left(1 + \frac{2r|S|}{n}\right). \tag{53}$$

A substitution of (53) into (52), and contrasting the resulting estimate with (51), yields

$$\frac{Lnd}{2} \left(1 + \frac{2r|S|}{n}\right) \geqslant \frac{|J| \log |J|}{16 \log d}. \tag{54}$$

It is elementary to check that if $a \in [1, \infty)$ and $b \in (1, \infty)$ satisfy $a \log a \leqslant b$ then $a \leqslant 2b/\log b$. By applying this with $a = |J|$ and $b = 8Lnd \log d(1 + 2r|S|/n) \geqslant n$, we see that (54) implies that

$$|J| \leqslant \frac{16nLd \log d}{\log n} \left(1 + \frac{2r|S|}{n}\right),$$

thus completing the proof of (46). □

Corollary 11. *Suppose that the following conditions hold true.*

$$d^{16(s-r)} \leqslant \frac{\phi|S|}{8} \qquad \text{and} \qquad L \leqslant \frac{\phi|S|\log n}{128\left(1+\frac{2r|S|}{n}\right)nd\log d}. \tag{55}$$

Then

$$\max_{x\in V_G} |\mathcal{B}_s(x)| < \frac{|S|}{2}. \tag{56}$$

Proof. Fix $x \in V_G$. If $\mathcal{B}_s(x)\cap S \neq \emptyset$ then it follows from (36) that there exist $y,z \in \mathcal{B}_s(x)\cap S$ with

$$d_G(y,z) \geqslant \frac{\log|\mathcal{B}_s(x)\cap S|}{4\log d}. \tag{57}$$

At the same time, since $y,z\in S$ and $F|_S = f$ is an isometry on $(S, d_{G_r(S)})$, we know that

$$\begin{aligned} d_G(y,z) + 2r \overset{(30)}{=} d_{G_r(S)}(y,z) &= \|F(y)-F(z)\|_{W_1(S,d_{G_r(S)})} \\ &\leqslant \|F(y)-F(x)\|_{W_1(S,d_{G_r(S)})} + \|F(x)-F(z)\|_{W_1(S,d_{G_r(S)})} \leqslant 2s, \end{aligned} \tag{58}$$

where in the last step of (58) we used the fact that $y,z\in\mathcal{B}_s(x)$. Contrasting (57) and (58) yields

$$|\mathcal{B}_s(x)\cap S| \leqslant d^{8(s-r)} \leqslant \sqrt{\frac{\phi|S|}{8}} \leqslant \frac{2|S|}{5}, \tag{59}$$

where we used the first assumption in (55) (and that $\phi \leqslant 1 \leqslant |S|$). It follows from (59) that $|S\smallsetminus \mathcal{B}_s(x)| \geqslant 3|S|/5$. By combining (45) and (46) with this lower bound on $|S\smallsetminus\mathcal{B}_s(x)|$ we see that

$$\min\left\{\frac{3|S|}{5}, |\mathcal{B}_s(x)|\right\} < \max\left\{\frac{4d^{16(s-r)}}{\phi}, \frac{64nLd\log d}{\phi\log n}\left(1+\frac{2r|S|}{n}\right)\right\}. \tag{60}$$

But, the two assumptions in (55) imply that $3|S|/5$ is greater than the right hand side of (60), so

$$|\mathcal{B}_s(x)| \leqslant \max\left\{\frac{4d^{16(s-r)}}{\phi}, \frac{64nLd\log d}{\phi\log n}\left(1+\frac{2r|S|}{n}\right)\right\} \overset{(55)}{\leqslant} \frac{|S|}{2}.$$

□

Corollary 12. *If the conditions in* (55) *are satisfied then*

$$L \geqslant \frac{\phi s}{2\left(1+\frac{\operatorname{diam}(G,d_G)}{2r}\right)\left(1+\frac{2r|S|}{n}\right)}. \tag{61}$$

Proof. For every $x\in V_G$ and $y \in V_G\smallsetminus\mathcal{B}_s(x)$ we have $\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} > s$. Hence,

$$\begin{aligned} \frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} &\geqslant \frac{1}{n^2}\sum_{x\in V_G}\sum_{y\in V_G\smallsetminus\mathcal{B}_s(x)}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} \\ &\geqslant \frac{s}{n^2}\sum_{x\in V_G}(n-|\mathcal{B}_s(x)|) \geqslant s\left(1-\frac{\max_{x\in V_G}|\mathcal{B}_s(x)|}{n}\right) \overset{(56)}{\geqslant} \frac{s}{2}. \end{aligned} \tag{62}$$

At the same time, by Lemma 8 we have

$$\begin{aligned} \frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} &\leqslant \frac{2r+\operatorname{diam}(S,d_G)}{(2r+1)\phi|E_G|}\sum_{\{x,y\}\in E_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} \\ &\overset{(43)}{\leqslant} \frac{L\left(1+\frac{\operatorname{diam}(G,d_G)}{2r}\right)}{\phi|E_G|}\sum_{\{x,y\}\in E_G} d_{G_r(S)}(x,y) \overset{(37)}{=} \frac{L\left(1+\frac{\operatorname{diam}(G,d_G)}{2r}\right)\left(1+\frac{2r|S|}{n}\right)}{\phi}. \end{aligned} \tag{63}$$

The desired estimate (61) follows by contrasting (62) with (63). $\square$

Theorem 13. *Continuing with the same notation as above, if $0 < r \leqslant \mathrm{diam}(G, d_G)$ then*

$$L \gtrsim \frac{\phi}{1+\frac{r|S|}{n}} \cdot \min\left\{\frac{|S|\log n}{nd\log d}, \frac{16r^2\log d + r\log(\phi|S|/8)}{\mathrm{diam}(G,d_G)\log d}\right\}. \tag{64}$$

Proof. If $16r\log d + \log(\phi|S|/8) \leqslant 0$ then (64) is vacuous. Supposing that $16r\log d+\log(\phi|S|/8) > 0$, choose $s = r + (\log(\phi|S|/8))/(16\log d)$, so $s > 0$ and $d^{16(s-r)} = \phi|S|/8$. The first inequality of (55) is therefore satisfied, so either the second inequality in (55) fails, corresponding to a lower bound on L, or by Corollary 12 the lower bound on L in (61) is satisfied. This simplifies to give (64). $\square$

For a wide range of the parameters (i.e., $n, d, \phi, \mathrm{diam}(G, d_G), |S|, r$), Theorem 13 yields a nontrivial lower bound on the Lipschitz constant of any mapping F that extends the 1-Lipschitz function f given in (42). Rather than treating the general case, we shall now proceed to compute an optimal setting of the parameters in (64) when G has the property that $\mathrm{diam}(G, d_G) \asymp (\log n)/\log d$ and $\phi(G) \asymp 1$. Such graphs exist for any degree $d \geqslant 3$ and arbitrarily large $n \in \mathbb{N}$, and in fact this holds true with probability tending to 1 as $n \to \infty$ when G is chosen uniformly at random from the finite set of all n-vertex d-regular graphs; for the diameter of random regular graphs see e.g. [BFdlV82] and for the expansion of random regular graphs see e.g. [HLW06, Theorem 4.16].

