Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

L_1 embeddings of the Heisenberg group and fast estimation of graph isoperimetry

Assaf Naor*

Abstract. We survey connections between the theory of bi-Lipschitz embeddings and the Sparsest Cut Problem in combinatorial optimization. The story of the Sparsest Cut Problem is a striking example of the deep interplay between analysis, geometry, and probability on the one hand, and computational issues in discrete mathematics on the other. We explain how the key ideas evolved over the past 20 years, emphasizing the interactions with Banach space theory, geometric measure theory, and geometric group theory. As an important illustrative example, we shall examine recently established connections to the structure of the Heisenberg group, and the incompatibility of its Carnot-Carathéodory geometry with the geometry of the Lebesgue space L_1.

Mathematics Subject Classification (2000). 46B85, 30L05, 46B80, 51F99.

Keywords. Bi-Lipschitz embeddings, Sparsest Cut Problem, Heisenberg group.

1. Introduction

Among the common definitions of the Heisenberg group $\mathbb{H}$, it will be convenient for us to work here with $\mathbb{H}$ modeled as $\mathbb{R}^3$, equipped with the group product $(a,b,c)\cdot(a',b',c') \stackrel{\mathrm{def}}{=} (a+a', b+b', c+c'+ab'-ba')$. The integer lattice $\mathbb{Z}^3$ is then a discrete cocompact subgroup of $\mathbb{H}$, denoted by $\mathbb{H}(\mathbb{Z})$, which is generated by the finite symmetric set $\{(\pm1,0,0),(0,\pm1,0),(0,0,\pm1)\}$. The word metric on $\mathbb{H}(\mathbb{Z})$ induced by this generating set will be denoted by d_W.

As noted by Semmes [66], a differentiability result of Pansu [61] implies that the metric space $(\mathbb{H}(\mathbb{Z}), d_W)$ does not admit a bi-Lipschitz embedding into $\mathbb{R}^n$ for any $n \in \mathbb{N}$. This was extended by Pauls [62] to bi-Lipschitz non-embeddability results of $(\mathbb{H}(\mathbb{Z}), d_W)$ into metric spaces with either lower or upper curvature bounds in the sense of Alexandrov. In [52, 27] it was observed that Pansu's differentiability argument extends to Banach space targets with the Radon-Nikodým property (see [14, Ch. 5]), and hence $\mathbb{H}$ does not admit a bi-Lipschitz embedding into, say, a Banach space which is either reflexive or is a separable dual; in particular $\mathbb{H}$ does not admit a bi-Lipschitz embedding into any $L_p(\mu)$ space, $1 < p < \infty$, or into the sequence space ℓ_1.

*Research supported in part by NSF grants CCF-0635078 and CCF-0832795, BSF grant 2006009, and the Packard Foundation.

The embeddability of $\mathbb{H}(\mathbb{Z})$ into the function space $L_1(\mu)$, when μ is non-atomic, turned out to be much harder to settle. This question is of particular importance since it is well understood that for μ non-atomic, $L_1(\mu)$ is a space for which the differentiability results quoted above manifestly break down. Nevertheless, Cheeger and Kleiner [26, 25] introduced a novel notion of differentiability for which they could prove a differentiability theorem for Lipschitz maps from the Heisenberg group to $L_1(\mu)$, thus establishing that $\mathbb{H}(\mathbb{Z})$ does not admit a bi-Lipschitz embedding into any $L_1(\mu)$ space.

Another motivation for the $L_1(\mu)$ embeddability question for $\mathbb{H}(\mathbb{Z})$ originates from [52], where it was established that it is connected to the Sparsest Cut Problem in the field of combinatorial optimization. For this application it was of importance to obtain quantitative estimates in the $L_1(\mu)$ non-embeddability results for $\mathbb{H}(\mathbb{Z})$. It turns out that establishing such estimates is quite subtle, as they require overcoming finitary issues that do not arise in the infinite setting of [25, 28]. The following two theorems were proved in [29, 30]. Both theorems follow painlessly from a more general theorem that is stated and discussed in Section 5.4.

Theorem 1.1. *There exists a universal constant $c > 0$ such that any embedding into $L_1(\mu)$ of the restriction of the word metric d_W to the $n \times n \times n$ grid $\{1, \dots, n\}^3$ incurs distortion $\gtrsim (\log n)^c$.*

Following Gromov [38], the compression rate of $f : \mathbb{H}(\mathbb{Z}) \to L_1(\mu)$, denoted $\omega_f(\cdot)$, is defined as the largest non-decreasing function such that for all $x, y \in \mathbb{H}(\mathbb{Z})$ we have $\|f(x) - f(y)\|_1 \geqslant \omega_f(d_W(x, y))$ (see [7] for more information on this topic).

Theorem 1.2. *There exists a universal constant $c > 0$ such that for every function $f : \mathbb{H}(\mathbb{Z}) \to L_1(\mu)$ which is 1-Lipschitz with respect to the word metric d_W, we have $\omega_f(t) \lesssim t/(\log t)^c$ for arbitrarily large t.*

Evaluating the supremum of those $c > 0$ for which Theorem 1.1 holds true remains an important open question, with geometric significance as well as importance to theoretical computer science. Conceivably we could get c in Theorem 1.1 to be arbitrarily close to $\frac{1}{2}$, which would be sharp since the results of [8, 64] imply (see the explanation in [41]) that the metric space $\left(\{1, \dots, n\}^3, d_W\right)$ embeds into ℓ_1 with distortion $\lesssim \sqrt{\log n}$. Similarly, we do not know the best possible c in Theorem 1.2; $\frac{1}{2}$ is again the limit here since it was shown in [69] that there exists a 1-Lipschitz mapping $f : \mathbb{H}(\mathbb{Z}) \to \ell_1$ for which $\omega_f(t) \gtrsim t/(\sqrt{\log t} \cdot \log \log t)$.

The purpose of this article is to describe the above non-embeddability results for the Heisenberg group. Since one of the motivations for these investigations is the application to the Sparsest Cut Problem, we also include here a detailed discussion of this problem from theoretical computer science, and its deep connections to metric geometry. Our goal is to present the ideas in a way that is accessible to mathematicians who do not necessarily have background in computer science.

Acknowledgements. I am grateful to the following people for helpful comments and suggestions on earlier versions of this manuscript: Tim Austin, Keith Ball, Subhash Khot, Bruce Kleiner, Russ Lyons, Manor Mendel, Mikhail Ostrovskii, Gideon Schechtman, Lior Silberman.

2. Embeddings

A metric space $(\mathscr{M}, d_{\mathscr{M}})$ is said to embed with distortion $D \geqslant 1$ into a metric space $(\mathscr{Y}, d_{\mathscr{Y}})$ if there exists a mapping $f : \mathscr{M} \to \mathscr{Y}$, and a scaling factor $s > 0$, such that for all $x, y \in \mathscr{M}$ we have $sd_{\mathscr{M}}(x,y) \leqslant d_{\mathscr{Y}}(f(x), f(y)) \leqslant Dsd_{\mathscr{M}}(x,y)$. The infimum over those $D \geqslant 1$ for which $(\mathscr{M}, d_{\mathscr{M}})$ embeds with distortion D into $(\mathscr{Y}, d_{\mathscr{Y}})$ is denoted by $c_{\mathscr{Y}}(\mathscr{M})$. If $(\mathscr{M}, d_{\mathscr{M}})$ does not admit a bi-Lipschitz embedding into $(\mathscr{Y}, d_{\mathscr{Y}})$, we will write $c_{\mathscr{Y}}(\mathscr{M}) = \infty$.

Throughout this paper, for $p \geqslant 1$, the space L_p will stand for $L_p([0,1], \lambda)$, where λ is Lebesgue measure. The spaces ℓ_p and ℓ_p^n will stand for the space of p-summable infinite sequences, and $\mathbb{R}^n$ equipped with the ℓ_p norm, respectively. Much of this paper will deal with bi-Lipschitz embeddings of *finite* metric spaces into L_p. Since every n-point subset of an $L_p(\Omega, \mu)$ space embeds isometrically into $\ell_p^{n(n-1)/2}$ (see the discussion in [12]), when it comes to embeddings of finite metric spaces, the distinction between different $L_p(\Omega, \mu)$ spaces is irrelevant. Nevertheless, later, in the study of the embeddability of the Heisenberg group, we will need to distinguish between sequence spaces and function spaces.

