THE ANDONI–KRAUTHGAMER–RAZENSHTEYN CHARACTERIZATION OF SKETCHABLE NORMS FAILS FOR SKETCHABLE METRICS

SUBHASH KHOT AND ASSAF NAOR

ABSTRACT. Andoni, Krauthgamer and Razenshteyn (AKR) proved (STOC 2015) that a finite-dimensional normed space $(X,\|\cdot\|_X)$ admits a $O(1)$ sketching algorithm (namely, with $O(1)$ sketch size and $O(1)$ approximation) if and only if for every $\varepsilon\in(0,1)$ there exist $\alpha\geqslant 1$ and an embedding $f:X\to\ell_{1-\varepsilon}$ such that $\|x-y\|_X\leqslant\|f(x)-f(y)\|_{1-\varepsilon}\leqslant\alpha\|x-y\|_X$ for all $x,y\in X$. The "if part" of this theorem follows from a sketching algorithm of Indyk (FOCS 2000). The contribution of AKR is therefore to demonstrate that the mere availability of a sketching algorithm implies the existence of the aforementioned geometric realization. Indyk's algorithm shows that the "if part" of the AKR characterization holds true for any metric space whatsoever, i.e., the existence of an embedding as above implies sketchability even when X is not a normed space. Due to this, a natural question that AKR posed was whether the assumption that the underlying space is a normed space is needed for their characterization of sketchability. We resolve this question by proving that for arbitrarily large $n\in\mathbb{N}$ there is an n-point metric space $(M(n),d_{M(n)})$ which is $O(1)$-sketchable yet for every $\varepsilon\in(0,\frac12)$, if $\alpha(n)\geqslant 1$ and $f_n:M(n)\to\ell_{1-\varepsilon}$ are such that $d_{M(n)}(x,y)\leqslant\|f_n(x)-f_n(y)\|_{1-\varepsilon}\leqslant\alpha(n)d_{M(n)}(x,y)$ for all $x,y\in M(n)$, then necessarily $\lim_{n\to\infty}\alpha(n)=\infty$.

1. INTRODUCTION

We shall start by recalling the notion of sketchability; it is implicit in seminal work [2] of Alon, Matias and Szegedy, though the formal definition that is described below was put forth by Saks and Sun [38]. This is a crucial and well-studied algorithmic primitive for analyzing massive date sets, with several powerful applications; surveying them here would be needlessly repetitive, so we refer instead to e.g. [17, 3] and the references therein.

Given a set X, a function $K:X\times X\to\mathbb{R}$ is called a nonnegative kernel if $K(x,y)\geqslant 0$ and $K(x,y)=K(y,x)$ for every $x,y\in X$. In what follows, we will be mainly interested in the geometric setting when the kernel $K=d_X$ is in fact a metric on X, but even for that purpose we will also need to consider nonnegative kernels that are not metrics.

Fix $D\geqslant 1$ and $s\in\mathbb{N}$. Say that a nonnegative kernel $K:X\times X\to[0,\infty)$ is (s,D)-sketchable if for every $r>0$ there is a mapping $\mathsf{R}=\mathsf{R}_r:\{0,1\}^s\times\{0,1\}^s\to\{0,1\}$ and a probability distribution over mappings $\mathsf{Sk}=\mathsf{Sk}_r:X\to\{0,1\}^s$ such that

$$\inf_{\substack{x,y\in X\\ K(x,y)\leqslant r}}\mathbf{Prob}\Big[\mathsf{R}\big(\mathsf{Sk}(x),\mathsf{Sk}(y)\big)=0\Big]\geqslant\frac35\qquad\text{and}\qquad\inf_{\substack{x,y\in X\\ K(x,y)>Dr}}\mathbf{Prob}\Big[\mathsf{R}\big(\mathsf{Sk}(x),\mathsf{Sk}(y)\big)=1\Big]\geqslant\frac35.\tag{1}$$

The value $\frac35$ in (1) can be replaced throughout by any constant that is strictly bigger than $\frac12$; we chose to fix an arbitrary value here in order to avoid the need for the notation to indicate dependence on a further parameter. A kernel (or, more formally, a family of kernels) is said to be sketchable if it is (s,D)-sketchable for some $s=O(1)$ and $D=O(1)$.

The way to interpret the above definition is to think of Sk as a randomized method to assign one of the 2^s labels $\{0,1\}^s$ to each point in X, and to think of R as a reconstruction algorithm that takes as input two such labels in $\{0,1\}^s$ and outputs either 0 or 1, which stand for "small" or "large," respectively. The meaning of (1) becomes that for every pair $x,y\in X$, if one applies the reconstruction algorithm to the random labels $\mathsf{Sk}(x)$ and $\mathsf{Sk}(y)$, then with substantially high probability its output is consistent with the value of the kernel $K(x,y)$ at scale r and approximation D, namely the algorithm declares "small" if $K(x,y)$ is at most r, and it declares "large" if $K(x,y)$ is greater than Dr.

Suppose that $\alpha,\beta,\theta>0$ and that $K:X\times X\to[0,\infty)$ and $L:Y\times Y\to[0,\infty)$ are nonnegative kernels on the sets X and Y, respectively. Suppose also that there is $f:Y\to X$ such that $\alpha L(x,y)^\theta\leqslant K(f(x),f(y))\leqslant\beta L(x,y)^\theta$ for all $x,y\in Y$. It follows formally from this assumption and the above definition that if K is (s,D)-sketchable for some $s\in\mathbb{N}$ and $D\geqslant 1$, then L is $(s,(\beta D/\alpha)^{1/\theta})$-sketchable. Such an "embedding approach" to deduce sketchability is used frequently in the literature. As an example of its many consequences, since ℓ_2 is sketchable by the works of Indyk and Motwani [18] and Kushilevitz, Ostrovsky and Rabani [27], so is any metric space of negative type, where we recall that a metric space (X,d) is said to be of negative type (see e.g. [15]) if the metric space (X,ρ) with $\rho=\sqrt{d}$ is isometric to a subset of ℓ_2.

S.K. was supported by NSF CCF-1422159 and the Simons Foundation. A.N. was supported by NSF CCF-1412958, the Packard Foundation and the Simons Foundation. This work was carried out under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank.

1.1. **The Andoni–Krauthgamer–Razenshteyn characterization of sketchable norms.** The following theorem from [3] is a remarkable result of Andoni, Krauthgamer and Razenshteyn (AKR) that characterizes those norms that are sketchable[1] in terms of their geometric embeddability into a classical kernel (which is not a metric).

Theorem 1 (AKR characterization of sketchability). *Fix $s \in \mathbb{N}$ and $D \geqslant 1$. A finite-dimensional normed space $(X, \|\cdot\|_X)$ is (s,D)-sketchable if and only if for any $\varepsilon \in (0,1)$ there exists $\alpha = \alpha(s,D,\varepsilon) > 0$ and an embedding $f : X \to \ell_{1-\varepsilon}$ such that*

$$\forall\, x, y \in X, \qquad \|x-y\|_X \leqslant \|f(x)-f(y)\|_{1-\varepsilon} \leqslant \alpha\|x-y\|_X.$$

Thus, a finite-dimensional normed space is sketchable if and only if it can be realized as a subset of a the classical sequence space $\ell_{1-\varepsilon}$ so that the kernel $\|\cdot\|_{1-\varepsilon}$ reproduces faithfully (namely, up to factor α) all the pairwise distances in X. See [3, Theorem 1.2] for an explicit dependence in Theorem 1 of $\alpha(s,D,\varepsilon)$ on the parameters s, D, ε.