So, if $\phi \asymp 1$ and $\mathrm{diam}(G, d_G) \asymp (\log n)/\log d$, and continuing with the assumption of Theorem 13 that $0 < r \leqslant \mathrm{diam}(G, d_G)$, the lower bound (64) becomes

$$L \gtrsim \frac{1}{1+\frac{r|S|}{n}} \cdot \min\left\{\frac{|S|\log n}{nd\log d}, \frac{r\left(r\log d + \log|S|\right)}{\log n}\right\}.$$

The optimal choice (up to constant factors) is to take $S \subseteq V_G$ with $|S| = \lfloor n\sqrt{d\log d}/\sqrt{\log n}\rfloor$ (this is allowed provided n is large enough so as to ensure that $|S| \leqslant n$, specifically it suffices to assume that $n \geqslant d^d$) and $r \asymp \sqrt{\log n}/\sqrt{d\log d}$. With these choices we see that $L \gtrsim \sqrt{\log n}/\sqrt{d\log d}$. Thus,

$$e_{\lfloor n\sqrt{d\log d}/\sqrt{\log n}\rfloor}\left(G_r(S), W_1\left(S, d_{G_r(S)}\right)\right) \gtrsim \frac{\sqrt{\log n}}{\sqrt{d\log d}}. \tag{65}$$

For fixed d (say, $d = 4$), the estimate (65) implies Theorem 1. $\square$

Remark 14. The above reasoning also provides a new example showing that $\mathfrak{ae}(\varepsilon) \gtrsim 1/\varepsilon$. Recalling (31), every distinct $x, y \in S$ satisfy $d_{G_r(S)}(x,y) \geqslant \varepsilon\mathrm{diam}(S, d_{G_r(S)})$ where $\varepsilon \asymp r/\mathrm{diam}(G, d_G)$. When $\mathrm{diam}(G, d_G) \asymp (\log n)/\log d$ and $r \asymp \sqrt{\log n}/\sqrt{d\log d}$, this becomes $\varepsilon \asymp \sqrt{\log d}/\sqrt{d\log n}$. The lower bound $L \gtrsim \sqrt{\log n}/\sqrt{d\log d}$ therefore becomes $L \gtrsim 1/(\varepsilon d)$.

4. Proof of Theorem 4

In Section 1.1 we defined the parameters $e^\alpha(X, Z)$ and $e^\alpha_n(X, Z)$, for every $\alpha \in (0, 1]$, $n \in \mathbb{N}$ and every two metric spaces (X, d_X) and (Z, d_Z). Below it will be convenient to also use the analogous notation $e^\alpha(X, S, Z)$ for $S \subseteq X$, i.e.,

$$e^\alpha(X, S, Z) \stackrel{\mathrm{def}}{=} e\big((X, d_X^\alpha), S, (Z, d_Z)\big).$$

Theorem 4 asserts the validity of the following four lower bounds for every $m, n \in \mathbb{N}$.

$$e^\alpha(\ell_\infty, \ell_2^m) \gtrsim m^{\frac{2\alpha-1}{4\alpha}} \qquad \text{and} \qquad e^\alpha_n(\ell_\infty, \ell_2) \gtrsim (\log n)^{\frac{2\alpha-1}{4\alpha}}, \tag{66}$$

and

$$e^\alpha(\ell_\infty, \ell_2^m) \gtrsim m^{\alpha^2-\frac12} \qquad \text{and} \qquad e^\alpha_n(\ell_\infty, \ell_2) \gtrsim \left(\frac{\log n}{\log\log n}\right)^{\alpha^2-\frac12}. \tag{67}$$

The proof of (66) appears in Section 4.1 below, and the proof of (67) appears in Section 4.2 below.

4.1. **Twisted unions of hypercubes.** For $m \in \mathbb{N}$ we identify the discrete hypercube $\{0,1\}^m$ with $\mathbb{F}_2^m$, the vector space of dimension m over the field of size two $\mathbb{F}_2$. We let $e_1, \ldots, e_m$ denote the standard basis of $\mathbb{F}_2^m$ and write $e = e_1 + \ldots + e_m$.

Lemma 15. *Suppose that $\alpha \in (\frac{1}{2}, 1]$ and $r, s \in (0,\infty)$ satisfy*

$$(2\alpha)^{2\alpha} s (2r)^{2\alpha-1} \geqslant \left((2\alpha)^{\frac{2\alpha}{2\alpha-1}} - 1\right)^{2\alpha-1}. \tag{68}$$

Define for every $(x,i),(y,j) \in \mathbb{F}_2^m \times \mathbb{F}_2$,

$$d((x,i),(y,j)) \stackrel{\mathrm{def}}{=} \begin{cases} \min\left\{s\|x-y\|_1, 2r + \|x-y\|_1^{\frac{1}{2\alpha}}\right\} & \text{if } i=j=1,\\ \|x-y\|_1^{\frac{1}{2\alpha}} & \text{if } i=j=0,\\ r + \min\left\{s\|x-y\|_1, \|x-y\|_1^{\frac{1}{2\alpha}}\right\} & \text{if } i \neq j.\end{cases} \tag{69}$$

Then d is a metric on $\mathbb{F}_2^m \times \mathbb{F}_2$.

Proof. Recall that if $\omega : [0,\infty) \to [0,\infty)$ is concave, nondecreasing, and satisfies $\omega(0) = 0$, then $\omega \circ d_X$ is a metric on X for every metric space (X, d_X). An application of this observation to the mapping $t \mapsto t^{1/(2\alpha)}$, which is concave because $\alpha > \frac{1}{2}$, shows that d is a metric on $\mathbb{F}_2^m \times \{0\}$. Also, since both of the mappings $t \mapsto st$ and $t \mapsto 2r + t^{1/(2\alpha)}$ are concave and increasing, the mapping $t \mapsto \min\{st, 2r + t^{1/(2\alpha)}\}$ is concave and increasing. Consequently, d is a metric on $\mathbb{F}_2^m \times \{1\}$.

It therefore remains to show that if $i,j,k \in \{0,1\}$ satisfy $\{i,j,k\} = \{0,1\}$ and $x,y,z \in \mathbb{F}_2^m$ then $d((x,i),(y,j)) \leqslant d((x,i),(z,k)) + d((z,k),(y,j))$. By translation and permutation of the coordinates we may assume that $x = 0$ and $y = e_1 + \ldots + e_a$ for some $a \in \{1,\ldots,m\}$. In this case, note that if z has a nonzero entry outside $\{1,\ldots,a\}$ then both $\|z\|_1$ and $\|y-z\|_1$ will decrease if we change that entry of z from 1 to 0. By the definition of d in (69), this shows that $d((x,i),(z,k)) + d((z,k),(y,j))$ will decrease if we set all the entries of z that are outside $\{1,\ldots,a\}$ to be equal 0. It therefore suffices to prove that for every $a \in \{1,\ldots,m\}$ and $b \in \{1,\ldots,a-1\}$ we have

$$d((0,i),(e_1+\ldots+e_a,j)) \leqslant d((0,i),(e_1+\ldots+e_b,k)) + d((e_1+\ldots+e_b,k),(e_1+\ldots+e_a,j)). \tag{70}$$

We shall establish the validity of (70) through the following case analysis.

Case 1: $i = j = 0$. In this case necessarily $k = 1$, so by (69) the desired inequality (70) becomes

$$a^{\frac{1}{2\alpha}} \leqslant 2r + \min\left\{sb, b^{\frac{1}{2\alpha}}\right\} + \min\left\{s(a-b), (a-b)^{\frac{1}{2\alpha}}\right\}. \tag{71}$$

The mapping $\omega : [0,\infty) \to [0,\infty)$ given by $\omega(t) \stackrel{\mathrm{def}}{=} \min\{st, t^{1/(2\alpha)}\}$ is concave, increasing and satisfies $\omega(0) = 0$. Consequently, $\omega(b) + \omega(a-b) \geqslant \omega(a)$. So, in order to prove (71) it suffices to show that $a^{1/(2\alpha)} \leqslant 2r + \omega(a) = 2r + \min\{sa, a^{1/(2\alpha)}\}$. We therefore need to show that $a^{1/(2\alpha)} \leqslant 2r + sa$. Define $\phi : [0,\infty) \to [0,\infty)$ by $\phi(t) = 2r + st - t^{1/(2\alpha)}$. The minimum of ϕ on $[0,\infty)$ is attained at $t_{\min} = 1/(2\alpha s)^{2\alpha/(2\alpha-1)}$. Hence, by our assumption (68) we have

$$\phi(a) \geqslant \phi(t_{\min}) = 2r + \frac{s}{(2\alpha s)^{\frac{2\alpha}{2\alpha-1}}} - \frac{1}{(2\alpha s)^{\frac{1}{2\alpha-1}}} = 2r - \frac{(2\alpha)^{\frac{2\alpha}{2\alpha-1}} - 1}{((2\alpha)^{2\alpha} s)^{\frac{1}{2\alpha-1}}} \stackrel{(68)}{\geqslant} 0.$$

Case 2: $i = j = 1$. In this case necessarily $k = 0$, so by (69) the desired inequality (70) becomes

$$\min\left\{sa, 2r + a^{\frac{1}{2\alpha}}\right\} \leqslant 2r + \min\left\{sb, b^{\frac{1}{2\alpha}}\right\} + \min\left\{s(a-b), (a-b)^{\frac{1}{2\alpha}}\right\}. \tag{72}$$

As in Case 1, by the subadditivity of ω we see that in order to prove (72) it suffices to show that $\min\{sa, 2r + a^{1/(2\alpha)}\} \leqslant 2r + \omega(a) = 2r + \min\{sa, a^{1/(2\alpha)}\}$, which is immediate.