For $p \geqslant 1$ we will use the shorter notation $c_p(\mathscr{M}) = c_{L_p}(\mathscr{M})$. The parameter $c_2(\mathscr{M})$ is known as the Euclidean distortion of $\mathscr{M}$. Dvoretzky's theorem says that if $\mathscr{Y}$ is an infinite dimensional Banach space then $c_{\mathscr{Y}}(\ell_2^n) = 1$ for all $n \in \mathbb{N}$. Thus, for every finite metric space $\mathscr{M}$ and every infinite dimensional Banach space $\mathscr{Y}$, we have $c_2(\mathscr{M}) \geqslant c_{\mathscr{Y}}(\mathscr{M})$.

The following famous theorem of Bourgain [15] will play a key role in what follows:

Theorem 2.1 (Bourgain's embedding theorem [15]). *For every n-point metric space $(\mathscr{M}, d_{\mathscr{M}})$, we have*

$$c_2(\mathscr{M}) \lesssim \log n. \tag{1}$$

Bourgain proved in [15] that the estimate (1) is sharp up to an iterated logarithm factor, i.e., that there exist arbitrarily large n-point metric spaces $\mathscr{M}_n$ for which $c_2(\mathscr{M}_n) \gtrsim \frac{\log n}{\log \log n}$. The $\log \log n$ term was removed in the important paper [56] of Linial, London and Rabinovich, who showed that the shortest path metric on bounded degree n-vertex expander graphs has Euclidean distortion $\gtrsim \log n$.

If one is interested only in embeddings into infinite dimensional Banach spaces, then Theorem 2.1 is stated in the strongest possible form: as noted above, it implies that for every infinite dimensional Banach space $\mathscr{Y}$, we have $c_{\mathscr{Y}}(\mathscr{M}) \lesssim \log n$. Below, we will actually use Theorem 2.1 for embeddings into L_1, i.e., we will use the fact that $c_1(\mathscr{M}) \lesssim \log n$. The expander based lower bound of Linial, London and Rabinovich [56] extends to embeddings into L_1 as well, i.e., even this weaker form of Bourgain's embedding theorem is asymptotically sharp. We refer to [58, Ch. 15] for a comprehensive discussion of these issues, as well as a nice presentation of the proof of Bourgain's embedding theorem.

3. L_1 as a metric space

Let (Ω,μ) be a measure space. Define a mapping $T : L_1(\Omega,\mu) \to L_\infty(\Omega\times\mathbb{R},\mu\times\lambda)$, where λ is Lebesgue measure, by:

$$T(f)(\omega,x) \stackrel{\mathrm{def}}{=} \begin{cases} 1 & 0 < x \leqslant f(\omega), \\ -1 & f(\omega) < x < 0, \\ 0 & \text{otherwise.} \end{cases}$$

For all $f,g \in L_1(\Omega,\mu)$ we have:

$$\Big| T(f)(\omega,x) - T(g)(\omega,x) \Big| = \begin{cases} 1 & g(\omega) < x \leqslant f(\omega) \text{ or } f(\omega) < x \leqslant g(\omega), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for all $p > 0$ we have,

$$\begin{aligned} \|T(f)-T(g)\|_{L_p(\Omega\times\mathbb{R},\mu\times\lambda)}^p &= \int_\Omega \left(\int_{(g(\omega),f(\omega)]\sqcup(f(\omega),g(\omega)]} d\lambda \right) d\mu(\omega) \\ &= \int_\Omega |f(\omega)-g(\omega)| d\mu(\omega) = \|f-g\|_{L_1(\Omega,\mu)}. \qquad (2) \end{aligned}$$

Specializing (2) to $p = 2$, we see that:

$$\|T(f)-T(g)\|_{L_2(\Omega\times\mathbb{R},\mu\times\lambda)} = \sqrt{\|f-g\|_{L_1(\Omega,\mu)}}.$$

Corollary 3.1. *The metric space* $\left(L_1(\Omega,\mu), \|f-g\|_{L_1(\Omega,\mu)}^{1/2}\right)$ *admits an isometric embedding into Hilbert space.*

Another useful corollary is obtained when (2) is specialized to the case $p = 1$. Take an arbitrary finite subset $X \subseteq L_1(\Omega,\mu)$. For every $(\omega,x) \in \Omega\times\mathbb{R}$ consider the set $S(\omega,x) = \{f \in X : \ x \leqslant f(\omega)\} \subseteq X$. For every $S \subseteq X$ we can define a measurable subset $E_S = \{(\omega,x) \in \Omega\times\mathbb{R} : \ S(\omega,x) = S\} \subseteq \Omega\times\mathbb{R}$. By the definition of T, for every $f,g \in X$ we have

$$\begin{aligned} \|f-g\|_{L_1(\Omega,\mu)} &\stackrel{(2)}{=} \|T(f)-T(g)\|_{L_1(\Omega\times\mathbb{R},\mu\times\lambda)} \\ &= \int_{\Omega\times\mathbb{R}} \Big| \mathbf{1}_{S(w,x)}(f) - \mathbf{1}_{S(w,x)}(g) \Big| d(\mu\times\lambda)(\omega,x) \\ &= \sum_{S\subseteq X} (\mu\times\lambda)(E_S) \Big| \mathbf{1}_S(f) - \mathbf{1}_S(g) \Big|, \end{aligned}$$

where here, and in what follows, $\mathbf{1}_S(\cdot)$ is the characteristic function of S. Writing $\beta_S = (\mu\times\lambda)(E_S)$, we have the following important corollary:

Corollary 3.2. *Let* $X \subseteq L_1(\Omega,\mu)$ *be a finite subset of* $L_1(\Omega,\mu)$*. Then there exist nonnegative numbers* $\{\beta_S\}_{S\subseteq X} \subseteq [0,\infty)$ *such that for all* $f,g \in X$ *we have:*

$$\|f-g\|_{L_1(\Omega,\mu)} = \sum_{S\subseteq X} \beta_S \Big| \mathbf{1}_S(f) - \mathbf{1}_S(g) \Big|. \qquad (3)$$

A metric space $(\mathscr{M}, d_{\mathscr{M}})$ is said to be of *negative type* if the metric space $\left(\mathscr{M}, d_{\mathscr{M}}^{1/2}\right)$ admits an isometric embedding into Hilbert space. Such metrics will play a crucial role in the ensuing discussion. This terminology (see e.g., [33]) is due to a classical theorem of Schoenberg [65], which asserts that $(\mathscr{M}, d_{\mathscr{M}})$ is of negative type if and only if for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$, the matrix $(d_{\mathscr{M}}(x_i, x_j))_{i,j=1}^n$ is negative semidefinite on the orthogonal complement of the main diagonal in $\mathbb{C}^n$, i.e., for all $\zeta_1, \ldots, \zeta_n \in \mathbb{C}$ with $\sum_{j=1}^n \zeta_j = 0$ we have $\sum_{i=1}^n \sum_{j=1}^n \zeta_i \overline{\zeta_j} d_{\mathscr{M}}(x_i, x_j) \leqslant 0$. Corollary (3.1) can be restated as saying that $L_1(\Omega, \mu)$ is a metric space of negative type.