L_p space notation. In Theorem 1 and below, we use the following standard notation for L_p spaces. If $p \in (0,\infty)$ and (Ω,μ) is a measure space, then $L_p(\mu)$ is the set of (equivalence classes up to measure 0 of) measurable functions $\varphi : \Omega \to \mathbb{R}$ with $\int_\Omega |\varphi(\omega)|^p \,\mathrm{d}\mu(\omega) < \infty$. When μ is the counting measure on $\mathbb{N}$, write $L_p(\mu) = \ell_p$. When μ is the counting measure on $\{1,\ldots,n\}$ for some $n \in \mathbb{N}$, write $L_p(\mu) = \ell_p^n$. When μ is the Lebesgue measure on $[0,1]$, write $L_p(\mu) = L_p$. When the underlying measure is clear from the context (e.g. counting measure or Lebesgue measure), one sometimes writes $L_p(\mu) = L_p(\Omega)$. The $L_p(\mu)$ (quasi)norm is defined by setting $\|\varphi\|_p^p = \int_\Omega |\varphi(\omega)|^p \,\mathrm{d}\mu(\omega)$ for $\varphi \in L_p(\mu)$. While if $p \geqslant 1$, then $(\varphi,\psi) \mapsto \|\varphi-\psi\|_p$ is a metric on $L_p(\mu)$, if $p = 1-\varepsilon$ for some $\varepsilon \in (0,1)$, then $\|\cdot\|_{1-\varepsilon}$ is not a metric; if $L_{1-\varepsilon}(\mu)$ is infinite dimensional, then $\|\cdot\|_{1-\varepsilon}$ is not even equivalent to a metric in the sense that there do not exist any $c, C \in (0,\infty)$ and a metric $d : L_{1-\varepsilon}(\mu) \times L_{1-\varepsilon}(\mu) \to [0,\infty)$ such that $cd(\varphi,\psi) \leqslant \|\varphi-\psi\|_{1-\varepsilon} \leqslant Cd(\varphi,\psi)$ for all $\varphi,\psi \in L_p(\mu)$. Nevertheless, $\|\cdot\|_{1-\varepsilon}$ is a nonnegative kernel on $L_p(\mu)$ and there is a canonical metric $\mathfrak{d}_{1-\varepsilon}$ on $L_p(\mu)$, which is given by

$$\forall\, \varphi, \psi \in L_{1-\varepsilon}(\mu), \qquad \mathfrak{d}_{1-\varepsilon}(\varphi,\psi) \stackrel{\mathrm{def}}{=} \|\varphi-\psi\|_{1-\varepsilon}^{1-\varepsilon} = \int_\Omega |\varphi(\omega)-\psi(\omega)|^{1-\varepsilon} \,\mathrm{d}\mu(\omega). \tag{2}$$

See the books [30, 31] and [21] for much more on the structure for $L_p(\mu)$ spaces when $p \geqslant 1$ and $0 < p < 1$, respectively.

1.1.1. *Beyond norms?* Fix $\varepsilon \in (0,1)$. The sketchability of the nonnegative kernel on $\ell_{1-\varepsilon}$ that is given by $\|\varphi-\psi\|_{1-\varepsilon}$ for $\varphi,\psi \in \ell_{1-\varepsilon}$ was proved by Indyk [17] (formally, using the above terminology it is sketchable provided ε is bounded away from 0; when $\varepsilon \to 0$ the space $s = s(\varepsilon)$ of Indyk's algorithm becomes unbounded). Thus, any metric space (M, d_M) for which there exists $\alpha \in [1,\infty)$ and an embedding $f : M \to \ell_{1-\varepsilon}$ that satisfies

$$\forall\, x, y \in M, \qquad d_M(x,y) \leqslant \|f(x)-f(y)\|_{1-\varepsilon} \leqslant \alpha d_M(x,y) \tag{3}$$

is sketchable with sketch size $O_\varepsilon(1)$ and approximation $O(\alpha)$. Therefore, the "if part" of Theorem 1 holds for any metric space whatsoever, not only for norms. The "only if" part of Theorem 1, namely showing that the mere availability of a sketching algorithm for a normed space implies that it can be realized faithfully as a subset of $\ell_{1-\varepsilon}$, is the main result of [3]. This major achievement demonstrates that a fundamental algorithmic primitive *coincides* with a geometric/analytic property that has been studied long before sketchability was introduced (other phenomena of this nature were discovered in the literature, but they are rare). The underlying reason for Theorem 1 is deep, as the proof in [3] relies on a combination of major results from the literature on functional analysis and communication complexity.

A natural question that Theorem 1 leaves open is whether one could obtain the same result for $\varepsilon = 0$, namely for embeddings into ℓ_1. As discussed in [3], this is equivalent to an old question [28] of Kwapień; a positive result in this direction (for a certain class of norms) is derived in [3] using classical partial progress of Kalton [20] on Kwapień's problem, but fully answering this longstanding question seems difficult (and it may very well have a negative answer).

Another natural question that Theorem 1 leaves open is whether its assumption that the underlying metric space is a norm is needed. Given that the "if part" of Theorem 1 holds for any metric space, this amounts to understanding whether a sketchable metric space (M, d_M) admits for every $\varepsilon \in (0,1)$ an embedding $f : M \to \ell_{1-\varepsilon}$ that satisfies (3). This was a central open question of [3]. Theorem 2 below resolves this question. It should be noted that the authors of [3] formulated their question while hinting that they suspect that the answer is negative, namely in [3, page 893] they wrote *"we are not aware of any counter-example to the generalization of Theorem 1.2 to general metrics"* (Theorem 1.2 in [3] corresponds to Theorem 1 here). One could therefore view Theorem 2 as a confirmation of a prediction of [3].

---

[1] In [3], the conclusion of Theorem 1 is proven under a formally weaker assumption, namely it uses a less stringent notion of sketchability which allows for the random sketches of the points $x, y \in X$ to be different from each other, and for the reconstruction algorithm to depend on the underlying randomness that was used to produce those sketches. Since our main result, namely Theorem 2 below, is an impossibility statement, it becomes only stronger if we use the simpler and stronger notion of sketchability that we stated above.

Theorem 2 (failure of the AKR characterization for general metrics)**.** *For arbitrarily large $n \in \mathbb{N}$ there exists an n-point metric space $(M(n), d_{M(n)})$ which is $\big(O(1), O(1)\big)$-sketchable, yet for every $\varepsilon \in \big(0, \frac{1}{2}\big)$ and $\alpha \geqslant 1$, if there were a mapping $f : M(n) \to \ell_{1-\varepsilon}$ that satisfies $d_{M(n)}(x,y) \leqslant \|f(x) - f(y)\|_{1-\varepsilon} \leqslant \alpha d_{M(n)}(x,y)$ for all $x, y \in M(n)$, then necessarily*

$$\alpha \gtrsim (\log\log n)^{\frac{1-2\varepsilon}{2(1-\varepsilon)}}. \tag{4}$$

Asymptotic notation. In addition to the usual "$O(\cdot), o(\cdot), \Omega(\cdot), \Theta(\cdot)$" notation, it will be convenient to use throughout this article (as we already did in (4)) the following (also standard) asymptotic notation. Given two quantities $Q, Q' > 0$, the notations $Q \lesssim Q'$ and $Q' \gtrsim Q$ mean that $Q \leqslant CQ'$ for some universal constant $C > 0$. The notation $Q \asymp Q'$ stands for $(Q \lesssim Q') \wedge (Q' \lesssim Q)$. If we need to allow for dependence on parameters, we indicate this by subscripts. For example, in the presence of auxiliary objects (e.g. numbers or spaces) $\phi, \mathfrak{Z}$, the notation $Q \lesssim_{\phi,\mathfrak{Z}} Q'$ means that $Q \leqslant C(\phi, \mathfrak{Z})Q'$, where $C(\phi, \mathfrak{Z}) > 0$ is allowed to depend only on $\phi, \mathfrak{Z}$; similarly for the notations $Q \gtrsim_{\phi,\mathfrak{Z}} Q'$ and $Q \asymp_{\phi,\mathfrak{Z}} Q'$.