Case 3: $\{i, j\} = \{0, 1\}$. In this case, by interchanging the roles of b and $a-b$, if necessary, in order to prove the desired inequality (70) one must show that the following two inequalities hold true

$$r + \min\left\{sa, a^{\frac{1}{2\alpha}}\right\} \leqslant b^{\frac{1}{2\alpha}} + r + \min\left\{s(a-b), (a-b)^{\frac{1}{2\alpha}}\right\},$$

and

$$r + \min\left\{sa, a^{\frac{1}{2\alpha}}\right\} \leqslant r + \min\left\{sb, b^{\frac{1}{2\alpha}}\right\} + \min\left\{s(a-b), 2r + (a-b)^{\frac{1}{2\alpha}}\right\}.$$

Thus, we need to check that

$$\min\left\{sa, a^{\frac{1}{2\alpha}}\right\} \leqslant b^{\frac{1}{2\alpha}} + \min\left\{s(a-b), (a-b)^{\frac{1}{2\alpha}}\right\}, \tag{73}$$

and

$$\min\left\{sa, a^{\frac{1}{2\alpha}}\right\} \leqslant \min\left\{sb, b^{\frac{1}{2\alpha}}\right\} + \min\left\{s(a-b), 2r + (a-b)^{\frac{1}{2\alpha}}\right\}. \tag{74}$$

Inequality (74) follows immediately from the subadditivity of ω as follows.

$$\min\left\{sb, b^{\frac{1}{2\alpha}}\right\} + \min\left\{s(a-b), 2r + (a-b)^{\frac{1}{2\alpha}}\right\} \geqslant \omega(b) + \omega(a-b) \geqslant \omega(a) = \min\left\{sa, a^{\frac{1}{2\alpha}}\right\}.$$

Since the mapping $t \mapsto t^{1/(2\alpha)}$ is subadditive, we have $b^{1/(2\alpha)} + (a-b)^{1/(2\alpha)} \geqslant a^{1/(2\alpha)}$. Therefore, in order to prove (73) it suffices to show that $b^{1/(2\alpha)} + s(a-b) \geqslant \min\{sa, a^{1/(2\alpha)}\}$. Define $\psi : [0, a] \to \mathbb{R}$ by $\psi(t) = t^{1/(2\alpha)} + s(a-t)$. Then ψ is concave. It follows that the minimum of ψ on the interval $[0, a]$ is attained at one of its endpoints, and therefore $\psi(b) \geqslant \min\{\psi(0), \psi(a)\} = \min\{sa, a^{1/(2\alpha)}\}$. $\square$

The proof of the following theorem is a variant of an argument of [JLS86].

Theorem 16. *Continuing with the notation of Lemma 15 and assuming from now on that the condition* (68) *is satisfied, define* $f : \mathbb{F}_2^m \times \{0\} \to \ell_2^m$ *by setting* $f(x, 0) = x$ *for every* $x \in \mathbb{F}_2^m$*. Then* $\|f(x, 0) - f(y, 0)\|_2 = d((x, 0), (y, 0))^\alpha$ *for every* $x, y \in \mathbb{F}_2^m$*. At the same time, if* $F : \mathbb{F}_2^m \times \mathbb{F}_2 \to \ell_2^m$ *extends* f *and satisfies* $\|F(u) - F(v)\|_2 \leqslant Ld(u, v)^\alpha$ *for some* $L \in (0, \infty)$ *and every* $u, v \in \mathbb{F}_2^m \times \mathbb{F}_2$ *then we necessarily have*

$$L \geqslant \frac{\sqrt{m}}{s^\alpha \sqrt{m} + 2r^\alpha}. \tag{75}$$

Proof. Note that for every $x, y \in \mathbb{F}_2^m$ we have

$$\|f(x, 0) - f(y, 0)\|_2 = \|x - y\|_2 = \sqrt{\|x - y\|_1} = d((x, 0), (y, 0))^\alpha,$$

by the definition of d in (69). Next, the assumptions on F imply that for every $x, y \in \mathbb{F}_2^m$ and every $j \in \{1, \ldots, m\}$ we have

$$\|F(x + e_j, 1) - F(x, 1)\|_2 \leqslant Ld((x + e_j, 1), (x, 1))^\alpha \overset{(69)}{=} L \min\{s^\alpha, (2r + 1)^\alpha\} \leqslant Ls^\alpha.$$

Hence,

$$\sum_{j=1}^{m} \sum_{x \in \mathbb{F}_2^m} \|F(x + e_j, 1) - F(x, 1)\|_2^2 \leqslant m 2^m L^2 s^{2\alpha}. \tag{76}$$

Also, the following estimate holds true for every $x \in \mathbb{F}_2^m$.

$$\begin{aligned}
&\|F(x + e, 1) - F(x, 1)\|_2 \\
&\qquad \geqslant \|F(x + e, 0) - F(x, 0)\|_2 - \|F(x + e, 1) - F(x + e, 0)\|_2 - \|F(x, 1) - F(x, 0)\|_2 \\
&\qquad \geqslant \|f(x + e, 0) - f(x, 0)\|_2 - Ld((x + e, 1), (x + e, 0))^\alpha - Ld((x, 1), (x, 0))^\alpha \\
&\qquad = \sqrt{m} - 2Lr^\alpha.
\end{aligned}$$

Hence,
$$\sum_{x\in\mathbb{F}_2^m} \|F(x+e,1)-F(x,1)\|_2^2 \geqslant 2^m \left(\max\left\{\sqrt{m}-2Lr^\alpha,0\right\}\right)^2. \tag{77}$$
By a classical inequality of Enflo [Enf69] (see also [Mat02, § 15.4]),
$$\sum_{x\in\mathbb{F}_2^m} \|F(x+e,1)-F(x,1)\|_2^2 \leqslant \sum_{j=1}^m\sum_{x\in\mathbb{F}_2^m} \|F(x+e_j,1)-F(x,1)\|_2^2. \tag{78}$$
A substitution of (76) and (77) into (78) now yields the estimate
$$\sqrt{m}-2Lr^\alpha \leqslant Ls^\alpha\sqrt{m} \implies L \geqslant \frac{\sqrt{m}}{s^\alpha\sqrt{m}+2r^\alpha}. \qquad \square$$

One should clearly choose those $r,s\in(0,\infty)$ that satisfy the constraint (68) and maximize the right hand side of (68). This yields better dependence (in terms of an α-dependent constant factor but not in term of dependence on n) than the following sub-optimal choices.
$$r \stackrel{\mathrm{def}}{=} m^{\frac{1}{4\alpha^2}} \qquad \text{and} \qquad s \stackrel{\mathrm{def}}{=} m^{-\frac{2\alpha-1}{4\alpha^2}}. \tag{79}$$
It is elementary to check that under these choices the constraint (68) is satisfied, yielding the estimate $L \gtrsim m^{(2\alpha-1)/(4\alpha)}$ in Theorem 16. This implies the validity of (66), i.e., the first half of Theorem 4. Namely, the following bounds hold true.