Corollary (3.2) is often called the *cut cone representation of L_1 metrics*. To explain this terminology, consider the set $\mathscr{C} \subseteq \mathbb{R}^{n^2}$ of all $n \times n$ real matrices $A = (a_{ij})$ such that there is a measure space (Ω, μ) and $f_1, \ldots, f_n \in L_1(\Omega, \mu)$ with $a_{ij} = \|f_i - f_j\|_{L_1(\Omega,\mu)}$ for all $i, j \in \{1, \ldots, n\}$. If $f_1, \ldots, f_n \in L_1(\Omega_1, \mu_1)$ and $g_1, \ldots, g_n \in L_1(\Omega_2, \mu_2)$ then for all $c_1, c_2 \geqslant 0$ and $i, j \in \{1, \ldots, n\}$ we have

$$c_1\|f_i - f_j\|_{L_1(\Omega_1,\mu_1)} + c_2\|g_i - g_j\|_{L_1(\Omega_2,\mu_2)} = \|h_i - h_j\|_{L_1(\Omega_1 \sqcup \Omega_2, \mu_1 \sqcup \mu_2)},$$

where $h_1, \ldots, h_n$ are functions defined on the disjoint union $\Omega_1 \sqcup \Omega_2$ as follows: $h_i(\omega) = c_1 f_i(\omega)\mathbf{1}_{\Omega_1}(\omega) + c_2 g_i(\omega)\mathbf{1}_{\Omega_2}(\omega)$. This observation shows that $\mathscr{C}$ is a cone (of dimension $n(n-1)/2$). Identity (3) says that the cone $\mathscr{C}$ is generated by the rays induced by cut semimetrics, i.e., by matrices of the form $a_{ij} = |\mathbf{1}_S(i) - \mathbf{1}_S(j)|$ for some $S \subseteq \{1, \ldots, n\}$. It is not difficult to see that these rays are actually the extreme rays of the cone $\mathscr{C}$. Carathéodory's theorem (for cones) says that we can choose the coefficients $\{\beta_S\}_{S \subseteq X}$ in (3) so that only $n(n-1)/2$ of them are non-zero.

4. The Sparsest Cut Problem

Given $n \in \mathbb{N}$ and two symmetric functions $C, D : \{1, \ldots, n\} \times \{1, \ldots, n\} \to [0, \infty)$ (called capacities and demands, respectively), and a subset $\emptyset \neq S \subsetneq \{1, \ldots, n\}$, write

$$\Phi(S) \stackrel{\mathrm{def}}{=} \frac{\sum_{i=1}^n \sum_{j=1}^n C(i,j) \cdot |\mathbf{1}_S(i) - \mathbf{1}_S(j)|}{\sum_{i=1}^n \sum_{j=1}^n D(i,j) \cdot |\mathbf{1}_S(i) - \mathbf{1}_S(j)|}. \tag{4}$$

The value

$$\Phi^*(C, D) \stackrel{\mathrm{def}}{=} \min_{\emptyset \neq S \subsetneq \{1,\ldots,n\}} \Phi(S) \tag{5}$$

is the minimum over all cuts (two-part partitions) of $\{1, \ldots, n\}$ of the ratio between the total capacity crossing the boundary of the cut and the total demand crossing the boundary of the cut.

Finding in polynomial time a cut for which $\Phi^*(C, D)$ is attained up to a definite multiplicative constant is called the Sparsest Cut problem. This problem is used as a subroutine in many approximation algorithms for NP-hard problems; see the survey articles [68, 22], as well as [53, 1] and the references in [6, 5] for some of the vast literature on this topic. Computing $\Phi^*(C, D)$ exactly has been long known to be NP-hard [67]. More recently, it was shown in [31] that there exists $\varepsilon_0 > 0$ such that it is NP-hard to approximate $\Phi^*(C,D)$ to within a factor smaller than $1+\varepsilon_0$. In [47, 24] it was shown that it is Unique Games hard to approximate $\Phi^*(C,D)$ to within any constant factor (see [44, 45] for more information on the Unique Games Conjecture; we will return to this issue in Section 4.3.3).

It is customary in the literature to highlight the support of the capacities function C: this allows us to introduce a particulary important special case of the Sparsest Cut Problem. Thus, a different way to formulate the above setup is via an n-vertex graph $G=(V,E)$, with a positive weight (called a capacity) $C(e)$ associated to each edge $e \in E$, and a nonnegative weight (called a demand) $D(u,v)$ associated to each pair of vertices $u,v \in V$. The goal is to evaluate in polynomial time (and in particular, while examining only a negligible fraction of the subsets of V) the quantity:

$$\Phi^*(C,D) = \min_{\emptyset \neq S \subsetneq V} \frac{\sum_{uv\in E} C(uv)\,|\mathbf{1}_S(u)-\mathbf{1}_S(v)|}{\sum_{u,v\in V} D(u,v)\,|\mathbf{1}_S(u)-\mathbf{1}_S(v)|}\,.$$

To get a feeling for the meaning of Φ^*, consider the case $C(e)=D(u,v)=1$ for all $e \in E$ and $u,v \in V$. This is an important instance of the Sparsest Cut problem which is called "Sparsest Cut with Uniform Demands". In this case Φ^* becomes:

$$\Phi^* = \min_{\emptyset \neq S \subsetneq V} \frac{\#\{\text{edges joining } S \text{ and } V\setminus S\}}{|S|\cdot|V\setminus S|}\,.$$

Thus, in the case of uniform demands, the Sparsest Cut problem essentially amounts to solving efficiently the combinatorial isoperimetric problem on G: determining the subset of the graph whose ratio of edge boundary to its size is as small as possible.

In the literature it is also customary to emphasize the size of the support of the demand function D, i.e., to state bounds in terms of the number k of pairs $\{i,j\} \subseteq \{1,\ldots,n\}$ for which $D(i,j)>0$. For the sake of simplicity of exposition, we will not adopt this convention here, and state all of our bounds in terms of n rather than the number of positive demand pairs k. We refer to the relevant references for the simple modifications that are required to obtain bounds in terms of k alone.

From now on, the Sparsest Cut problem will be understood to be with general capacities and demands; when discussing the special case of uniform demands we will say so explicitly. In applications, general capacities and demands are used to tune the notion of "interface" between S and $V\setminus S$ to a wide variety of combinatorial optimization problems, which is one of the reasons why the Sparsest Cut problem is so versatile in the field of approximation algorithms.

4.1. Reformulation as an optimization problem over L_1. Although the Sparsest Cut Problem clearly has geometric flavor as a discrete isoperimetric problem, the following key reformulation of it, due to [11, 56], explicitly relates it to the geometry of L_1.

Lemma 4.1. *Given symmetric* $C, D : \{1,\ldots,n\}\times\{1,\ldots,n\}\to[0,\infty)$*, we have:*

$$\Phi^*(C,D)=\min_{f_1,\ldots,f_n\in L_1}\frac{\sum_{i=1}^n\sum_{j=1}^n C(i,j)\|f_i-f_j\|_1}{\sum_{i=1}^n\sum_{j=1}^n D(i,j)\|f_i-f_j\|_1}. \tag{6}$$

Proof. Let ϕ denote the right hand side of (6), and write $\Phi^*=\Phi^*(C,D)$. Given a subset $S\subseteq\{1,\ldots,n\}$, by considering $f_i=\mathbf{1}_S(i)\in\{0,1\}\subseteq L_1$ we see that that $\phi\leqslant\Phi^*$. In the reverse direction, if $X=\{f_1,\ldots,f_n\}\subseteq L_1$ then let $\{\beta_S\}_{S\subseteq X}$ be the non-negative weights from Corollary 3.2. For $S\subseteq X$ define a subset of $\{1,\ldots,n\}$ by $S'=\{i\in\{1,\ldots,n\}:\ f_i\in S\}$. It follows from the definition of Φ^* that for all $S\subseteq X$ we have,

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n C(i,j)|\mathbf{1}_S(f_i)-\mathbf{1}_S(f_j)| &\stackrel{(4)}{=}\Phi(S')\sum_{i=1}^n\sum_{j=1}^n D(i,j)|\mathbf{1}_S(f_i)-\mathbf{1}_S(f_j)|\\ &\stackrel{(5)}{\geqslant}\Phi^*\sum_{i=1}^n\sum_{j=1}^n D(i,j)|\mathbf{1}_S(f_i)-\mathbf{1}_S(f_j)|. \end{aligned}\tag{7}$$

Thus

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^n C(i,j)\|f_i-f_j\|_1 &\stackrel{(3)}{=}\sum_{S\subseteq X}\beta_S\sum_{i=1}^n\sum_{j=1}^n C(i,j)|\mathbf{1}_S(f_i)-\mathbf{1}_S(f_j)|\\ &\stackrel{(7)}{\geqslant}\Phi^*\sum_{S\subseteq X}\beta_S\sum_{i=1}^n\sum_{j=1}^n D(i,j)|\mathbf{1}_S(f_i)-\mathbf{1}_S(f_j)|\stackrel{(3)}{=}\sum_{i=1}^n\sum_{j=1}^n D(i,j)\|f_i-f_j\|_1.\end{aligned}$$

It follows that $\phi\geqslant\Phi^*$, as required. □

4.2. The linear program. Lemma 4.1 is a reformulation of the Sparsest Cut Problems in terms of a continuous optimization problem on the space L_1. Being a reformulation, it shows in particular that solving L_1 optimization problems such as the right hand side of (6) is NP-hard.