We will see that the metric spaces $\{(M(n), d_{M(n)})\}_{n=1}^{\infty}$ of Theorem 2 are of negative type, so by the above discussion their sketchability follows from the sketchability of Hilbert space [18, 27]. In fact, these metric spaces are (subsets of) the metric spaces of negative type that were considered by Devanur, Khot, Saket and Vishnoi in [14] as integrality gap examples for the Goemans–Linial semidefinite relaxation of the Sparsest Cut problem with uniform demands. Hence, our contribution is the geometric aspect of Theorem 2, namely demonstrating the non-embeddability into $\ell_{1-\varepsilon}$, rather than its algorithmic component (sketchability). This is a special case of the more general geometric phenomenon of Theorem 7 below, which is our main result. It amounts to strengthening our work [23] which investigated the ℓ_1 non-embeddability of quotients of metric spaces using Fourier-analytic techniques. Here, we derive the (formally stronger) non-embeddability into ℓ_1 of snowflakes of such quotients (the relevant terminology is recalled in Section 1.2 below). It suffices to mention at this juncture (with further discussion in Section 1.2.4 below) that on a conceptual level, the strategy of [23] (as well as that of [26, 14]) for proving non-embeddability using the classical theorem [19] of Kahn, Kalai and Linial (KKL) on influences of variables does not imply the required ℓ_1 non-embeddability of snowflakes of quotients. Instead, we revisit the use of Bourgain's noise sensitivity theorem [7], which was applied for other (non-embeddability) purposes in [24, 23], but subsequent work [26, 14] realized that one could use the much simpler KKL theorem in those contexts (even yielding quantitative improvements). Thus, prior to the present work it seemed that, after all, Bourgain's theorem does not have a decisive use in metric embedding theory, but here we see that in fact it has a qualitative advantage over the KKL theorem in some geometric applications.

The present work also shows that the Khot–Vishnoi approach [24] to the Sparsest Cut integrality gap has a further qualitative advantage (beyond its relevance to the case of uniform demands) over the use of the Heisenberg group for this purpose [29], which yields a better [12] (essentially sharp [35]) lower bound. Indeed, the Heisenberg group is a $O(1)$-doubling metric space (see e.g. [16]), and by Assouad's embedding theorem [5] any such space admits for any $\varepsilon \in (0,1)$ an embedding into $\ell_{1-\varepsilon}$ which satisfies (3) with $\alpha \lesssim_{\varepsilon} 1$ (for the connection to Assouad's theorem, which may not be apparent at this point, see Fact 6 below). Thus, despite its quantitative superiority as an integrality gap example for Sparsest Cut with general demands, the Heisenberg group cannot yield Theorem 2 while the Khot–Vishnoi spaces do (strictly speaking, we work here with a simpler different construction than that of [24], but an inspection of the ensuing proof reveals that one could have also used the metric spaces of [24] to answer the question of [3]).

Question 3. The obvious question that is left open by Theorem 2 is to understand what happens when $\varepsilon \in \big[\frac{1}{2}, 1\big)$. While we established a marked qualitative gap vis à vis sketchability between the behaviors of general normed spaces and general metric spaces, the possibility remains that there exists some $\varepsilon_0 \in \big[\frac{1}{2}, 1\big)$ such that any sketchable metric space (M, d_M) admits an embedding into $\ell_{1-\varepsilon_0}$ that satisfies (3) with $\alpha = O(1)$; perhaps one could even take $\varepsilon_0 = \frac{1}{2}$ here. This possibility is of course tantalizing, as it would be a complete characterization of sketchable metric spaces that is nevertheless qualitatively different from its counterpart for general normed spaces. At present, there is insufficient evidence to speculate that this is so, and it seems more likely that other counterexamples could yield a statement that is analogous to Theorem 2 also in the range $\varepsilon \in \big[\frac{1}{2}, 1\big)$, though a new idea would be needed for that.

Question 4. Even in the range $\varepsilon \in \big(0, \frac{1}{2}\big)$ of Theorem 2, it would be interesting to determine if one could improve (4) to $\alpha \gtrsim (\log n)^{c(\varepsilon)}$ for some $c(\varepsilon) > 0$ (see Remark 5 below for a technical enhancement that yields an asymptotic improvement of (4) but does not achieve such a bound). For the corresponding question when $\varepsilon = 0$, namely embeddings into ℓ_1, it follows from [35] that one could improve (4) to $\alpha \gtrsim \sqrt{\log n}$. However, the example that exhibits this stronger lower bound for $\varepsilon = 0$ is a doubling metric space, and hence by Assouad's theorem [5] for every $\varepsilon > 0$ it does admit an embedding into $\ell_{1-\varepsilon}$ that satisfies (4) with $\alpha \lesssim_{\varepsilon} 1$. Note that by [34, 4] we see that if an n-point metric space (M, d_M) is sketchable for the reason that for some $\theta \in (0,1]$ the metric space (M, d_M^{θ}) is bi-Lipschitz to a subset of ℓ_2, then (4) holds for $\varepsilon = 0$ and $\alpha \lesssim (\log n)^{1/2+o(1)}$. It would be worthwhile to determine if this upper bound on α (for $\varepsilon = 0$) holds for any sketchable metric space whatsoever, i.e., not only for those whose sketchability is due to the fact that some power of the metric is Hilbertian. It seems plausible that the latter question is accessible using available methods.

Remark 5. The lower bound (4) can be improved by incorporating the "enhanced short code argument" of Kane and Meka [22] (which is in essence a derandomization step) into the ensuing reasoning. This yields a more complicated construction for which (4) can be improved to $\alpha \geqslant \exp\big(c(1-2\varepsilon)\sqrt{\log\log n}\big)$ for some universal constant $c > 0$. Because it becomes a significantly more intricate case-specific argument that does not pertain to the more general geometric phenomenon that we study in Theorem 7, we will not include the technical details of this quantitative enhancement of Theorem 2 in the present extended abstract (the full version will contain more information).

1.2. **Metric embeddings.** The distortion of a metric space (U, d_U) in a metric space (V, d_V) is a numerical invariant that is denoted $\mathsf{c}_{(V,d_V)}(U, d_U)$ and defined to be the infimum over those $\alpha \in [1,\infty]$ for which there exist an embedding $f : U \to V$ and a scaling factor $\lambda \in (0,\infty)$ such that $\lambda d_U(x, y) \leqslant d_V\big(f(x), f(y)\big) \leqslant \alpha\lambda d_U(x, y)$ for all distinct $x, y \in U$. Given $p \geqslant 1$, the infimum of $\mathsf{c}_{(V,d_V)}(U, d_U)$ over all possible[2] $L_p(\mu)$ spaces (V, d_V) is denoted $\mathsf{c}_p(U, d_U)$.

1.2.1. *Snowflakes.* Because for every $\varepsilon \in (0,1)$ the quasi-norm $\|\cdot\|_{1-\varepsilon}$ does not induce a metric on $\ell_{1-\varepsilon}$, the embedding requirement (3) does not fit into the above standard metric embedding framework. However, as we explain in Fact 6 below, it is possible to situate (3) within this framework (even without mentioning $\ell_{1-\varepsilon}$ at all) by considering embeddings of the $(1-\varepsilon)$-snowflake of a finite metric space into ℓ_1. Recall the commonly used terminology (see e.g. [13]) that the $(1-\varepsilon)$-snowflake of a metric space (M, d_M) is the metric space $(M, d_M^{1-\varepsilon})$.

Fact 6. *Let (M, d_M) be a finite[3] metric space and fix $\varepsilon \in (0,1)$. The quantity $\mathsf{c}_1(M, d_M^{1-\varepsilon})^{\frac{1}{1-\varepsilon}}$ is equal to the infimum over those $\alpha \geqslant 1$ for which there exists an embedding $f : M \to \ell_{1-\varepsilon}$ that satisfies* (3).