Corollary 17. *Continuing with the notation of Lemma 15, we have*
$$e^\alpha(\ell_\infty,\ell_2^m) \geqslant e^\alpha\big((\mathbb{F}_2^m\times\mathbb{F}_2,d),\mathbb{F}_2^m\times\{0\},\ell_2^m\big) \gtrsim m^{\frac{2\alpha-1}{4\alpha}}.$$
Also, since $m \asymp \log|\mathbb{F}_2^m|$, it follows that for arbitrarily large $n\in\mathbb{N}$ we have
$$e_n^\alpha(\ell_\infty,\ell_2) \gtrsim (\log n)^{\frac{2\alpha-1}{4\alpha}}.$$

Remark 18. Consider the mappings $\omega_0,\omega_1:[0,\infty)\to[0,\infty)$ given by
$$\forall\, t\geqslant 0,\qquad \omega_0(t)=t^{1/(2\alpha)} \qquad \text{and} \qquad \omega_1(t)=\min\{st,2r+\omega_0(t)\}.$$
Both ω_0 and ω_1 are concave, increasing and vanish at the origin. The metric d given in (69) satisfies $d((x,0),d(y,0))=\omega_0(\|x-y\|_1)$ and $d((x,1),(y,1))=\omega_1(\|x-y\|_1)$. As explained in Remark 5.5 of [MN15], it follows that both of the metric spaces $(\mathbb{F}_2^m\times\{0\},d)$ and $(\mathbb{F}_2^m\times\{1\},d)$ embed into ℓ_1 with $O(1)$ distortion. We do not know whether the metric space $(\mathbb{F}_2^m\times\mathbb{F}_2,d)$ admits an embedding into ℓ_1 with $O(1)$ distortion. If it were the case that any embedding of $(\mathbb{F}_2^m\times\mathbb{F}_2,d)$ (say, when $\alpha=1$) into ℓ_1 must incur bi-Lipschitz distortion that tends to ∞ as $n\to\infty$, this would be the first known example of a metric space that can be partitioned into two subsets, each of which well-embeds into ℓ_1 yet the entire space does not. For more on such questions, see [MN13b].

Remark 19. The metric d in (69), as well as the magnification of a metric space that was described in Section 3.1, are both special cases of a generalization of a construction that was used in [JLS86] for the purpose of proving a Lipschitz-nonextendability result. Variants of this construction were also used in [Lan99, CKR05]. The general idea is the following procedure to "glue" two metric spaces. Suppose that (X,d_X) and (Y,d_Y) are finite metric spaces with X and Y disjoint as sets. Suppose also that we are given a mapping $\sigma:X\to Y$ and $r\in(0,\infty)$. Define a weighted graph structure on $X\cup Y$ as follows. If $x_1,x_2\in X$ then join x_1 and x_2 by an edge of weight $d_X(x_1,x_2)$. Similarly, if $y_1,y_2\in Y$ then join y_1 and y_2 by an edge of weight $d_Y(y_1,y_2)$. Also, for every $x\in X$ join x and $\sigma(x)$ by an edge of weight r (a further generalization of this procedure could allow the weight of the edge $\{x,\sigma(x)\}$ to depend on x). Consider now the shortest-path metric that this weighted graph induces on $X\cup Y$. One can check that the above metric spaces can also be described as subsets of this general construction, whose usefulness is probably yet to be fully exploited.

4.2. **Linearization.** Our goal here is to prove (67), thus completing the proof of Theorem 4. The argument below is based on the linearization procedure of [JL84], with several modifications that yield quantitative improvements and, more importantly, allow us to to treat Hölder mappings.

In what follows, given a Banach space $(W, \|\cdot\|_W)$ we denote

$$B_W \stackrel{\mathrm{def}}{=} \{x \in W:\ \|x\|_W \leqslant 1\} \qquad \text{and} \qquad S_W \stackrel{\mathrm{def}}{=} \partial B_W = \{x \in W:\ \|x\|_W = 1\}.$$

If $V \subseteq W$ is a linear subspace then we will always consider it as being equipped with the norm inherited from W, thus slightly abusing notation by denoting the sets $V \cap B_W$ and $V \cap S_W$ by B_V and S_V, respectively. A mapping $h : W_1 \to W_2$ between two real vector spaces W_1, W_2 is said to be positively homogeneous if $h(\lambda x) = \lambda h(x)$ for every $\lambda \in [0,\infty)$.

Lemma 20 below is a variant of Lemma 5 in [JL84], with the difference being that in [JL84] the conclusion corresponding to (81) is weaker in the sense that the factor $\sqrt{m}D$ is replaced by m. The difference between our proof and the proof in [JL84] is that we perform averaging in the image of T rather than in X itself, allowing for a more refined estimate that we shall use later.

Lemma 20. *Fix $m \in \mathbb{N}$ and $D \in [1,\infty)$. Let $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ be Banach spaces and $X \subseteq Y$ an m-dimensional linear subspace of Y. Suppose that $T : X \to \ell_2^m$ is a linear operator that satisfies*

$$\forall\, x \in X, \qquad \|x\|_Y \leqslant \|Tx\|_2 \leqslant D\|x\|_Y. \tag{80}$$

Suppose also that we are given a positively homogeneous Lipschitz mapping $\psi : X \to Z$ and a linear operator $U : X \to Z$. Then there exists a positively homogeneous mapping $\Psi : Y \to Z$ that satisfies

$$\|\Psi\|_{\mathrm{Lip}} \lesssim \|\psi\|_{\mathrm{Lip}} \qquad \text{and} \qquad \|\Psi|_X - U\|_{\mathrm{Lip}} \lesssim \sqrt{m}D \sup_{z \in S_X} \|\psi(z) - Uz\|_Z. \tag{81}$$

Proof. Write $B_2^m = B_{\ell_2^m}$ and

$$v_m = \mathrm{vol}\,(B_2^m) = \frac{\pi^{\frac{m}{2}}}{\Gamma\left(1 + \frac{m}{2}\right)}. \tag{82}$$

Define $\phi : Y \to Z$ as follows.

$$\forall\, y \in Y, \qquad \phi(y) \stackrel{\mathrm{def}}{=} \frac{1}{v_m}\int_{B_2^m} \psi\left(y + T^{-1}w\right) dw = \frac{1}{v_m}\int_{Ty + B_2^m} \psi\left(T^{-1}w\right) dw.$$

Then, because ϕ is obtained by applying an averaging operator to ψ, we have

$$\|\phi\|_{\mathrm{Lip}} \leqslant \|\psi\|_{\mathrm{Lip}}. \tag{83}$$

Also, because $\psi(0) = 0$ (recall that ψ is positively homogeneous), we have

$$\sup_{y \in S_Y} \|\phi(y)\|_Z \leqslant \|\psi\|_{\mathrm{Lip}} \sup_{(y,w) \in S_Y \times B_2^m} \|y + T^{-1}w\|_Y \stackrel{(80)}{\leqslant} 2\|\psi\|_{\mathrm{Lip}}. \tag{84}$$

Similarly,

$$\begin{aligned}\sup_{y \in S_X} \|\phi(y) - Uy\|_Z &= \sup_{y \in S_X} \left\|\frac{1}{v_m}\int_{B_2^m} (\psi - U)\left(y + T^{-1}w\right) dw\right\|_Z \\ &\leqslant \sup_{(y,w) \in S_X \times B_2^m} \|(\psi - U)(y + T^{-1}w)\|_Z \leqslant \sup_{z \in 2B_X} \|\psi(z) - Uz\|_Z = 2\sup_{z \in S_X} \|\psi(z) - Uz\|_Z,\end{aligned} \tag{85}$$

where the penultimate step in (85) holds true because, due to (80), we have $T^{-1}B_2^m \subseteq B_X$, so $S_X + T^{-1}B_2^m \subseteq 2B_X$. The final step of (85) holds true because $\psi - U$ is positively homogeneous.