In the beautiful paper [53] of Leighton and Rao it was shown that there exists a polynomial time algorithm that, given an n-vertex graph $G=(V,E)$, computes a number which is guaranteed to be within a factor of $\lesssim\log n$ of the uniform Sparsest Cut value (4). The Leighton-Rao algorithm uses combinatorial ideas which do not apply to Sparsest Cut with general demands. A breakthrough result, due to Linial-London-Rabinovich [56] and Aumann-Rabani [9], introduced embedding methods to this field, yielding a polynomial time algorithm which computes $\Phi^*(C,D)$ up to a factor $\lesssim\log n$ for all $C,D:\{1,\ldots,n\}\times\{1,\ldots,n\}\to[0,\infty)$.

The key idea of [56, 9] is based on replacing the finite subset $\{f_1,\ldots,f_n\}$ of L_1 in (6) by an *arbitrary* semimetric on $\{1,\ldots,n\}$. Specifically, by homogeneity we can always assume that the denominator in (6) equals 1, in which case Lemma 4.1 says that $\Phi^*(C,D)$ equals the minimum of $\sum_{i=1}^n\sum_{j=1}^n C(i,j)d_{ij}$, given that $\sum_{i=1}^n\sum_{j=1}^n D(i,j)d_{ij}=1$ and there exist $f_1,\ldots,f_n\in L_1$ for which

$d_{ij} = \|f_i - f_j\|_1$ for all $i, j \in \{1, \ldots, n\}$. We can now ignore the fact that d_{ij} was a semimetric that came from a subset of L_1, i.e., we can define $M^*(C, D)$ to be the minimum of $\sum_{i=1}^n \sum_{j=1}^n C(i,j)d_{ij}$, given that $\sum_{i=1}^n \sum_{j=1}^n D(i,j)d_{ij} = 1$, $d_{ii} = 0$, $d_{ij} \geqslant 0$, $d_{ij} = d_{ji}$ for all $i, j \in \{1, \ldots, n\}$ ($n(n-1)/2$ symmetry constraints) and $d_{ij} \leqslant d_{ik} + d_{kj}$ for all $i, j, k \in \{1, \ldots, n\}$ ($\leqslant n^3$ triangle inequality constraints).

Clearly $M^*(C, D) \leqslant \Phi^*(C, D)$, since we are minimizing over all semimetrics rather than just those arising from subsets of L_1. Moreover, $M^*(C, D)$ can be computed in polynomial time up to arbitrarily good precision [40], since it is a linear program (minimizing a linear functional in the variables (d_{ij}) subject to polynomially many linear constraints).

The linear program produces a semimetric d^*_{ij} on $\{1, \ldots, n\}$ which satisfies $M^*(C, D) = \sum_{i=1}^n \sum_{j=1}^n C(i,j)d^*_{ij}$ and $\sum_{i=1}^n \sum_{j=1}^n D(i,j)d^*_{ij} = 1$ (ignoring arbitrarily small errors). By Lemma 4.1 we need to somehow relate this semimetric to L_1. It is at this juncture that we see the power of Bourgain's embedding theorem 2.1: the constraints of the linear program only provide us the information that d^*_{ij} is a semimetric, and nothing else. So, we need to be able to somehow handle arbitrary metric spaces—precisely what Bourgain's theorem does, by furnishing $f_1, \ldots, f_n \in L_1$ such that for all $i, j \in \{1, \ldots, n\}$ we have

$$\frac{d^*_{ij}}{\log n} \lesssim \|f_i - f_j\|_1 \leqslant d^*_{ij}. \tag{8}$$

Now,

$$\begin{aligned} \Phi^*(C, D) &\overset{(6)}{\leqslant} \frac{\sum_{i=1}^n \sum_{j=1}^n C(i,j)\|f_i - f_j\|_1}{\sum_{i=1}^n \sum_{j=1}^n D(i,j)\|f_i - f_j\|_1} \\ &\overset{(8)}{\lesssim} \log n \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n C(i,j)d^*_{ij}}{\sum_{i=1}^n \sum_{j=1}^n D(i,j)d^*_{ij}} = \log n \cdot M^*(C, D). \end{aligned} \tag{9}$$

Thus, $\frac{\Phi^*(C,D)}{\log n} \lesssim M^*(C, D) \leqslant \Phi^*(C, D)$, i.e., the polynomial time algorithm of computing $M^*(C, D)$ is guaranteed to produce a number which is within a factor $\lesssim \log n$ of $\Phi^*(C, D)$.

Remark 4.2. In the above argument we only discussed the algorithmic task of fast estimation of the number $\Phi^*(C, D)$, rather than the problem of producing in polynomial time a subset $\emptyset \neq S \subsetneq \{1, \ldots, n\}$ for which $\Phi^*(S)$ is close up to a certain multiplicative guarantee to the optimum value $\Phi^*(C, D)$. All the algorithms discussed in this paper produce such a set S, rather than just approximating the number $\Phi^*(C, D)$. In order to modify the argument above to this setting, one needs to go into the proof of Bourgain's embedding theorem, which as currently stated as just an existential result for $f_1, \ldots, f_n$ as in (8). This issue is addressed in [56], which provides an algorithmic version of Bourgain's theorem. Ensuing algorithms in this paper can be similarly modified to produce a good cut S, but we will ignore this issue from now on, and continue to focus solely on algorithms for approximate computation of $\Phi^*(C, D)$.

4.3. The semidefinite program. We have already stated in Section 2 that the logarithmic loss in the application (8) of Bourgain's theorem cannot be improved. Thus, in order to obtain a polynomial time algorithm with approximation guarantee better than $\lesssim \log n$, we need to impose additional geometric restrictions on the metric d_{ij}^*; conditions that will hopefully yield a class of metric spaces for which one can prove an L_1 distortion bound that is asymptotically smaller than the $\lesssim \log n$ of Bourgain's embedding theorem. This is indeed possible, based on a quadratic variant of the discussion in Section 4.2; an approach due to Goemans and Linial [37, 55, 54].

The idea of Goemans and Linial is based on Corollary 3.1, i.e., on the fact that the metric space L_1 is of negative type. We define $M^{**}(C,D)$ to be the minimum of $\sum_{i=1}^n \sum_{j=1}^n C(i,j)d_{ij}$, subject to the constraint that $\sum_{i=1}^n \sum_{j=1}^n D(i,j)d_{ij} = 1$ and d_{ij} is a semimetric of negative type on $\{1,\ldots,n\}$. The latter condition can be equivalently restated as the requirement that, in addition to d_{ij} being a semimetric on $\{1,\ldots,n\}$, there exist vectors $v_1,\ldots,v_n \in L_2$ such that $d_{ij} = \|v_i - v_j\|_2^2$ for all $i,j \in \{1,\ldots,n\}$. Equivalently, there exists a symmetric positive semidefinite $n\times n$ matrix (a_{ij}) (the Gram matrix of $v_1,\ldots,v_n$), such that $d_{ij} = a_{ii} + a_{jj} - 2a_{ij}$ for all $i,j \in \{1,\ldots,n\}$.