Proof. Suppose that $f : M \to \ell_{1-\varepsilon}$ satisfies (3). Then, recalling the notation (2) for the metric $\mathfrak{d}_{1-\varepsilon}$ on $\ell_{1-\varepsilon}$, we have $d_M(x, y)^{1-\varepsilon} \leqslant \mathfrak{d}_{1-\varepsilon}(f(x), f(y)) \leqslant \alpha^{1-\varepsilon} d_M(x, y)^{1-\varepsilon}$ for all $x, y \in M$. It follows from general principles [9, 39] that the metric space $(\ell_{1-\varepsilon}, \mathfrak{d}_{1-\varepsilon})$ admits an isometric embedding into an $L_1(\mu)$ space (an explicit formula for such an embedding into $L_1(\mathbb{R}^2)$ can be found in [32, Remark 5.10]). Hence, $\mathsf{c}_1(M, d_M^{1-\varepsilon}) \leqslant \alpha^{1-\varepsilon}$. Conversely, there is an explicit embedding (see equation (2) in [33]) $T : \ell_1 \to L_{1-\varepsilon}(\mathbb{N}\times\mathbb{R})$ which is an isometry when one takes the metric $\mathfrak{d}_{1-\varepsilon}$ on $L_{1-\varepsilon}(\mathbb{N}\times\mathbb{R})$. Hence, if $\beta > \mathsf{c}_1(M, d_M^{1-\varepsilon})$, then take an embedding $g : M \to \ell_1$ such that $d_M(x, y)^{1-\varepsilon} \leqslant \|g(x) - g(y)\|_1 \leqslant \beta d_M(x, y)^{1-\varepsilon}$ for all $x, y \in M$ and consider the embedding $T \circ g$ which satisfies (3) with $\alpha = \beta^{1/(1-\varepsilon)}$, except that the target space is $L_{1-\varepsilon}(\mathbb{N}\times\mathbb{R})$ rather than $\ell_{1-\varepsilon}$. By an approximation by simple functions we obtain the desired embedding into $\ell_{1-\varepsilon}$. $\square$

1.2.2. *Quotients.* Suppose that G is a group that acts on a metric space (X, d_X) by isometries. The quotient space $X/G = \{Gx\}_{x\in X}$ of all the orbits of G can be equipped with the following quotient metric $d_{X/G} : (X/G) \times (X/G) \to [0,\infty)$.

$$\forall\, x, y \in X, \qquad d_{G/X}(Gx, Gy) \stackrel{\mathrm{def}}{=} \inf_{(u,v)\in(Gx)\times(Gy)} d_X(u, v) = \inf_{g\in G} d_X(gx, y). \tag{5}$$

See [10, Section 5.19] for more on this basic construction (in particular, for a verification that (5) indeed gives a metric).

Given $k \in \mathbb{N}$, we will consider the Hamming cube to be the vector space $\mathbb{F}_2^k$ over the field of two elements $\mathbb{F}_2$, equipped with the Hamming metric $d_{\mathbb{F}_2^k} : \mathbb{F}_2^k \times \mathbb{F}_2^k \to \mathbb{N}\cup\{0\}$ that is given by

$$\forall\, x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k) \in \mathbb{F}_2^k, \qquad d_{\mathbb{F}_2^k}(x, y) = |\{j \in \{1, \ldots, k\} :\ x_j \neq y_j\}|.$$

Below, $\mathbb{F}_2^k$ will always be assumed to be equipped with the metric $d_{\mathbb{F}_2^k}$. The standard basis of $\mathbb{F}_2^k$ is denoted $e_1, \ldots, e_k$.

If G is a group acting on $\mathbb{F}_2^k$ by isometries, and if it isn't too large, say, $|G| \leqslant 2^{k/2}$, then all but an exponentially small fraction of the pairs $(x, y) \in \mathbb{F}_2^k \times \mathbb{F}_2^k$ satisfy $d_{\mathbb{F}_2^k}(Gx, Gy) \gtrsim k$. Specifically, there is a universal constant $\eta > 0$ such that

$$|G| \leqslant 2^{\frac{k}{2}} \implies \left|\left\{(x, y) \in \mathbb{F}_2^k \times \mathbb{F}_2^k :\ d_{\mathbb{F}_2^k/G}(x, y) \leqslant \eta k\right\}\right| \leqslant 2^{\frac{5}{3}k}. \tag{6}$$

A simple counting argument which verifies (6) appears in the proof of [23, Lemma 3.2].

---

[2] When (U, d_U) is a finite metric space, it suffices to consider embeddings into ℓ_p rather than a general $L_p(\mu)$ space, as follows via a straightforward approximation by simple functions. We warn that this is not so for general (infinite) separable metric spaces, in which case one must consider embeddings into L_p; by [11, Corollary 1.5] there is even a doubling subset of L_1 that does not admit a bi-Lipschitz embedding into ℓ_1.

[3] The only reason for the finiteness assumption here (the present article deals only with finite metric space) is to ensure that the embedding is into $\ell_{1-\varepsilon}$ rather than a more general $L_{1-\varepsilon}(\mu)$ space. For embeddings of finite-dimensional normed spaces, i.e., the setting of [3], a similar reduction to embeddings into $\ell_{1-\varepsilon}$ is possible using tools from [36, 1, 6].

The symmetric group S_k acts isometrically on $\mathbb{F}_2^k$ by permuting the coordinates, namely for each permutation g of $\{1,\ldots,k\}$ and $x\in\mathbb{F}_2^k$ we write $gx=(x_{g^{-1}(1)},x_{g^{-1}(2)},\ldots,x_{g^{-1}(k)})$. A subgroup $G\leqslant S_k$ of S_k therefore acts by isometries on $\mathbb{F}_2^k$; below we will only consider quotients of the form $(\mathbb{F}_2^k/G,d_{\mathbb{F}_2^K/G})$ when G is a transitive subgroup of S_k.

1.2.3. *ℓ_1 non-embeddability of snowflakes of (subsets of) hypercube quotients.* In [23] we studied the ℓ_1 embeddability of quotients of $\mathbb{F}_2^k$. In particular, [23, Corollary 3] states that if G is a transitive subgroup of S_k with $|G|\leqslant 2^{k/2}$, then

$$\mathsf{c}_1\big(\mathbb{F}_2^k/G,d_{\mathbb{F}_2^k/G}\big)\gtrsim \log k. \tag{7}$$

In Remark 4 of [23] we (implicitly) asked about the sketchability of $\mathbb{F}_2^k/G$, by inquiring whether its $(1/2)$-snowflake embeds into a Hilbert space with $O(1)$ distortion, as a possible alternative approach for obtaining integrality gaps (quantitatively stronger than what was known at the time) for the Goemans–Linial semidefinite relaxation of the Sparsest Cut problem. This hope was realized in [14] for the special case when $G=\langle\mathfrak{S}_k\rangle\leqslant S_k$ is the cyclic group that is generated by the cyclic shift $\mathfrak{S}_k=(1,2,\ldots,k)\in S_k$. Specifically, it follows from [14] that there exists a large subset $M\subseteq\mathbb{F}_2^k$, namely $|\mathbb{F}_2^k\smallsetminus M|\lesssim 2^k/k^2$, and a metric ρ on $M/\langle\mathfrak{S}_k\rangle$ satisfying $\rho(\mathcal{O},\mathcal{O}')\asymp d_{\mathbb{F}_2^k/\langle\mathfrak{S}_k\rangle}(\mathcal{O},\mathcal{O}')$ for all pairs of orbits $\mathcal{O},\mathcal{O}'\in M/\langle\mathfrak{S}_k\rangle$, and such that the metric space $(M/\langle\mathfrak{S}_k\rangle,\sqrt{\rho})$ embeds isometrically into ℓ_2. Strictly speaking, a stronger statement than this was obtained in [14] for a larger metric space (namely, for the quotient of $\mathbb{F}_2^k\times\mathbb{F}_2^k$ by the group $\langle\mathfrak{S}_k\rangle\times\langle\mathfrak{S}_k\rangle$), but here it suffices to consider the above smaller metric space which inherits the stated properties.

Recalling Fact 6, this discussion leads naturally, as a strategy towards proving Theorem 2, to investigating whether a lower bound as (7) holds for the $(1-\varepsilon)$-snowflake of the hypercube quotient $\mathbb{F}_2^k/G$ rather than that quotient itself. We will see that the method of [23] does not yield any such lower bound that tends to ∞ as $k\to\infty$ for fixed $\varepsilon>0$, but we do obtain the desired statement here, albeit with an asymptotically weaker lower bound than the $\log k$ of (7). Note that an application of Theorem 7 below to the above subset $M\subseteq\mathbb{F}_2^k$ from [14] yields Theorem 2, because of Fact 6.