Fix distinct $x, y \in S_X$ and write $2r \stackrel{\mathrm{def}}{=} \|x-y\|_Y$. Then, since U and T are linear,

$$
\begin{aligned}
&(\phi-U)(x)-(\phi-U)(y)\\
&\quad=\frac{1}{v_m}\int_{(Tx+B_2^m)\setminus(Ty+B_2^m)}(\psi-U)(T^{-1}(w))dw-\frac{1}{v_n}\int_{(Ty+B_2^m)\setminus(Tx+B_2^n)}(\psi-U)(T^{-1}(w))dw.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
\|(\phi-U)(x)-(\phi-U)(y)\|_Z &\leqslant \frac{1}{v_m}\int_{(Tx+B_2^m)\triangle(Ty+B_2^m)}\|(\psi-U)(T^{-1}(w))\|_Z dw\\
&\leqslant \frac{\mathrm{vol}\left((Tx+B_2^m)\triangle(Ty+B_2^m)\right)}{v_n}\cdot \sup_{z\in(x+T^{-1}B_2^m)\triangle(y+T^{-1}B_2^m)}\|\psi(z)-Uz\|_Z. \quad (86)
\end{aligned}
$$

We note that

$$
\mathrm{vol}\left((Tx+B_2^m)\cup(Ty+B_2^m)\right)\leqslant v_m+\|Tx-Ty\|_2 v_{m-1}. \quad (87)
$$

Indeed, by rotation invariance, it suffices to verify (87) when $Tx = re_1$ and $Ty = -re_1$, where $e_1, \ldots, e_m$ is the standard basis of $\mathbb{R}^m$ and we recall that $\|Tx-Ty\|_2 = 2r$. Identifying $\mathbb{R}^m$ with $\mathbb{R}\times\mathbb{R}^{m-1}$, the estimate (87) is a consequence of the fact that $(re_1+B_2^m)\cup(-re_1+B_2^m)$ is contained in the union of the three sets $([r,\infty)\times\mathbb{R}^{m-1})\cap(re_1+B_2^m)$, $((-\infty,-r]\times\mathbb{R}^{m-1})\cap(-re_1+B_2^m)$ and $[-r,r]\times B_2^{m-1}$, whose interiors are disjoint.

Now, due to (87) we have

$$
\begin{aligned}
\mathrm{vol}\left((Tx+B_2^m)\triangle(Ty+B_2^m)\right) &= 2\mathrm{vol}\left((Tx+B_2^m)\cup(Ty+B_2^m)\right)-\mathrm{vol}\left(Tx+B_2^m\right)-\mathrm{vol}\left(Ty+B_2^m\right)\\
&\leqslant 2\|Tx-Ty\|_2 v_{m-1}.
\end{aligned}
$$

So, the first term in the right hand side of (86) can be bounded as follows.

$$
\frac{\mathrm{vol}\left((Tx+B_2^m)\triangle(Ty+B_2^m)\right)}{v_n}\leqslant\frac{2v_{m-1}}{v_m}\|Tx-Ty\|_2\overset{(80)}{\leqslant}\sqrt{m}D\|x-y\|_Y, \quad (88)
$$

where we used the fact that $2v_{m-1}/v_m \leqslant \sqrt{m}$, which follows from (82) and Stirling's formula (in addition, we actually have $\lim_{m\to\infty} 2v_{m-1}/v_m = \sqrt{2/\pi}$).

To bound the second term in the right hand side of (86), note that $T^{-1}B_2^m \subseteq B_X$, by (80). Therefore $(x+T^{-1}B_2^m)\triangle(y+T^{-1}B_2^m)\subseteq 2B_X$, since $x, y\in S_X$. So,

$$
\sup_{z\in(x+T^{-1}B_2^m)\triangle(y+T^{-1}B_2^m)}\|\psi(z)-Uz\|_Z\leqslant\sup_{z\in 2B_X}\|\psi(z)-Uz\|_Z=2\sup_{z\in S_X}\|\psi(z)-Uz\|_Z, \quad (89)
$$

where we used the fact that $\psi - U$ is positively homogeneous.

A substitution of (88) and (89) into (86) shows that

$$
\|(\phi-U)|_{S_X}\|_{\mathrm{Lip}}\leqslant 4\sqrt{m}D\sup_{z\in S_X}\|\psi(z)-Uz\|_Z. \quad (90)
$$

Define $\Psi : Y \to Z$ to be the positively homogeneous extension of $\phi|_{S_Y}$, i.e., $\Psi(y) \stackrel{\mathrm{def}}{=} \|y\|_Y\phi(y/\|y\|_Y)$ for every $y \in Y \setminus \{0\}$ and $\Psi(0) = 0$. Then a simple computation (that is carried out in detail in [JL84, Lemma 2]) shows that

$$
\|\Psi\|_{\mathrm{Lip}}\leqslant 2\|\phi|_{S_Y}\|_{\mathrm{Lip}}+\sup_{y\in S_Y}\|\phi(y)\|_Z\overset{(83)\wedge(84)}{\leqslant}4\|\psi\|_{\mathrm{Lip}}.
$$

For the same reason, since $\Psi - U$ is the positively homogeneous extension of $(\phi - U)|_{S_X}$ to X,

$$
\|\Psi|_X-U\|_{\mathrm{Lip}}\leqslant 2\|(\phi-U)|_{S_X}\|_{\mathrm{Lip}}+\sup_{y\in S_X}\|\phi(y)-Uy\|_Z\overset{(85)\wedge(90)}{\leqslant}10\sqrt{m}D\sup_{z\in S_X}\|\psi(z)-Uz\|_Z. \qquad \square
$$

We shall use below the following lemma of Begun [Beg99]; alternatively one could use below an argument of Bourgain [Bou87] that Begun significantly simplified in [Beg99].

Lemma 21. *Fix $\delta, L, \tau \in (0,\infty)$ and $m \in \mathbb{N}$. Suppose that $(Y, \|\cdot\|_Y)$ is an m-dimensional normed space and that $(Z, \|\cdot\|_Z)$ is a Banach space. Let $K \subseteq Y$ be a convex set and suppose that we are given a mapping $F : K + \tau B_Y \to Z$ that satisfies*

$$\forall\, x, y \in K + \tau B_Y, \qquad \|F(x) - F(y)\|_Z \leqslant L(\|x-y\|_Y + \delta).$$

Define $h : K \to Z$ by

$$\forall\, x \in K, \qquad h(x) \stackrel{\mathrm{def}}{=} \frac{1}{\mathrm{vol}(\tau B_Y)} \int_{\tau B_Y} F(x+y)dy, \tag{91}$$

where the integration and $\mathrm{vol}(\cdot)$ are interpreted through an identification of Y with $\mathbb{R}^m$. Then

$$\|h\|_{\mathrm{Lip}} \leqslant L\left(1 + \frac{\delta m}{2\tau}\right).$$

Corollary 22. *Fix $\tau \in (0,\infty)$, $\alpha \in (0,1]$ and $m \in \mathbb{N}$. Suppose that $(Y, \|\cdot\|_Y)$ is an m-dimensional normed space and that $(Z, \|\cdot\|_Z)$ is a Banach space. Let $K \subseteq Y$ be a convex set and suppose that $F : K + \tau B_Y \to Z$ is α-Hölder. Then there exists $h : K \to Z$ that satisfies*

$$\sup_{x\in K} \|h(x) - F(x)\|_Z \leqslant \tau^\alpha \|F\|_{\mathrm{Lip}(\alpha)} \qquad \text{and} \qquad \|h\|_{\mathrm{Lip}} \leqslant \frac{\alpha^2 (m/\tau)^{1-\alpha} + 1}{\alpha} \|F\|_{\mathrm{Lip}(\alpha)}.$$

Proof. Denote $L = \|F\|_{\mathrm{Lip}(\alpha)}$. Note that for every $\delta \in (0,\infty)$ and every $x, y \in K + \tau B_Y$ we have

$$\|F(x) - F(y)\|_Z \leqslant L\|x-y\|_Y^\alpha \leqslant L(\|x-y\|_Z + \delta) \sup_{t\geqslant 0} \frac{t^\alpha}{t+\delta} = \frac{(1-\alpha)^{1-\alpha}\alpha^\alpha L}{\delta^{1-\alpha}} (\|x-y\|_Z + \delta).$$