Thus, $M^{**}(C,D)$ is the minimum of $\sum_{i=1}^n \sum_{j=1}^n C(i,j)(a_{ii} + a_{jj} - 2a_{ij})$, a linear function in the variables (a_{ij}), subject to the constraint that (a_{ij}) is a symmetric positive semidefinite matrix, in conjunction with the linear constraints $\sum_{i=1}^n \sum_{j=1}^n D(i,j)(a_{ii} + a_{jj} - 2a_{ij}) = 1$ and for all $i,j,k \in \{1,\ldots,n\}$, the triangle inequality constraint $a_{ii} + a_{jj} - 2a_{ij} \leqslant (a_{ii} + a_{kk} - 2a_{ik}) + (a_{kk} + a_{jj} - 2a_{kj})$. Such an optimization problem is called a semidefinite program, and by the methods described in [40], $M^{**}(C,D)$ can be computed with arbitrarily good precision in polynomial time.

Corollary 3.1 and Lemma 4.1 imply that $M^*(C,D) \leqslant M^{**}(C,D) \leqslant \Phi^*(C,D)$. The following breakthrough result of Arora, Rao and Vazirani [6] shows that for Sparsest Cut with uniform demands the Goemans-Linial approach does indeed yield an improved approximation algorithm:

Theorem 4.3 ([6]). *In the case of uniform demands, i.e., if $C(i,j) \in \{0,1\}$ and $D(i,j) = 1$ for all $i,j \in \{1,\ldots,n\}$, we have*

$$\frac{\Phi^*(C,D)}{\sqrt{\log n}} \lesssim M^{**}(C,D) \leqslant \Phi^*(C,D). \tag{10}$$

In the case of general demands we have almost the same result, up to lower order factors:

Theorem 4.4 ([5]). *For all symmetric $C,D : \{1,\ldots,n\} \times \{1,\ldots,n\} \to [0,\infty)$ we have*

$$\frac{\Phi^*(C,D)}{(\log n)^{\frac{1}{2}+o(1)}} \lesssim M^{**}(C,D) \leqslant \Phi^*(C,D). \tag{11}$$

The $o(1)$ term in (11) is $\lesssim \frac{\log\log\log n}{\log\log n}$. We conjecture that it could be removed altogether, though at present it seems to be an inherent artifact of complications in the proof in [5].

Before explaining some of the ideas behind the proofs of Theorem 4.3 and Theorem 4.4 (the full details are quite lengthy and are beyond the scope of this survey), we prove, following [58, Prop. 15.5.2], a crucial identity (attributed in [58] to Y. Rabinovich) which reformulates these results in terms of an L_1 embeddability problem.

Lemma 4.5. *We have*

$$\sup\left\{\frac{\Phi^*(C,D)}{M^{**}(C,D)}:\ C,D:\{1,\ldots,n\}\times\{1,\ldots,n\}\to(0,\infty)\right\}$$
$$=\sup\Big\{c_1\big(\{1,\ldots,n\},d\big):\ d \text{ is a metric of negative type}\Big\}. \tag{12}$$

Proof. The proof of the fact that the left hand side of (12) is at most the right hand side of (12) is identical to the way (9) was deduced from (8).

In the reverse direction, let d^* be a metric of negative type on $\{1,\ldots,n\}$ for which $c_1(\{1,\ldots,n\},d^*)\stackrel{\mathrm{def}}{=}c$ is maximal among all such metrics. Let $\mathscr{C}\subseteq\mathbb{R}^{n^2}$ be the cone in the space of $n\times n$ symmetric matrices from the last paragraph of Section 3, i.e., $\mathscr{C}$ consists of all matrices of the form $(\|f_i-f_j\|_1)$ for some $f_1,\ldots,f_n\in L_1$.

Fix $\varepsilon\in(0,c-1)$ and let $\mathscr{K}_\varepsilon\subseteq\mathbb{R}^{n^2}$ be the set of all symmetric matrices (a_{ij}) for which there exists $s>0$ such that $sd^*(i,j)\leqslant a_{ij}\leqslant(c-\varepsilon)sd^*(i,j)$ for all $i,j\in\{1,\ldots,n\}$. By the definition of c, the convex sets $\mathscr{C}$ and $\mathscr{K}_\varepsilon$ are disjoint, since otherwise d^* would admit an embedding into L_1 with distortion $c-\varepsilon$. It follows that there exists a symmetric matrix $(h_{ij}^\varepsilon)\in\mathbb{R}^{n^2}\setminus\{0\}$ and $\alpha\in\mathbb{R}$, such that $\sum_{i=1}^n\sum_{j=1}^n h_{ij}^\varepsilon a_{ij}\leqslant\alpha$ for all $(a_{ij})\in\mathscr{K}_\varepsilon$, and $\sum_{i=1}^n\sum_{j=1}^n h_{ij}^\varepsilon b_{ij}\geqslant\alpha$ for all $(b_{ij})\in\mathscr{C}$. Since both $\mathscr{C}$ and $\mathscr{K}_\varepsilon$ are closed under multiplication by positive scalars, necessarily $\alpha=0$.

Define $C^\varepsilon(i,j)\stackrel{\mathrm{def}}{=}h_{ij}^\varepsilon\mathbf{1}_{\{h_{ij}^\varepsilon\geqslant0\}}$ and $D^\varepsilon(i,j)\stackrel{\mathrm{def}}{=}|h_{ij}^\varepsilon|\mathbf{1}_{\{h_{ij}^\varepsilon\leqslant0\}}$. By definition of $M^{**}(C^\varepsilon,D^\varepsilon)$,

$$\sum_{i=1}^n\sum_{j=1}^n C^\varepsilon(i,j)d_{ij}^*\geqslant M^{**}(C^\varepsilon,D^\varepsilon)\cdot\sum_{i=1}^n\sum_{j=1}^n D^\varepsilon(i,j)d_{ij}^*. \tag{13}$$

By considering $a_{ij}\stackrel{\mathrm{def}}{=}\Big((c-\varepsilon)\mathbf{1}_{\{h_{ij}^\varepsilon\geqslant0\}}+\mathbf{1}_{\{h_{ij}^\varepsilon<0\}}\Big)d^*(i,j)\in\mathscr{K}_\varepsilon$, the inequality $\sum_{i=1}^n\sum_{j=1}^n h_{ij}^\varepsilon a_{ij}\leqslant0$ becomes:

$$\sum_{i=1}^n\sum_{j=1}^n D^\varepsilon(i,j)d_{ij}^*\geqslant(c-\varepsilon)\sum_{i=1}^n\sum_{j=1}^n C^\varepsilon(i,j)d_{ij}^*. \tag{14}$$

A combination of (13) and (14) implies that $(c-\varepsilon)M^{**}(C^\varepsilon,D^\varepsilon)\leqslant1$. At the same time, for all $f_1,\ldots,f_n\in L_1$, the inequality $\sum_{i=1}^n\sum_{j=1}^n h_{ij}^\varepsilon\|f_i-f_j\|_1\geqslant0$ is the same as $\sum_{i=1}^n\sum_{j=1}^n C^\varepsilon(i,j)\|f_i-f_j\|_1\geqslant\sum_{i=1}^n\sum_{j=1}^n D^\varepsilon(i,j)\|f_i-f_j\|_1$, which by Lemma 6 means that $\Phi^*(C^\varepsilon,D^\varepsilon)\geqslant1$. Thus $\Phi^*(C^\varepsilon,D^\varepsilon)/M^{**}(C^\varepsilon,D^\varepsilon)\geqslant c-\varepsilon$, and since this holds for all $\varepsilon\in(0,c-1)$, the proof of Lemma 4.5 is complete. □

In the case of Sparsest Cut with uniform demands, we have the following result which is analogous to Lemma 4.5, where the L_1 bi-Lipschitz distortion is replaced by the smallest possible factor by which 1-Lipschitz functions into L_1 can distort the *average distance*. The proof is a slight variant of the proof of Lemma 4.5; the simple details are left to the reader. This connection between Sparsest Cut with uniform demands and embeddings that preserve the average distance is due to Rabinovich [63].