Theorem 7 (non-embeddability of snowflakes of quotients of large subsets of the hypercube)**.** *Fix $k\in\mathbb{N}$ and $\varepsilon\in(0,\frac12)$. Let G be a transitive subgroup of S_k with $|G|\leqslant 2^{k/2}$. Then, every $M\subseteq\mathbb{F}_2^k$ with $|\mathbb{F}_2^k\smallsetminus M|\leqslant 2^k/\sqrt{\log k}$ satisfies*

$$\mathsf{c}_1\Big(M/G,d_{\mathbb{F}_2^k/G}^{1-\varepsilon}\Big)\gtrsim(\log k)^{\frac12-\varepsilon}. \tag{8}$$

It would be interesting to determine the asymptotically sharp behavior (up to universal constant factors) in (8) for $M=\mathbb{F}_2^k$, though understanding the dependence on the transitive subgroup $G\leqslant S_k$ may be challenging; see [8] for investigations along these lines. Even in the special case $G=\langle\mathfrak{S}_k\rangle$ we do not know the sharp bound, and in particular how it transitions from the $(\log k)^{1/2-\varepsilon}$ of (8) to the $\log k$ of (7) as $\varepsilon\to0$ (it could be that neither bound is tight).

1.2.4. *Bourgain's Fourier tails versus the Kahn–Kalai–Linial influence of variables.* In [23, Theorem 3.8] we applied the important theorem [19] of Kahn, Kalai and Linial on the influence of variables on Boolean functions to show that if G is a transitive subgroup of S_k, then every $f:\mathbb{F}_2^k/G\to\ell_1$ satisfies the following Cheeger/Poincaré inequality.

$$\frac{1}{4^k}\sum_{(x,y)\in\mathbb{F}_2^k\times\mathbb{F}_2^k}\|f(Gx)-f(Gy)\|_1\lesssim\frac{1}{\log k}\sum_{j=1}^k\frac{1}{2^k}\sum_{x\in\mathbb{F}_2^k}\big\|f\big(G(x+e_j)\big)-f(Gx)\big\|_1. \tag{9}$$

Fix $(\varepsilon,\alpha)\in(0,1)\times[1,\infty)$. If $|G|\leqslant 2^{k/2}$ and $d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}\leqslant\|f(Gx)-f(Gy)\|_1\leqslant\alpha d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}$ for $x,y\in\mathbb{F}_2^k$, then

$$\begin{aligned}k^{1-\varepsilon}&\overset{(6)}{\lesssim}\frac{1}{4^k}\sum_{(x,y)\in\mathbb{F}_2^k\times\mathbb{F}_2^k}d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}\leqslant\frac{1}{4^k}\sum_{(x,y)\in\mathbb{F}_2^k\times\mathbb{F}_2^k}\|f(Gx)-f(Gy)\|_1\overset{(9)}{\lesssim}\frac{1}{\log k}\sum_{j=1}^k\frac{1}{2^k}\sum_{x\in\mathbb{F}_2^k}\big\|f\big(G(x+e_j)\big)-f(Gx)\big\|_1\\&\leqslant\frac{\alpha}{\log k}\sum_{j=1}^k\frac{1}{2^k}\sum_{x\in\mathbb{F}_2^k}d_{\mathbb{F}_2^k/G}\big(G(x+e_j),Gy\big)^{1-\varepsilon}\overset{(5)}{\leqslant}\frac{\alpha}{\log k}\sum_{j=1}^k\frac{1}{2^k}\sum_{x\in\mathbb{F}_2^k}d_{\mathbb{F}_2^k}(x+e_j,y)^{1-\varepsilon}=\frac{\alpha k}{\log k}.\end{aligned}$$

It follows that

$$\mathsf{c}_1\Big(\mathbb{F}_2^k/G,d_{\mathbb{F}_2^k/G}^{1-\varepsilon}\Big)\gtrsim\frac{\log k}{k^{\varepsilon}}. \tag{10}$$

This is how (7) was derived in [23], but the right hand side of (10) tends to ∞ as $k\to\infty$ only if $\varepsilon\lesssim(\log\log k)/\log k$.

Following the above use of the KKL theorem [23], it was used elsewhere in place of applications [24, 23] of a more substantial theorem of Bourgain [7] on the Fourier tails of Boolean functions that are not close to juntas; notably this was first done by Krauthgamer and Rabani [26] to obtain an asymptotically improved analysis of the Khot–Vishnoi integrality gap [24] for Sparsest Cut. We have seen above that the KKL-based approach does not yield Theorem 7 (though, of course, one cannot rule out the availability of a more sophisticated application of KKL that does), but our use of Bourgain's theorem in the ensuing proof of Theorem 7 shows that this theorem does sometime provide qualitatively stronger geometric information. One should note here that (8) follows from an application of a sharp form of Bourgain's theorem that was more recently obtained by Kindler, Kirshner, and O'Donnell [25]; an application of Bourgain's original formulation yields a bound that is asymptotically weaker by a lower-order factor.

2. PROOF OF THEOREM 7

Here we will prove Theorem 7, thereby completing the justification of Theorem 2 as well.

2.1. **Fourier-analytic preliminaries.** We will include here some basic facts and notation related to Fourier analysis on the hypercube $\mathbb{F}_2^k$; an extensive treatment of this topic can be found in e.g. the monograph [37]. Fix $k\in\mathbb{N}$. From now on, let $\mu=\mu_k$ denote the normalized counting measure on $\mathbb{F}_2^k$. Given $A\subseteq\{1,\ldots,k\}$, the Walsh function $\mathsf{W}_A:\mathbb{F}_2^k\to\{-1,1\}$ and Fourier coefficient $\widehat{\varphi}(A)\in\mathbb{R}$ of a function $\varphi:\mathbb{F}_2^k\to\mathbb{R}$ are defined by

$$\forall\, x\in\mathbb{F}_2^k,\qquad \mathsf{W}_A(x)=(-1)^{\sum_{j=1}^n x_j}\qquad\text{and}\qquad \widehat{\varphi}(A)=\int_{\mathbb{F}_2^k}\varphi(x)\mathsf{W}_A(x)\,\mathrm{d}\mu(x).$$

The convolution $\varphi*\psi:\mathbb{F}_2^k\to\mathbb{R}$ of two functions $\varphi,\psi:\mathbb{F}_2^k\to\mathbb{R}$ is defined by

$$\forall\, x\in\mathbb{F}_2^k,\qquad (\varphi*\psi)(x)=\int_{\mathbb{F}_2^k}\varphi(y)\psi(x+y)\,\mathrm{d}\mu(y)=\sum_{A\subseteq\{1,\ldots,k\}}\widehat{\varphi}(A)\widehat{\psi}(A)\mathsf{W}_A(x),$$

where the last equality is valid because the 2^k Walsh functions $\{\mathsf{W}_A\}_{A\subseteq\{1,\ldots,k\}}$ consist of all of the characters of the additive group $\mathbb{F}_2^k$, hence forming an orthonormal basis of $L_2(\mu)$. Suppose that $g\in\mathsf{GL}(\mathbb{F}_2^k)$ is an automorphism of $\mathbb{F}_2^k$. If $\varphi:\mathbb{F}_2^k\to\mathbb{R}$ is a g-invariant function, i.e., $\varphi(gy)=\varphi(y)$ for all $y\in\mathbb{F}_2^k$, then for every $\psi:\mathbb{F}_2^k\to\mathbb{R}$ and $x\in\mathbb{F}_2^k$,

$$\begin{aligned}(\varphi*\psi)(x)&=\int_{\mathbb{F}_2^k}\varphi(y)\psi(x+y)\,\mathrm{d}\mu(y)=\int_{\mathbb{F}_2^k}\varphi(gy)\psi(x+y)\,\mathrm{d}\mu(y)\\&=\int_{\mathbb{F}_2^k}\varphi(z)\psi\big(x+g^{-1}z\big)\,\mathrm{d}\mu(z)=\int_{\mathbb{F}_2^k}\varphi(z)\psi\big(g^{-1}(gx+z)\big)\,\mathrm{d}\mu(z)=\Big(\varphi*\big(\psi\circ g^{-1}\big)\Big)(gx).\end{aligned}$$