We may therefore apply Lemma 21, deducing that for h defined as in (91) we have

$$\|h\|_{\mathrm{Lip}} \leqslant \frac{(1-\alpha)^{1-\alpha}\alpha^\alpha L}{\delta^{1-\alpha}} \left(1 + \frac{\delta m}{2\tau}\right). \tag{92}$$

The value of δ that minimizes the right hand side of (92) is $\delta = 2\tau(1-\alpha)/(\alpha m)$, which gives

$$\|h\|_{\mathrm{Lip}} \leqslant (1-\alpha)^{1-\alpha}\alpha^\alpha \left(\frac{\alpha^{1-\alpha}}{(1-\alpha)^{1-\alpha}} \left(\frac{m}{2\tau}\right)^{1-\alpha} + \frac{1-\alpha}{\alpha}\right) L \leqslant \frac{\alpha^2(m/\tau)^{1-\alpha} + 1}{\alpha} L.$$

It remains to note that for every $x \in K$ we have

$$\|h(x) - F(x)\|_Z = \left\| \frac{1}{\mathrm{vol}(\tau B_Y)} \int_{\tau B_Y} (F(x+y) - F(x)) dy \right\|_Z \leqslant \frac{1}{\mathrm{vol}(\tau B_Y)} \int_{\tau B_Y} L\|y\|_Y^\alpha dy \leqslant L\tau^\alpha. \qquad \square$$

Theorem 23. *There is a universal constant $\kappa \in (0,1)$ with the following properties. Fix $\alpha \in (0,1]$, $M, m \in \mathbb{N}$ and $C, D \in [1,\infty)$. Suppose that $(Y, \|\cdot\|_Y)$ is an M-dimensional normed space and that $(Z, \|\cdot\|_Z)$ is an m-dimensional normed space. Let $X \subseteq Y$ be an m-dimensional subspace of Y whose Banach–Mazur distance to ℓ_2^m is at most D, i.e., there exists a linear operator $T : X \to \ell_2^m$ satisfying (80). Suppose also that $U : X \to Z$ is a linear operator with*

$$\forall x \in X, \qquad \|x\|_Y \leqslant \|Ux\|_Z \leqslant C\|x\|_Y. \tag{93}$$

For every $\varepsilon \in (0,1)$ let $\mathcal{N}_\varepsilon$ be an ε-net in S_X. Then, under the additional assumption that

$$\varepsilon \leqslant \left(\frac{\kappa}{\sqrt{m}CDe^\alpha(Y, \mathcal{N}_\varepsilon, Z)}\right)^{\frac{1}{\alpha}}, \tag{94}$$

there exists a linear operator $V : Y \to Z$ that extends U and whose operator norm satisfies

$$\|V\|_{Y\to Z} \lesssim \alpha M^{1-\alpha} m^{\frac{1-\alpha}{2\alpha}} C^{\frac{1}{\alpha}} D^{\frac{1-\alpha}{\alpha}} e^{\alpha}(Y, \mathcal{N}_\varepsilon, Z)^{\frac{1}{\alpha}} + \frac{C}{\alpha} e^{\alpha}(Y, \mathcal{N}_\varepsilon, Z). \tag{95}$$

Proof. If $x, y \in B_X$ then $\|x-y\|_Y \leqslant 2$, hence $\|Ux - Uy\|_Z \leqslant C\|x-y\|_Y \leqslant 2^{1-\alpha}C\|x-y\|_Y^{\alpha}$. This implies that the restriction of U to $\mathcal{N}_\varepsilon$ is α-Hölder with constant $2C$. Denoting $L = e^{\alpha}(Y, \mathcal{N}_\varepsilon, Z)$, it follows that there exists $g : Y \to Z$ that extends $U|_{\mathcal{N}_\varepsilon}$ and is α-Hölder with constant $2CL$.

Let $\rho : Z \to CB_Z$ be the radial retraction of Z onto CB_Z, i.e., $\rho(z) = z$ if $\|z\|_Z \leqslant C$ and $\rho(z) = Cz/\|z\|_Z$ if $\|z\|_Z > C$. It is straightforward to check that ρ is 2-Lipschitz. Therefore, if we define $F = \rho \circ g$ then $\|F\|_{\mathrm{Lip}(\alpha)} \leqslant 2\|g\|_{\mathrm{Lip}(\alpha)} \leqslant 4CL$ and $\|F(y)\|_Z \leqslant C$ for every $y \in Y$. Moreover, for every $a \in \mathcal{N}_\varepsilon$ we have $\|g(a)\|_Z = \|Ua\|_Z \leqslant C$, so $F(a) = \rho(g(a)) = g(a)$. Thus $F : Y \to CB_Z$ is also an extension of $U|_A$. Fix $\tau \in (0,1)$ that will be determined later. By Corollary 22 (applied with $K = Y$) there exists $h : Y \to Z$ that satisfies

$$\forall\, y \in Y \qquad \|h(y) - F(y)\|_Z \leqslant 4CL\tau^{\alpha}, \tag{96}$$

and

$$\|h\|_{\mathrm{Lip}} \leqslant 4\frac{\alpha^2(M/\tau)^{1-\alpha} + 1}{\alpha} CL. \tag{97}$$

Then

$$\sup_{y\in S_Y} \|h(y)\|_Z \overset{(96)}{\leqslant} \sup_{y \in S_Y} \|F(y)\|_Z + 4CL\tau^{\alpha} \leqslant C + 4CL\tau^{\alpha}, \tag{98}$$

where in the last step of (98) we used the fact that F takes values in CB_Z.

Denote by $\psi : Y \to Z$ the positively homogeneous extension of $h|_{S_Y}$ to Y. By [JL84, Lemma 2],

$$\|\psi\|_{\mathrm{Lip}} \leqslant 2\|h|_{S_Y}\|_{\mathrm{Lip}} + \sup_{y\in S_Y}\|h(y)\|_Z \overset{(97)\wedge(98)}{\leqslant} 4\frac{\alpha^2(M/\tau)^{1-\alpha}+1}{\alpha}CL + C + 4CL\tau^{\alpha}. \tag{99}$$

Take $x \in S_X$ (thus $\psi(x) = h(x)$) and $a \in \mathcal{N}_\varepsilon$ with $\|a - x\|_Y \leqslant \varepsilon$. Then $h(a) = F(a) = Ua$. Hence,

$$\begin{aligned}\|\psi(x) - Ux\|_Z &\leqslant \|h(x) - F(x)\|_Z + \|F(x) - F(a)\|_Z + \|Ua - Ux\|_Z \\ &\leqslant 4CL\tau^{\alpha} + 4CL\|x-a\|_Y^{\alpha} + C\|x-a\|_Y \leqslant 4CL\tau^{\alpha} + 4CL\varepsilon^{\alpha} + C\varepsilon \leqslant 5CL(\tau^{\alpha}+\varepsilon^{\alpha}). \end{aligned} \tag{100}$$

Due to (99) and (100), Lemma 20 implies that there exists $\Psi : Y \to Z$ that satisfies

$$\|\Psi\|_{\mathrm{Lip}} \lesssim \frac{\alpha^2(M/\tau)^{1-\alpha}+1}{\alpha}CL + C + CL\tau^{\alpha} \lesssim \frac{\alpha^2(M/\tau)^{1-\alpha}+1}{\alpha}CL,$$

and

$$\|\Psi|_X - U\|_{\mathrm{Lip}} \lesssim \sqrt{m}CDL(\tau^{\alpha}+\varepsilon^{\alpha}).$$