Lemma 4.6. *The supremum of $\Phi^*(C,D)/M^{**}(C,D)$ over all instances of uniform demands, i.e., when $C(i,j)\in\{0,1\}$ and $D(i,j)=1$ for all $i,j\in\{1,\dots,n\}$, equals the infimum over $A>0$ such that for all metrics d on $\{1,\dots,n\}$ of negative type, there exist $f_1,\dots,f_n\in L_1$ satisfying $\|f_i-f_j\|_1\leqslant d(i,j)$ for all $i,j\in\{1,\dots,n\}$ and $A\sum_{i=1}^n\sum_{j=1}^n\|f_i-f_j\|_1\geqslant\sum_{i=1}^n\sum_{j=1}^n d(i,j)$.*

4.3.1. L_2 embeddings of negative type metrics. The proof of Theorem 4.3 in [6] is based on a clever geometric partitioning procedure for metrics of negative type. Building heavily on ideas of [6], in conjunction with some substantial additional combinatorial arguments, an alternative approach to Theorem 4.3 was obtained in [59], based on a purely graph theoretical statement which is of independent interest. We shall now sketch this approach, since it is modular and general, and as such it is useful for additional geometric corollaries. We refer to [59] for more information on these additional applications, as well as to [6] for the original proof of Theorem 4.3.

Let $G=(V,E)$ be an n-vertex graph. The *vertex expansion* of G, denoted $h(G)$, is the largest $h\geqslant 0$ such that every $S\subseteq V$ with $|S|\leqslant n/2$ has at least $h|S|$ neighbors in $V\setminus S$. The *edge expansion* of G, denoted $\alpha(G)$, is the largest $\alpha\geqslant 0$ such that for every $S\subseteq V$ with $|S|\leqslant n/2$, the number of edges joining S and $V\setminus S$ is at least $\alpha|S|\cdot\frac{|E|}{n}$. The main combinatorial statement of [59] relates these two notions of expansion of graphs:

Theorem 4.7 (Edge Replacement Theorem [59]). *For every graph $G=(V,E)$ with $h(G)\geqslant\frac12$ there is a set of edges E' on V with $\alpha(V,E')\gtrsim 1$, and such that for every $uv\in E'$ we have $d_G(u,v)\lesssim\sqrt{\log|V|}$. Here d_G is the shortest path metric on G (with respect to the original edge set E), and all implicit constants are universal.*

It is shown in [59] that the $\lesssim\sqrt{\log|V|}$ bound on the length of the new edges in Theorem 4.7 is asymptotically tight. The proof of Theorem 4.7 is involved, and cannot be described here: it has two components, a combinatorial construction, as well a purely Hilbertian geometric argument based on, and simpler than, the original algorithm of [6]. We shall now explain how Theorem 4.7 implies Theorem 4.3 (this is somewhat different from the deduction in [59], which deals with a different semidefinite program for Sparsest Cut with uniform demands).

Proof of Theorem 4.3 assuming Theorem 4.7. An application of (the easy direction of) Lemma 4.6 shows that in order to prove Theorem 4.3 it suffices to show that if $(\mathscr{M},d)$ is an n-point metric space of negative type, with $\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}d(x,y)=1$, then there exists a mapping $F:\mathscr{M}\to\mathbb{R}$ which is 1-Lipschitz and such that

$\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}|F(x)-F(y)|\gtrsim 1/\sqrt{\log n}$. In what follows we use the standard notation for closed balls: for $x\in\mathscr{M}$ and $t\geqslant 0$, set $B(x,t)=\{y\in\mathscr{M}:\ d(x,y)\leqslant t\}$.

Choose $x_0\in\mathscr{M}$ with $\frac{1}{n}\sum_{y\in\mathscr{M}}d(x_0,y)=r\stackrel{\mathrm{def}}{=}\min_{x\in\mathscr{M}}\frac{1}{n}\sum_{y\in\mathscr{M}}d(x,y)$. Then $r\leqslant\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}d(x,y)=1$, implying $1\geqslant\frac{1}{n}\sum_{y\in\mathscr{M}}d(x_0,y)>\frac{2}{n}|\mathscr{M}\setminus B(x_0,2)|$, or $|B(x_0,2)|>n/2$. Similarly $|B(x_0,4)|>3n/4$.

Assume first that $\frac{1}{n^2}\sum_{x,y\in B(x_0,4)}d(x,y)\leqslant\frac14$ (this will be the easy case). Then

$$\begin{aligned}1=\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}d(x,y)&\leqslant\frac14+\frac{2}{n^2}\sum_{x\in\mathscr{M}}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}\Big(d(x,x_0)+d(x_0,y)\Big)\\&=\frac14+\frac{2r}{n}|\mathscr{M}\setminus B(x_0,4)|+\frac{2}{n}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}d(x_0,y)\leqslant\frac34+\frac{2}{n}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}d(x_0,y),\end{aligned}$$

or $\frac{1}{n}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}d(x_0,y)\geqslant\frac18$. Define a 1-Lipschitz mapping $F:\mathscr{M}\to\mathbb{R}$ by $F(x)=d\big(x,B(x_0,2)\big)=\min_{y\in B(x_0,2)}d(x,y)$. The triangle inequality implies that for every $y\in\mathscr{M}\setminus B(x_0,4)$ we have $F(y)\geqslant\frac12 d(y,x_0)$. Thus

$$\begin{aligned}\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}|F(x)-F(y)|&\geqslant\frac{|B(x_0,2)|}{n^2}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}d\big(y,B(x_0,2)\big)\\&>\frac{1}{2n}\sum_{y\in\mathscr{M}\setminus B(x_0,4)}\frac12 d(y,x_0)\gtrsim 1=\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}d(x,y).\end{aligned}$$

This completes the easy case, where there is even no loss of $1/\sqrt{\log n}$ (and we did not use yet the assumption that d is a metric of negative type).

We may therefore assume from now on that $\frac{1}{n^2}\sum_{x,y\in B(x_0,4)}d(x,y)\geqslant\frac14$. The fact that d is of negative type means that there are vectors $\{v_x\}_{x\in\mathscr{M}}\subseteq L_2$ such that $d(x,y)=\|v_x-v_y\|_2^2$ for all $x,y\in\mathscr{M}$.

We will show that for a small enough universal constant $\varepsilon>0$, there are two sets $S_1,S_2\subseteq B(x_0,4)$ such that $|S_1|,|S_2|\geqslant\varepsilon n$ and $d(S_1,S_2)\geqslant\varepsilon^2/\sqrt{\log n}$. Once this is achieved, the mapping $F:\mathscr{M}\to\mathbb{R}$ given by $F(x)=d(x,S_1)$ will satisfy $\frac{1}{n^2}\sum_{x,y\in\mathscr{M}}|F(x)-F(y)|\geqslant\frac{2}{n^2}|S_1|\cdot|S_2|\frac{\varepsilon^2}{\sqrt{\log n}}\geqslant\frac{2\varepsilon^4}{\sqrt{\log n}}$, as desired.

Assume for contradiction that no such S_1,S_2 exist. Define a set of edges E_0 on $B(x_0,4)$ by $E_0\stackrel{\mathrm{def}}{=}\Big\{\{x,y\}\subseteq B(x_0,4):\ x\neq y\ \wedge\ d(x,y)<\varepsilon^2/\sqrt{\log n}\Big\}$. Our contrapositive assumption says that any two subsets $S_1,S_2\subseteq B(x_0,4)$ with $|S_1|,|S_2|\geqslant\varepsilon n\geqslant\varepsilon|B(x_0,4)|$ are joined by an edge from E_0. By a (simple) general graph theoretical lemma (see [59, Lem 2.3]), this implies that, provided $\varepsilon\leqslant 1/10$, there exists a subset $V\subseteq B(x_0,4)$ with $|V|\geqslant(1-\varepsilon)|B(x_0,4)|\gtrsim n$, such that the graph induced by E_0 on V, i.e., $G=\Big(V,E=E_0\cap\binom{V}{2}\Big)$, has $h(G)\geqslant\frac12$.