In particular, under the above invariance assumption we have the identity

$$\|\varphi*\psi\|_{L_2(\mu)}=\Big\|\varphi*\big(\psi\circ g^{-1}\big)\Big\|_{L_2(\mu)}.\tag{11}$$

Given $\mathfrak{p}\in[0,1]$, let $\vartheta^{\mathfrak{p}}:2^{\mathbb{F}_2^k\times\mathbb{F}_2^k}\to[0,1]$ be the probability measure that is defined by setting for each $(x,y)\in\mathbb{F}_2^k\times\mathbb{F}_2^k$,

$$\vartheta^{\mathfrak{p}}(x,y)\stackrel{\mathrm{def}}{=}\frac{\mathfrak{p}^{d_{\mathbb{F}_2^k}(x,y)}(1-\mathfrak{p})^{k-d_{\mathbb{F}_2^k}(x,y)}}{2^k}=\frac{1}{4^k}\prod_{j=1}^k\big(1+(1-2\mathfrak{p})(-1)^{x_j+y_j}\big)=\frac{1}{4^k}\sum_{A\subseteq\{1,\ldots,k\}}(1-2\mathfrak{p})^{|A|}\mathsf{W}_A(x+y).\tag{12}$$

In other words, $\vartheta^{\mathfrak{p}}(x,y)$ is equal to the probability that the ordered pair (x,y) is the outcome of the following randomized selection procedure: The first element $x\in\mathbb{F}_2^k$ is chosen uniformly at random, and the second element $y\in\mathbb{F}_2^k$ is obtained by changing the sign of each entry of x independently with probability $\mathfrak{p}$. Note in passing that both marginals of $\vartheta^{\mathfrak{p}}$ are equal to μ, i.e., $\vartheta^{\mathfrak{p}}(\Omega\times\mathbb{F}_2^k)=\vartheta^{\mathfrak{p}}(\mathbb{F}_2^k\times\Omega)=\mu(\Omega)$ for every $\Omega\subseteq\mathbb{F}_2^k$. Also, for every $\Omega\subseteq\mathbb{F}_2^k$ we have

$$\begin{aligned}\vartheta^{\mathfrak{p}}\big(\Omega\times(\mathbb{F}_2^k\smallsetminus\Omega)\big)&=\frac18\int_{\mathbb{F}_2^k\times\mathbb{F}_2^k}\Big((-1)^{\mathbf{1}_\Omega(x)}-(-1)^{\mathbf{1}_\Omega(y)}\Big)^2\,\mathrm{d}\vartheta^{\mathfrak{p}}(x,y)\\&=\frac14\bigg(1-\int_{\mathbb{F}_2^k\times\mathbb{F}_2^k}(-1)^{\mathbf{1}_\Omega}(x)(-1)^{\mathbf{1}_\Omega}(y)\,\mathrm{d}\vartheta^{\mathfrak{p}}(x,y)\bigg)=\frac14\sum_{A\subseteq\{1,\ldots,k\}}\Big(1-(1-2\mathfrak{p})^{|A|}\Big)\Big(\widehat{(-1)^{\mathbf{1}_\Omega}}(A)\Big)^2,\end{aligned}\tag{13}$$

where the last equality in (13) is a direct consequence of Parseval's identity and the final expression in (12) for $\vartheta^{\mathfrak{p}}(\cdot,\cdot)$.

For $\varphi:\mathbb{F}_2^k\to\mathbb{R}$ and $j,m\in\{1,\ldots,k\}$, the level-m influence of the j'th variable on φ, denoted $\mathsf{Inf}_j^{\leqslant m}[\varphi]$, is the quantity

$$\mathsf{Inf}_j^{\leqslant m}[\varphi]=\sum_{\substack{A\subseteq\{1,\ldots,k\}\smallsetminus\{j\}\\|A|\leqslant m-1}}\widehat{\varphi}(A\cup\{j\})^2=\Big\|\varphi*\mathcal{R}_j^{\leqslant m}\Big\|_{L_2(\mu)}^2,\tag{14}$$

where the last equality is a consequence of Parseval's identity, using the notation

$$\mathcal{R}_j^{\leqslant m} \stackrel{\mathrm{def}}{=} \sum_{\substack{A\subseteq\{1,\ldots,k\}\smallsetminus\{j\}\\ |A|\leqslant m-1}} \mathsf{W}_{A\cup\{j\}}. \tag{15}$$

It follows from the first equation in (14) that

$$\sum_{j=1}^{k} \mathsf{Inf}_j^{\leqslant m}[\varphi] = \sum_{\substack{B\subseteq\{1,\ldots,k\}\\ |B|\leqslant m}} |B|\widehat{\varphi}(B)^2 \leqslant m \sum_{\substack{B\subseteq\{1,\ldots,k\}\\ B\neq\varnothing}} \widehat{\varphi}(B)^2 = m\bigg(\int_{\mathbb{F}_2^k} \varphi^2\,\mathrm{d}\mu - \widehat{\varphi}(\varnothing)^2\bigg) = m\mathsf{Var}_\mu[\varphi], \tag{16}$$

where $\mathsf{Var}_\mu[\cdot]$ denotes the variance with respect to the probability measure μ. By considering the symmetric group S_k as a subgroup of $\mathsf{GL}(\mathbb{F}_2^k)$, where the action is permutation of coordinates, an inspection of definition (15) reveals that $\mathcal{R}_j^{\leqslant m}\circ g = \mathcal{R}_{gj}^{\leqslant m}$ for $g\in S_k$ and $j,m\in\{1,\ldots,k\}$. By (11) and the second equality in (14), if $\varphi:\mathbb{F}_2^k\to\mathbb{R}$ is g-invariant, then

$$\forall\, j,m\in\{1,\ldots,k\},\qquad \mathsf{Inf}_j^{\leqslant m}[\varphi] = \mathsf{Inf}_{g^{-1}j}^{\leqslant m}[\varphi].$$

A combination of this observation with (16) yields the following statement, which we record for ease of later reference.

Fact 8. *Fix $k\in\mathbb{N}$. Let G be a subgroup of S_k that acts transitively on the coordinates $\{1,\ldots,k\}$. Suppose that $\varphi:\mathbb{F}_2^k\to\mathbb{R}$ is a G-invariant function, i.e., $f(gx)=f(x)$ for every $g\in G$ and $x\in\mathbb{F}_2^n$. Then, for every $m\in\{1,\ldots,k\}$ we have*

$$\max_{j\in\{1,\ldots,k\}} \mathsf{Inf}_j^{\leqslant m}[\varphi] \leqslant \frac{m}{k}\mathsf{Var}_\mu[\varphi].$$

Throughout what follows, given a subgroup $G\leqslant S_k$, we denote by $\pi_G:\mathbb{F}_2^k\to\mathbb{F}_2^k/G$ its associated quotient mapping, i.e., $\pi_G(x)=Gx$ for all $x\in\mathbb{F}_2^k$. We denote by $\mu_{\mathbb{F}_2^k/G}=\mu\circ\pi_G^{-1}$ the probability measure on $\mathbb{F}_2^k/G$ that is given by

$$\forall\, \mathcal{O}\in\mathbb{F}_2^k/G,\qquad \mu_{\mathbb{F}_2^k/G}(\mathcal{O})=\mu(\mathcal{O}).$$

In a similar vein, for every $\mathfrak{p}\in[0,1]$ the probability measure $\vartheta^{\mathfrak{p}}$ on $\mathbb{F}_2^k\times\mathbb{F}_2^k$ that is given in (12) descends to a probability measure $\vartheta^{\mathfrak{p}}_{\mathbb{F}_2^k/G}=\vartheta^{\mathfrak{p}}\circ(\pi_G\times\pi_G)^{-1}$ on $(\mathbb{F}_2^k/G)\times(\mathbb{F}_2^k/G)$ by setting

$$\forall\, \mathcal{O},\mathcal{O}'\subseteq\mathbb{F}_2^k/G,\qquad \vartheta^{\mathfrak{p}}_{\mathbb{F}_2^k/G}(\mathcal{O},\mathcal{O}')=\vartheta^{\mathfrak{p}}(\mathcal{O}\times\mathcal{O}').$$

2.2. **A Cheeger/Poincaré inequality for transitive quotients.** Our main technical result is the following inequality.