By Proposition 1 in [JL84], which relies on an important linearization result due to Lindenstrauss [Lin64], there exists a linear operator $S : Y \to Z$ such that

$$\|S\|_{Y\to Z} \leqslant \|\Psi\|_{\mathrm{Lip}} \lesssim \frac{\alpha^2(M/\tau)^{1-\alpha}+1}{\alpha}CL, \tag{101}$$

and

$$\|S|_X - U\|_{X\to Z} \leqslant \|\Psi|_X - U\|_{\mathrm{Lip}} \lesssim \sqrt{m}CDL(\tau^{\alpha}+\varepsilon^{\alpha}). \tag{102}$$

Our assumption (94) says that $\sqrt{m}CDL\varepsilon^{\alpha} \leqslant \kappa$. So, if we choose $\tau = (\kappa/(\sqrt{m}CDL))^{1/\alpha}$ and $\kappa > 0$ is a small enough universal constant then it would follow from (101) and (102) that

$$\|S\|_{Y\to Z} \lesssim \alpha M^{1-\alpha} m^{\frac{1-\alpha}{2\alpha}} C^{\frac{1}{\alpha}} D^{\frac{1-\alpha}{\alpha}} L^{\frac{1}{\alpha}} + \frac{CL}{\alpha} \qquad \text{and} \qquad \|S|_X - U\|_{X\to Z} \leqslant \frac{1}{2}. \tag{103}$$

Letting $I_Z : Z \to Z$ denote the identity mapping on Z, we have

$$\left\|(S|_X)U^{-1} - I_Z\right\|_{Z\to Z} = \left\|(S|_X - U)U^{-1}\right\|_{Z\to Z} \leqslant \|S|_X - U\|_{X\to Z}\|U^{-1}\|_{Z\to X} \overset{(93)\wedge(103)}{\leqslant} \frac{1}{2}.$$

Consequently, the operator $(S|_X)U^{-1} : Z \to Z$ is invertible and $\|((S|_X)U^{-1})^{-1}\|_{Z\to Z} \leqslant 2$. Define

$$V \stackrel{\mathrm{def}}{=} ((S|_X)U^{-1})^{-1}S : Y \to Z.$$

Then V extends S and

$$\|V\|_{Y\to Z} \leqslant \left\|((S|_X)U^{-1})^{-1}\right\|_{Z\to Z}\|S\|_{Y\to Z} \leqslant 2\|S\|_{Y\to Z} \overset{(103)}{\lesssim} \alpha M^{1-\alpha} m^{\frac{1-\alpha}{2\alpha}} C^{\frac{1}{\alpha}} D^{\frac{1-\alpha}{\alpha}} L^{\frac{1}{\alpha}} + \frac{CL}{\alpha}. \qquad \square$$

Given a normed space $(Y, \|\cdot\|_Y)$ and a linear subspace $X \subseteq Y$, the projection constant of X relative to Y is denoted $\lambda(X,Y)$. Thus $\lambda(X,Y)$ is the infimum over those $\lambda \in [1,\infty]$ for which there exists a linear projection P from Y onto X with $\|P\|_{Y\to X} \leqslant \lambda$.

Corollary 24. *Let κ be the universal constant from Theorem 23. Fix $\alpha \in (0,1]$, $M, m \in \mathbb{N}$ and $D \in [1,\infty)$. Let $X \subseteq Y$ be an m-dimensional subspace of Y whose Banach–Mazur distance to ℓ_2^m is at most D. For every $\varepsilon \in (0,1)$ let $\mathcal{N}_\varepsilon$ be an ε-net in S_X. Then,*

$$\varepsilon \leqslant \frac{\kappa^{\frac{1}{\alpha}} M^{1-\alpha}}{\sqrt{m}D\lambda(X,Y)} \implies e^\alpha(Y, \mathcal{N}_\varepsilon, \ell_2^m) \gtrsim \frac{\lambda(X,Y)^\alpha}{m^{\frac{1-\alpha}{2}} M^{\alpha(1-\alpha)} D^{2-\alpha}}. \tag{104}$$

In particular, there exists $n \in \mathbb{N}$ satisfying $\log n \asymp m\left(\frac{1}{\alpha} + \log m\right)$ for which

$$\min\left\{e^\alpha(Y, \ell_2^m), e_n^\alpha(Y, \ell_2^m)\right\} \gtrsim \frac{\lambda(X,Y)^\alpha}{m^{\frac{1-\alpha}{2}} M^{\alpha(1-\alpha)} D^{2-\alpha}}. \tag{105}$$

Proof. Since the Banach–Mazur distance of X to ℓ_2^m is at most D, the lower bound on $e^\alpha(Y, \mathcal{N}_\varepsilon, \ell_2^m)$ that appears in (104) will follow if we show that

$$e^\alpha(Y, \mathcal{N}_\varepsilon, X) \gtrsim \frac{\lambda(X,Y)^\alpha}{m^{\frac{1-\alpha}{2}} M^{\alpha(1-\alpha)} D^{1-\alpha}}. \tag{106}$$

We may assume from now on that

$$e^\alpha(Y, \mathcal{N}_\varepsilon, X) \leqslant \frac{\lambda(X,Y)^\alpha}{m^{\frac{1-\alpha}{2}} M^{\alpha(1-\alpha)} D^{1-\alpha}}, \tag{107}$$

since otherwise there is nothing to prove. By the assumption on ε in (104) combined with (107),

$$\varepsilon \leqslant \frac{\kappa^{\frac{1}{\alpha}} M^{1-\alpha}}{\sqrt{m}D\lambda(X,Y)} \leqslant \left(\frac{\kappa}{\sqrt{m}De^\alpha(Y, \mathcal{N}_\varepsilon, Z)}\right)^{\frac{1}{\alpha}}.$$

We can therefore apply Theorem 23 with $Z = X$ and $U = I_X$ being the identity on X (in particular, $C = 1$). The operator $V : Y \to X$ thus obtained is a projection onto X, since $V|_X = I_X$. Hence,

$$\lambda(X,Y) \leqslant \|V\|_{Y\to X} \overset{(95)}{\lesssim} \alpha M^{1-\alpha} m^{\frac{1-\alpha}{2\alpha}} D^{\frac{1-\alpha}{\alpha}} e^\alpha(Y, \mathcal{N}_\varepsilon, X)^{\frac{1}{\alpha}} + \frac{e^\alpha(Y, \mathcal{N}_\varepsilon, X)}{\alpha}$$
$$\lesssim \frac{M^{1-\alpha} m^{\frac{1-\alpha}{2\alpha}} D^{\frac{1-\alpha}{\alpha}} e^\alpha(Y, \mathcal{N}_\varepsilon, X)^{\frac{1}{\alpha}}}{\alpha},$$

which simplifies to give the desired estimate (106).

To deduce (105), note that by the Kadec′–Snobar theorem [KS71] we have $\lambda(X,Y) \leqslant \sqrt{m}$ and by John's theorem [Joh48] we have $D \leqslant \sqrt{m}$. Therefore, if we choose $\varepsilon = \kappa^{1/\alpha}/m^{3/2}$ then the upper bound on ε that appears in (104) holds true. It remains to note that if we set $n = |\mathcal{N}_\varepsilon|$ then

$\min\{e^\alpha(X,\ell_2^m), e_n^\alpha(X,\ell_2^m)\} \geqslant e^\alpha(Y,\mathcal{N}_\varepsilon,X)$, and by standard estimates (e.g. [MS86]) on the size of ε-nets in m-dimensional normed spaces we have $\log n \asymp m\log(1/\varepsilon) \asymp m(\log m + 1/\alpha)$. $\square$

The following corollary implies (67), thus completing the proof of Theorem 4.