We are now in position to apply the Edge Replacement Theorem, i.e., Theorem 4.7. We obtain a new set of edges E' on V such that $\alpha(V,E')\gtrsim 1$ and for every $xy\in E'$ we have $d_G(x,y)\lesssim\sqrt{\log n}$. The latter condition means that there exists a path $\{x=x_0,x_1,\dots,x_m=y\}\subseteq V$ such that $m\lesssim\sqrt{\log n}$ and $x_ix_{i-1}\in E$

for every $i \in \{1, \dots, m\}$. By the definition of E, this implies that

$$xy \in E' \implies d(x,y) \leqslant \sum_{i=1}^{n} d(x_i, x_{i-1}) \leqslant m\frac{\varepsilon^2}{\sqrt{\log n}} \lesssim \varepsilon^2. \tag{15}$$

It is a standard fact (the equivalence between edge expansion and a Cheeger inequality) that for every $f : V \to L_1$ we have

$$\frac{1}{|E'|}\sum_{xy\in E'} \|f(x)-f(y)\|_1 \geqslant \frac{\alpha(V,E')}{2|V|^2}\sum_{x,y\in V}\|f(x)-f(y)\|_1. \tag{16}$$

For a proof of (16) see [59, Fact 2.1]: this is a simple consequence of the cut cone representation, i.e., Corollary 3.2, since the identity (3) shows that it suffices to prove (16) when $f(x) = \mathbf{1}_S(x)$ for some $S \subseteq V$, in which case the desired inequality follows immediately from the definition of the edge expansion $\alpha(V, E')$.

Since L_2 is isometric to a subset of L_1 (see, e.g., [71]), it follows from (16) and the fact that $\alpha(V, E') \gtrsim 1$ that

$$\begin{aligned}\varepsilon \overset{(15)}{\gtrsim} \frac{1}{|E'|}\sum_{xy\in E'}\sqrt{d(x,y)} &= \frac{1}{|E'|}\sum_{xy\in E'}\|v_x - v_y\|_2 \\ &\gtrsim \frac{1}{|V|^2}\sum_{x,y\in V}\|v_x-v_y\|_2 \gtrsim \frac{1}{n^2}\sum_{x,y\in V}\sqrt{d(x,y)}.\end{aligned} \tag{17}$$

Now comes the point where we use the assumption $\frac{1}{n^2}\sum_{x,y\in B(x_0,4)} d(x,y) \geqslant \frac14$. Since for any $x, y \in B(x_0, 4)$ we have $d(x, y) \leqslant 8$, it follows that the number of pairs $(x, y) \in B(x_0, 4) \times B(x_0, 4)$ with $d(x, y) \geqslant 1/8$ is at least $n^2/64$. Since $|V| \geqslant (1-\varepsilon)|B(x_0, 4)|$, the number of such pairs which are also in $V \times V$ is at least $\frac{n^2}{64} - 3\varepsilon n^2 \gtrsim n^2$, provided ε is small enough. Thus $\frac{1}{n^2}\sum_{x,y\in V}\sqrt{d(x,y)} \gtrsim 1$, and (17) becomes a contradiction for small enough ε. □

Remark 4.8. The above proof of Theorem 4.7 used very little of the fact that d is a metric of negative type. In fact, all that was required was that d admits a quasisymmetric embedding into L_2; see [59].

It remains to say a few words about the proof of Theorem 4.4. Unfortunately, the present proof of this theorem is long and involved, and it relies on a variety of results from metric embedding theory. It would be of interest to obtain a simpler proof. Lemma 4.5 implies that Theorem 4.4 is a consequence of the following embedding result:

Theorem 4.9 ([5]). *Every n-point metric space of negative type embeds into Hilbert space with distortion $\lesssim (\log n)^{\frac12 + o(1)}$.*

Theorem 4.9 improves over the previously known [23] bound of $\lesssim (\log n)^{3/4}$ on the Euclidean distortion of n-point metric spaces of negative type. As we shall explain below, Theorem 4.9 is tight up to the $o(1)$ term.

The proof of Theorem 4.9 uses the following notion from [5]:

Definition 4.10 (Random zero-sets [5]). Fix $\Delta,\ \zeta > 0$, and $p \in (0,1)$. A metric space $(\mathscr{M}, d)$ is said to admit a random zero set at scale Δ, which is ζ-spreading with probability p, if there is a probability distribution μ over subsets $Z \subseteq \mathscr{M}$ such that $\mu\left(\{Z:\ y \in Z\ \wedge\ d(x,Z) \geqslant \Delta/\zeta\}\right) \geqslant p$ for every $x, y \in \mathscr{M}$ with $d(x,y) \geqslant \Delta$. We denote by $\zeta(\mathscr{M}; p)$ the least $\zeta > 0$ such that for every $\Delta > 0$, $\mathscr{M}$ admits a random zero set at scale Δ which is ζ-spreading with probability p.

The connection to metrics of negative type is due to the following theorem, which can be viewed as the main structural consequence of [6]. Its proof uses [6] in conjunction with two additional ingredients: an analysis of the algorithm of [6] due to [50], and a clever iterative application of the algorithm of [6], due to [23], while carefully reweighting points at each step.

Theorem 4.11 (Random zero sets for negative type metrics). *There exists a universal constant $p > 0$ such that any n-point metric space $(\mathscr{M}, d)$ of negative type satisfies $\zeta(\mathscr{M}; p) \lesssim \sqrt{\log n}$.*

Random zero sets are related to embeddings as follows. Fix $\Delta > 0$. Let $(\mathscr{M}, d)$ be a finite metric space, and fix $S \subseteq \mathscr{M}$. By the definition of $\zeta(S;p)$, there exists a distribution μ over subsets $Z \subseteq S$ such that for every $x, y \in S$ with $d(x,y) \geqslant \Delta$ we have $\mu\left(\{Z \subseteq S:\ y \in Z\ \wedge\ d(x,Z) \geqslant \Delta/\zeta(S;p)\}\right) \geqslant p$. Define $\varphi_{S,\Delta} : \mathscr{M} \to L_2(\mu)$ by $\varphi_{S,\Delta}(x) = d(x,Z)$. Then $\varphi_{S,\Delta}$ is 1-Lipschitz, and for every $x, y \in S$ with $d(x,y) \geqslant \Delta$,

$$\|\varphi_{S,\Delta}(x) - \varphi_{S,\Delta}(y)\|_{L_2(\mu)} = \left(\int_{2^S} [d(x,Z) - d(y,Z)]^2\, d\mu(Z)\right)^{1/2} \geqslant \frac{\Delta\sqrt{p}}{\zeta(S;p)}. \quad (18)$$

The remaining task is to "glue" the mappings $\{\varphi_{S,\Delta}:\ \Delta > 0,\ S \subseteq \mathscr{M}\}$ to form an embedding of $\mathscr{M}$ into Hilbert space with the distortion claimed in Theorem 4.9. A key ingredient of the proof of Theorem 4.9 is the embedding method called "Measured Descent", that was developed in [48]. The results of [48] were stated as embedding theorems rather than a gluing procedure; the realization that a part of the arguments of [48] can be formulated explicitly as a general "gluing lemma" is due to [50]. In [5] it was necessary to enhance the Measured Descent technique in order to prove the following key theorem, which together with (18) and Theorem 4.11 implies Theorem 4.9. See also [4] for a different enhancement of Measured Descent, which also implies Theorem 4.9. The proof of Theorem 4.12 is quite intricate; we refer to [5] for the details.

Theorem 4.12. *Let $(\mathscr{M}, d)$ be an n-point metric space. Suppose that there is $\varepsilon \in [1/2, 1]$ such that for every $\Delta > 0$, and every subset $S \subseteq \mathscr{M}$, there exists a 1-Lipschitz map $\varphi_{S,\Delta} : \mathscr{M} \to L_2$ with $||\varphi_{S,\Delta}(x) - \varphi_{S,\Delta}(y)||_2 \gtrsim \Delta/(\log|S|)^{\varepsilon}$ whenever $x, y \in S$ and $d(x,y) \geqslant \Delta$. Then $c_2(\mathscr{M}) \lesssim (\log n)^{\varepsilon} \log\log n$.*

The following corollary is an obvious consequence of Theorem 4.9, due to the fact that L_1 is a metric space of negative type.