Lemma 9. *There is a universal constant $\beta\in(0,1)$ with the following property. Fix an integer $k\geqslant 55$ and a transitive subgroup G of S_k. Suppose that $X\subseteq\mathbb{F}_2^k/G$ is a sufficiently large subset in the following sense.*

$$\mu_{\mathbb{F}_2^k/G}(X)\geqslant 1-\frac{1}{\sqrt{\log k}}. \tag{17}$$

Then there is a further subset $Y\subseteq X$ with $\mu_{\mathbb{F}_2^k/G}(Y)\geqslant\frac34\mu_{\mathbb{F}_2^k/G}(X)$ such that every function $f:Y\to\ell_1$ satisfies

$$\iint_{Y\times Y}\|f(\mathcal{O})-f(\mathcal{O}')\|_1\,\mathrm{d}\mu_{\mathbb{F}_2^k/G}(\mathcal{O})\,\mathrm{d}\mu_{\mathbb{F}_2^k/G}(\mathcal{O}')\lesssim\sqrt{\log k}\iint_{Y\times Y}\|f(\mathcal{O})-f(\mathcal{O}')\|_1\,\mathrm{d}\vartheta^{\frac{1}{\beta\log k}}_{\mathbb{F}_2^n/G}(\mathcal{O},\mathcal{O}'). \tag{18}$$

Prior to proving Lemma 9 we shall assume its validity for the moment and proceed to prove Theorem 7.

Proof of Theorem 7 assuming Lemma 9. Fix $\alpha\geqslant 1$ and suppose that $f:M/G\to\ell_1$ satisfies

$$\forall\, x,y\in M,\qquad d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}\leqslant\|f(Gx)-f(Gy)\|_1\leqslant\alpha d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}. \tag{19}$$

Our task is to bound α from below by the right hand side of (8).

An application of Lemma 9 to $X=M/\mathbb{F}_2$, which satisfies the requirement (17) by the assumption of Theorem 7, produces a subset Y with $\mu(\pi_G^{-1}(Y))\geqslant\frac12$ for which (18) holds true. It follows that

$$\begin{aligned}\iint_{\pi_G^{-1}(Y)\times\pi_G^{-1}(Y)} d_{\mathbb{F}_2^k/G}(Gx,Gy)^{1-\varepsilon}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) &\overset{(18)\wedge(19)}{\lesssim} \alpha\sqrt{\log k}\iint_{(\mathbb{F}_2^k/G)\times(\mathbb{F}_2^k/G)} d_{\mathbb{F}_2^k/G}(\mathcal{O},\mathcal{O}')^{1-\varepsilon}\,\mathrm{d}\vartheta^{\frac{1}{\beta\log k}}_{\mathbb{F}_2^n/G}(\mathcal{O},\mathcal{O}')\\ &\overset{(5)}{\leqslant}\alpha\sqrt{\log k}\int_{\mathbb{F}_2^k\times\mathbb{F}_2^k} d_{\mathbb{F}_2^k}(x,y)^{1-\varepsilon}\,\mathrm{d}\vartheta^{\frac{1}{\beta\log k}}(x,y)\overset{(12)}{=}\alpha\sqrt{\log k}\sum_{\ell=0}^{k}\ell^{1-\varepsilon}\binom{k}{\ell}\left(\frac{\beta}{\log k}\right)^{\ell}\left(1-\frac{\beta}{\log k}\right)^{k-\ell}\leqslant\alpha\sqrt{\log k}\left(\frac{\beta k}{\log k}\right)^{1-\varepsilon}.\end{aligned} \tag{20}$$

Since $|G| \leqslant 2^{k/2}$, by (6) there exists $\eta \gtrsim 1$ such that, since $\mu(\pi_G^{-1}(Y)) \gtrsim 1$, we have

$$\begin{aligned}\mu \times \mu\Big(\big\{(x,y) \in \pi_G^{-1}(Y) \times \pi_G^{-1}(Y) :\ d_{F_2^k}(Gx,Gy) > \eta k\big\}\Big)& \\ \geqslant \mu\big(\pi_G^{-1}(Y)\big)^2 - \mu \times \mu\Big(\big\{(x,y) \in \mathbb{F}_2^k \times \mathbb{F}_2^k :\ d_{F_2^k}(Gx,Gy) \leqslant \eta k\big\}\Big) &\geqslant \mu\big(\pi_G^{-1}(Y)\big)^2 - 2^{-\frac{k}{3}} \gtrsim 1.\end{aligned}$$

So, the first quantity in (20) is at least a constant multiple of $k^{1-\varepsilon}$, and the desired lower bound on α follows. $\square$

Proof of Lemma 9. Suppose that $Z \subseteq X$ satisfies

$$\frac{1}{4} \leqslant \frac{\mu_{\mathbb{F}^k/G}(Z)}{\mu_{\mathbb{F}^k/G}(X)} \leqslant \frac{2}{3}. \tag{21}$$

Writing $\mathfrak{q} = \mu_{\mathbb{F}^k/G}(Z)$, the function $(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}} : \mathbb{F}_2^k \to \{-1,1\}$ is G-invariant and its variance is equal to $4\mathfrak{q}(1-\mathfrak{q}) \asymp 1$. Let $\beta \in (2/\log k, 1)$ be a small enough universal constant that will be determined later. Also, let $C \in (1,\infty)$ be a large enough universal constant, specifically take C to be the universal constant that appears in the statement of [25, Theorem 3.1]. If we denote $m = \lceil \beta \log k \rceil$, then it follows from Fact 8 that, provided β is a sufficiently small constant, we have

$$\max_{j\in\{1,\ldots,k\}} \mathsf{Inf}_j^{\leqslant m}\Big[(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}\Big] \leqslant \frac{m}{k}\mathsf{Var}\Big[(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}\Big] \leqslant \frac{\mathsf{Var}\Big[(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}\Big]^4}{C^m}.$$

This is precisely the assumption of [25, Theorem 3.1], from which we deduce the following Fourier tail bound.

$$\sum_{\substack{A\subseteq\{1,\ldots,k\}\\ |A|>\lceil\beta\log k\rceil}} \Big(\widehat{(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}}(A)\Big)^2 = \sum_{\substack{A\subseteq\{1,\ldots,k\}\\ |A|>m}} \Big(\widehat{(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}}(A)\Big)^2 \gtrsim \frac{\mathsf{Var}\Big[(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}\Big]}{\sqrt{m}} \asymp \frac{1}{\sqrt{\beta\log k}}. \tag{22}$$

Next, by the identity (13), we have

$$\begin{aligned}\vartheta^{\frac{1}{\beta\log k}}\Big(\pi_G^{-1}(Z) \times \big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(Z)\big)\Big) &= \frac{1}{4}\sum_{A\subseteq\{1,\ldots,k\}} \bigg(1-\Big(1-\frac{2}{\beta\log k}\Big)^{|A|}\bigg)\Big(\widehat{(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}}(A)\Big)^2 \\ &\geqslant \frac{1}{4}\bigg(1-\Big(1-\frac{2}{\beta\log k}\Big)^{\lceil\beta\log k\rceil+1}\bigg) \sum_{\substack{A\subseteq\{1,\ldots,k\}\\ |A|>\lceil\beta\log k\rceil}} \Big(\widehat{(-1)^{\mathbf{1}_{\pi_G^{-1}(Z)}}}(A)\Big)^2 \overset{(22)}{\geqslant} \frac{\gamma}{\sqrt{\beta\log k}},\end{aligned} \tag{23}$$

for some universal constant $\gamma \in (0,1)$. Therefore,

$$\begin{aligned}\vartheta^{\frac{1}{\beta\log k}}\Big(\pi_G^{-1}(Z) \times \big(\pi_G^{-1}(X) \smallsetminus \pi_G^{-1}(Z)\big)\Big) &\geqslant \vartheta^{\frac{1}{\beta\log k}}\Big(\pi_G^{-1}(Z) \times \big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(Z)\big)\Big) - \vartheta^{\frac{1}{\beta\log k}}\Big(\mathbb{F}_2^k \times \big((\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(X)\big)\Big) \\ &= \vartheta^{\frac{1}{\beta\log k}}\Big(\pi_G^{-1}(Z) \times \big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(Z)\big)\Big) - \mu\big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(X)\big) \overset{(17)\wedge(23)}{\geqslant} \frac{\gamma}{\sqrt{\beta\log k}} - \frac{1}{\sqrt{\log k}} \overset{(21)}{\asymp} \frac{1}{\sqrt{\log k}} \cdot \frac{\mu_{\mathbb{F}^k/G}(Z)}{\mu_{\mathbb{F}^k/G}(X)},\end{aligned} \tag{24}$$

where the final step of (24) holds provided $1 \asymp \beta \leqslant \gamma^2/4$, which is our final requirement from the universal constant β.