Corollary 25. *For arbitrarily large $m,n\in\mathbb{N}$ and every $\alpha\in(1/2,1]$ we have*

$$e^\alpha(\ell_1,\ell_2^m) \gtrsim m^{\alpha^2-\frac12} \qquad \text{and} \qquad e_n^\alpha(\ell_1,\ell_2) \gtrsim \left(\frac{\log n}{\log\log n}\right)^{\alpha^2-\frac12}.$$

Proof. By [FLM77, Kaš77], for every $m\in\mathbb{N}$ there exists an integer $M=O(m)$ and an m-dimensional subspace X of ℓ_1^M whose Banach–Mazur distance to ℓ_2^m is $O(1)$. By [Rut65], we have $\lambda(X,\ell_1^M)\gtrsim\sqrt{m}$. Therefore, an application of Corollary 24 with $D=O(1)$, and $M=O(m)$ shows that there exists $n\in\mathbb{N}$ with $\log n\asymp m\log m$ such that

$$\min\left\{e^\alpha(\ell_1^M,\ell_2^m), e_n^\alpha(\ell_1^M,\ell_2^m)\right\} \gtrsim \frac{m^{\frac{\alpha}{2}}}{m^{\frac{1-\alpha}{2}+\alpha(1-\alpha)}} = m^{\alpha^2-\frac12} \asymp \left(\frac{\log n}{\log\log n}\right)^{\alpha^2-\frac12}. \qquad \square$$

References

[Ach03] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003. Special issue on PODS 2001 (Santa Barbara, CA).

[AFH+04] A. Archer, J. Fakcharoenphol, C. Harrelson, R. Krauthgamer, K. Talwar, and É. Tardos. Approximate classification via earthmover metrics. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1079–1087 (electronic). ACM, New York, 2004.

[Bal92] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[Bal13] K. Ball. The Ribe programme. *Astérisque*, (352):Exp. No. 1047, viii, 147–159, 2013. Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058.

[Beg99] B. Begun. A remark on almost extensions of Lipschitz functions. *Israel J. Math.*, 109:151–155, 1999.

[BFdlV82] B. Bollobás and W. Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2(2):125–134, 1982.

[BL00] 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. ISBN 0-8218-0835-4.

[Bou81] J. Bourgain. A counterexample to a complementation problem. *Compositio Math.*, 43(1):133–144, 1981.

[Bou87] 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. doi:10.1007/BFb0078143.

[CKNZ05] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM J. Discrete Math.*, 18(3):608–625, 2004/05.

[CKR05] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. *SIAM J. Comput.*, 34(2):358–372, 2004/05.

[Die10] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Heidelberg, fourth edition, 2010. ISBN 978-3-642-14278-9. doi:10.1007/978-3-642-14279-6.

[DJP+94] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiterminal cuts. *SIAM J. Comput.*, 23(4):864–894, 1994.

[Dvo61] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

[FHRT03] J. Fakcharoenphol, C. Harrelson, S. Rao, and K. Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 257–265. ACM, New York, 2003.

[FJS88] T. Figiel, W. B. Johnson, and G. Schechtman. Factorizations of natural embeddings of l_p^n into L_r. I. *Studia Math.*, 89(1):79–103, 1988.

[FLM77] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[FTJ79] T. Figiel and N. Tomczak-Jaegermann. Projections onto Hilbertian subspaces of Banach spaces. *Israel J. Math.*, 33(2):155–171, 1979.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[HW71] T. L. Hayden and J. H. Wells. On the extension of Lipschitz-Hölder maps of order β. *J. Math. Anal. Appl.*, 33:627–640, 1971.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[JLS86] W. B. Johnson, J. Lindenstrauss, and G. Schechtman. Extensions of Lipschitz maps into Banach spaces. *Israel J. Math.*, 54(2):129–138, 1986.

[Joh48] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Kal04] N. J. Kalton. Spaces of Lipschitz and Hölder functions and their applications. *Collect. Math.*, 55(2):171–217, 2004.

[Kal12] N. J. Kalton. The uniform structure of Banach spaces. *Math. Ann.*, 354(4):1247–1288, 2012.

[Kar98] A. V. Karzanov. Minimum 0-extensions of graph metrics. *European J. Combin.*, 19(1):71–101, 1998.

[Kaš77] B. S. Kašin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977.

[Kir34] M. D. Kirszbraun. Über die zusammenziehenden und Lipschitzchen Transformationen. *Fundam. Math.*, 22:77–108, 1934.

[KKMR09] H. Karloff, S. Khot, A. Mehta, and Y. Rabani. On earthmover distance, metric labeling, and 0-extension. *SIAM J. Comput.*, 39(2):371–387, 2009.

[KRTJ80] H. König, J. R. Retherford, and N. Tomczak-Jaegermann. On the eigenvalues of $(p, 2)$-summing operators and constants associated with normed spaces. *J. Funct. Anal.*, 37(1):88–126, 1980.

[KS71] M. Ĭ. Kadec′ and M. G. Snobar. Certain functionals on the Minkowski compactum. *Mat. Zametki*, 10:453–457, 1971.

[Lan99] U. Lang. Extendability of large-scale Lipschitz maps. *Trans. Amer. Math. Soc.*, 351(10):3975–3988, 1999.

[Lew78] D. R. Lewis. Finite dimensional subspaces of L_p. *Studia Math.*, 63(2):207–212, 1978.

[Lin64] J. Lindenstrauss. On nonlinear projections in Banach spaces. *Michigan Math. J.*, 11:263–287, 1964.

[LN03] J. R. Lee and A. Naor. Metric decomposition, smooth measures, and clustering, 2003. Preprint, available on request.

[LN04] J. R. Lee and A. Naor. Absolute Lipschitz extendability. *C. R. Math. Acad. Sci. Paris*, 338(11):859–862, 2004.

[LN05] J. R. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[Mat97] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. ISBN 0-387-95373-6. doi:10.1007/978-1-4613-0039-7.

[Mau74] B. Maurey. *Théorèmes de factorisation pour les opérateurs linéaires à valeurs dans les espaces L^p*. Société Mathématique de France, Paris, 1974. With an English summary, Astérisque, No. 11.

[Men27] K. Menger. Zur allgemeinen Kurventheorie. *Fundam. Math.*, 10:96–115, 1927.

[Min70] G. J. Minty. On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Amer. Math. Soc.*, 76:334–339, 1970.

[MM10] K. Makarychev and Y. Makarychev. Metric extension operators, vertex sparsifiers and Lipschitz extendability. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science FOCS 2010*, pages 255–264. IEEE Computer Soc., Los Alamitos, CA, 2010.

[MN13a] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces*, 1:163–199, 2013.

[MN13b] M. Mendel and A. Naor. Ultrametric skeletons. *Proc. Natl. Acad. Sci. USA*, 110(48):19256–19262, 2013.

[MN15] M. Mendel and A. Naor. Expanders with respect to Hadamard spaces and random graphs, 2015. To appear in Duke Math. J.

[MP84] M. B. Marcus and G. Pisier. Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.*, 152(3-4):245–301, 1984.

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. ISBN 3-540-16769-2. With an appendix by M. Gromov.

[Nao01] A. Naor. A phase transition phenomenon between the isometric and isomorphic extension problems for Hölder functions between L_p spaces. *Mathematika*, 48(1-2):253–271 (2003), 2001.

[Nao12] A. Naor. An introduction to the Ribe program. *Jpn. J. Math.*, 7(2):167–233, 2012.

[NPSS06] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov-hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.

[NRS05] A. Naor, Y. Rabani, and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. *J. Funct. Anal.*, 227(2):273–303, 2005.

[Pis79] G. Pisier. Estimations des distances à un espace euclidien et des constantes de projection des espaces de Banach de dimension finie; d'après H. König et al. In *Séminaire d'Analyse Fonctionnelle (1978–1979)*, pages Exp. No. 10, 21. École Polytech., Palaiseau, 1979.

[Rut65] D. Rutovitz. Some parameters associated with finite-dimensional Banach spaces. *J. London Math. Soc.*, 40:241–255, 1965.

[Vil03] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. ISBN 0-8218-3312-X.

[Woj91] P. Wojtaszczyk. *Banach spaces for analysts*, volume 25 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1991.

Mathematics Department, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA.

E-mail address: naor@math.princeton.edu

The Rachel and Selim Benin School of Computer Science and Engineering, Jerusalem 91904, Israel.

E-mail address: yrabani@cs.huji.ac.il