Observe that

$$\begin{aligned}\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}(X \times X) &\geqslant \vartheta^{\frac{1}{\beta\log k}}\big(\mathbb{F}_2^k \times \mathbb{F}_2^k\big) - \vartheta^{\frac{1}{\beta\log k}}\Big(\big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(X)\big) \times \mathbb{F}_2^k\Big) - \vartheta^{\frac{1}{\beta\log k}}\Big(\mathbb{F}_2^k \times \big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(X)\big)\Big) \\ &= 1 - 2\mu\big(\mathbb{F}_2^k \smallsetminus \pi_G^{-1}(X)\big) = 1 - 2\big(1-\mu_{\mathbb{F}_2^k/G}(X)\big) \overset{(17)}{\geqslant} 1 - \frac{2}{\sqrt{\log k}} \asymp 1.\end{aligned} \tag{25}$$

Hence,

$$\frac{\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}\Big(\big(Z \times (X \smallsetminus Z)\big) \cup \big((X \smallsetminus Z) \times Z\big)\Big)}{\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}(X \times X)} = \frac{2\vartheta^{\frac{1}{\beta\log k}}\Big(\pi_G^{-1}(Z) \times \big(\pi_G^{-1}(X) \smallsetminus \pi_G^{-1}(Z)\big)\Big)}{\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}(X \times X)} \overset{(24)}{\gtrsim} \frac{1}{\sqrt{\log k}} \cdot \frac{\mu_{\mathbb{F}^k/G}(Z)}{\mu_{\mathbb{F}^k/G}(X)}. \tag{26}$$

We are now in position to apply [23, Lemma 6] with the parameters $\delta = \frac{1}{4}$, $\alpha \asymp 1/\sqrt{\log k}$, and the probability measures

$$\sigma \overset{\mathrm{def}}{=} \frac{\mu_{\mathbb{F}_2^k/G}}{\mu_{\mathbb{F}_2^k/G}(X)} : 2^X \to [0,1] \qquad \text{and} \qquad \tau \overset{\mathrm{def}}{=} \frac{\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}}{\vartheta_{\mathbb{F}_2^k/G}^{\frac{1}{\beta\log k}}(X \times X)} : 2^{X\times X} \to [0,1]. \tag{27}$$

Due to (26), by the proof of [23, Lemma 6] (specifically, equation (7) in [23]) there exists a subset $Y \subseteq \mathbb{F}_2^k/G$ with $\sigma(Y) \geqslant 3/4$, i.e., $\mu_{\mathbb{F}_2^k/G}(Y) \geqslant 3\mu_{\mathbb{F}_2^k/G}(X)/4$, such that every $f : Y \to L_1$ satisfies

$$\begin{aligned}\iint_{Y\times Y} \|f(\mathcal{O}) - f(\mathcal{O}')\|_1 \,\mathrm{d}\mu_{\mathbb{F}_2^k/G}(\mathcal{O})\,\mathrm{d}\mu_{\mathbb{F}_2^k/G}(\mathcal{O}') &\overset{(21)\wedge(27)}{\asymp} \iint_{Y\times Y} \|f(\mathcal{O}) - f(\mathcal{O}')\|_1 \,\mathrm{d}\sigma(\mathcal{O})\,\mathrm{d}\sigma(\mathcal{O}')\\ &\lesssim \sqrt{\log k}\iint_{Y\times Y} \|f(\mathcal{O}) - f(\mathcal{O}')\|_1 \,\mathrm{d}\tau(\mathcal{O},\mathcal{O}') \overset{(25)\wedge(27)}{\asymp} \sqrt{\log k}\iint_{Y\times Y} \|f(\mathcal{O}) - f(\mathcal{O}')\|_1 \,\mathrm{d}\vartheta^{\frac{1}{\beta\log k}}_{\mathbb{F}_2^n/G}(\mathcal{O},\mathcal{O}'). \qquad \square\end{aligned}$$

REFERENCES

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

[2] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *J. Comput. System Sci.*, 58(1, part 2):137–147, 1999. Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996).

[3] A. Andoni, R. Krauthgamer, and I. Razenshteyn. Sketching and Embedding are Equivalent for Norms. *SIAM J. Comput.*, 47(3):890–916, 2018.

[4] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21, 2008.

[5] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[6] 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.

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

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

[9] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine. Fonctions de type positif sur les espaces L^p. *C. R. Acad. Sci. Paris*, 261:2153–2156, 1965.

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

[11] J. Cheeger and B. Kleiner. Realization of metric spaces as inverse limits, and bilipschitz embedding in L_1. *Geom. Funct. Anal.*, 23(1):96–133, 2013.

[12] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to L_1. *Acta Math.*, 207(2):291–373, 2011.

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

[14] N. R. Devanur, S. A. Khot, R. Saket, and N. K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 537–546. ACM, New York, 2006.

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

[16] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.

[17] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.

[18] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98 (Dallas, TX)*, pages 604–613. ACM, New York, 1999.

[19] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions (extended abstract). In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 68–80. IEEE Computer Society, 1988.

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

[21] N. J. Kalton, N. T. Peck, and J. W. Roberts. *An F-space sampler*, volume 89 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1984.

[22] D. Kane and R. Meka. A PRG for Lipschitz functions of polynomials with applications to sparsest cut. In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing*, pages 1–10. ACM, New York, 2013.

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

[24] S. A. Khot and N. K. Vishnoi. The unique games conjecture, integrability gap for cut problems and embeddability of negative-type metrics into ℓ_1. *J. ACM*, 62(1):Art. 8, 39, 2015.

[25] G. Kindler, N. Kirshner, and R. O'Donnell. Gaussian noise sensitivity and Fourier tails. *Israel J. Math.*, 225(1):71–109, 2018.

[26] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. *SIAM J. Comput.*, 38(6):2487–2498, 2009.

[27] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474, 2000.

[28] S. Kwapień. Unsolved Problems. *Studia Math.*, 38:467–483, 1970. Problem 3, page 469.

[29] J. R. Lee and A. Naor. L_p metrics on the Heisenberg group and the Goemans–Linial conjecture. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 99–108. IEEE Computer Society, 2006.

[30] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. I.* Springer-Verlag, Berlin-New York, 1977. Sequence spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 92.

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

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

[33] A. Naor. L_1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1549–1575. Hindustan Book Agency, New Delhi, 2010.

[34] 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.

[35] A. Naor and R. Young. Vertical perimeter versus horizontal perimeter. *Ann. of Math. (2)*, 188(1):171–279, 2018.

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

[37] R. O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014.

[38] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 360–369. ACM, New York, 2002.

[39] J. H. Wells and L. R. Williams. *Embeddings and extensions in analysis*. Springer-Verlag, New York-Heidelberg, 1975. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84.

(S.K.) DEPARTMENT OF COMPUTER SCIENCE, COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012-1185, USA
E-mail address: `khot@cs.nyu.edu`

(A.N.) MATHEMATICS DEPARTMENT, PRINCETON UNIVERSITY, FINE HALL, WASHINGTON ROAD, PRINCETON, NJ 08544-1000, USA
E-mail address: `naor@math.princeton.edu`