Corollary 4.13. *Every $X \subseteq L_1$ embeds into L_2 with distortion $\lesssim (\log|X|)^{\frac{1}{2}+o(1)}$.*

We stated Corollary 4.13 since it is of special importance: in 1969, Enflo [34] proved that the Hamming cube, i.e., $\{0,1\}^k$ equipped with the metric induced from ℓ_1^k, has Euclidean distortion $\sqrt{k}$. Corollary 4.13 says that up to lower order factors, the Hamming cube is among the most non-Euclidean subsets of L_1. There are very few known results of this type, i.e., (almost) sharp evaluations of the largest Euclidean distortion of an n-point subset of a natural metric space. A notable such result is Matoušek's theorem [57] that any n-point subset of the infinite binary tree has Euclidean distortion $\lesssim \sqrt{\log\log n}$, and consequently, due to [20], the same holds true for n-point subsets of, say, the hyperbolic plane. This is tight due to Bourgain's matching lower bound [16] for the Euclidean distortion of finite depth complete binary trees.

4.3.2. The Goemans-Linial conjecture. Theorem 4.4 is the best known approximation algorithm for the Sparsest Cut Problem (and Theorem 4.3 is the best known algorithm in the case of uniform demands). But, a comparison of Lemma 4.5 and Theorem 4.9 reveals a possible avenue for further improvement: Theorem 4.9 produces an embedding of negative type metrics into L_2 (for which the bound of Theorem 4.9 is sharp up to lower order factors), while for Lemma 4.5 all we need is an embedding into the larger space L_1. It was conjectured by Goemans and Linial (see [37, 55, 54] and [58, pg. 379–380]) that any finite metric space of negative type embeds into L_1 with distortion $\lesssim 1$. If true, this would yield, via the Goemans-Linial semidefinite relaxation, a constant factor approximation algorithm for Sparsest Cut.

As we shall see below, it turns out that the Goemans-Linial conjecture is false, and in fact there exist [30] arbitrarily large n-point metric spaces $\mathscr{M}_n$ of negative type for which $c_1(\mathscr{M}_n) \geqslant (\log n)^c$, where c is a universal constant. Due to the duality argument in Lemma 4.5, this means that the algorithm of Section 4.3 is doomed to make an error of at least $(\log n)^c$, i.e., there exist capacity and demand functions $C_n, D_n : \{1,\ldots,n\} \times \{1,\ldots,n\} \to [0,\infty)$ for which we have $M^{**}(C_n,D_n) \lesssim \Phi^*(C_n,D_n)/(\log n)^c$. Such a statement is referred to in the literature as the fact that the *integrality gap* of the Goemans-Linial semidefinite relaxation of Sparsest Cut is at least $(\log n)^c$.

4.3.3. Unique Games hardness and the Khot-Vishnoi integrality gap. Khot's Unique Games Conjecture [44] is that for every $\varepsilon > 0$ there exists a prime $p = p(\varepsilon)$ such that there is no polynomial time algorithm that, given $n \in \mathbb{N}$ and a system of m-linear equations in n-variables of the form $x_i - x_j = c_{ij} \mod p$ for some $c_{ij} \in \mathbb{N}$, determines whether there exists an assignment of an integer value to each variable x_i such that at least $(1-\varepsilon)m$ of the equations are satisfied, or whether no assignment of such values can satisfy more than εm of the equations (if neither of these possibilities occur, then an arbitrary output is allowed). This formulation of the conjecture is due to [46], where it is shown that it is equivalent to the original formulation in [44]. The Unique Games Conjecture is by now a common assumption that has numerous applications in computational complexity; see the survey [45] (in this collection) for more information.

In [47, 24] it was shown that the existence of a polynomial time constant factor approximation algorithm for Sparsest Cut would refute the Unique Games Conjecture, i.e., one can use a polynomial time constant factor approximation algorithm for Sparsest Cut to solve in polynomial time the above algorithmic task for linear equations.

For a period of time in 2004, this computational hardness result led to a strange situation: either the complexity theoretic Unique Games Conjecture is true, or the purely geometric Goemans-Linial conjecture is true, but not both. In a remarkable tour de force, Khot and Vishnoi [47] delved into the proof of their hardness result and managed to construct from it a concrete family of arbitrarily large n-point metric spaces $\mathscr{M}_n$ of negative type for which $c_1(\mathscr{M}_n) \gtrsim (\log\log n)^c$, where c is a universal constant, thus refuting the Goemans-Linial conjecture. Subsequently, these Khot-Vishnoi metric spaces $\mathscr{M}_n$ were analyzed in [49], resulting in the lower bound $c_1(\mathscr{M}_n) \gtrsim \log\log n$. Further work in [32] yielded a $\gtrsim \log\log n$ integrality gap for Sparsest Cut with uniform demands, i.e., "average distortion" L_1 embeddings (in the sense of Lemma 4.6) of negative type metrics were ruled out as well.

4.3.4. The Bretagnolle, Dacunha-Castelle, Krivine theorem and invariant metrics on Abelian groups. A combination of Schoenberg's classical characterization [65] of metric spaces that are isometric to subsets of Hilbert space, and a theorem of Bretagnolle, Dacunha-Castelle and Krivine [18] (see also [70]), implies that if $p \in [1,2]$ and $(X, \|\cdot\|_X)$ is a separable Banach space such that the metric space $(X, \|x-y\|_X^{p/2})$ is isometric to a subset of Hilbert space, then X is (linearly) isometric to a subspace of L_p. Specializing to $p=1$ we see that the Goemans-Linial conjecture is true for Banach spaces. With this motivation for the Goemans-Linial conjecture in mind, one notices that the Goemans-Linial conjecture is part of a natural one parameter family of conjectures which attempt to extend the theorem of Bretagnolle, Dacunha-Castelle and Krivine to general metric spaces rather than Banach spaces: is it true that for $p \in [1,2)$ any metric space $(\mathscr{M}, d)$ for which $(\mathscr{M}, d^{p/2})$ is isometric to a subset of L_2 admits a bi-Lipschitz embedding into L_p? This generalized Goemans-Linial conjecture turns out to be false for all $p \in [1,2)$; our example based on the Heisenberg group furnishes counter-examples for all p.

It is also known that certain invariant metrics on Abelian groups satisfy the Goemans-Linial conjecture:

Theorem 4.14 ([10]). *Let G be a finite Abelian group, equipped with an invariant metric ρ. Suppose that $2 \leqslant m \in \mathbb{N}$ satisfies $mx = 0$ for all $x \in G$. Denote $D = c_2\left(G, \sqrt{\rho}\right)$. Then $c_1(G,\rho) \lesssim D^4 \log m$.*

It is an interesting open question whether the dependence on the exponent m of the group G in Theorem 4.14 is necessary. Can one construct a counter-example to the Goemans-Linial conjecture which is an invariant metric on the cyclic group C_n of order n? Or, is there for every $D \geqslant 1$ a constant $K(D)$ such that for every invariant metric ρ on C_n for which $c_2\left(G, \sqrt{\rho}\right) \leqslant D$ we have $c_1(G,\rho) \leqslant K(D)$?

One can view the above discussion as motivation for why one might consider the Heisenberg group as a potential counter-example to the Goemans-Linial conjecture. Assuming that we are interested in invariant metrics on groups, we wish to depart from the setting of Abelian groups or Banach spaces, and if at the same time we would like our example to have some useful analytic properties (such as invariance under rescaling and the availability of a group norm), the Heisenberg group suggests itself as a natural candidate. This plan is carried out in Section 5.

5. Embeddings of the Heisenberg group

The purpose of this section is to discuss Theorem 1.1 and Theorem 1.2 from the introduction. Before doing so, we have an important item of unfinished business: relating the Heisenberg group to the Sparsest Cut Problem. We will do this in Section 5.1, following [52].

In preparation, we need to recall the Carnot-Carathéodory geometry of the continuous Heisenberg group $\mathbb{H}$, i.e., $\mathbb{R}^3$ equipped with the non-commutative product $(a,b,c)\cdot(a',b',c') = (a+a', b+b', c+c'+ab'-ba')$. Due to lack of space, this will have to be a crash course, and we refer to the relevant introductory sections of [29] for a more thorough discussion.

The identity element of $\mathbb{H}$ is $e=(0,0,0)$, and the inverse element of $(a,b,c)\in\mathbb{H}$ is $(-a,-b,-c)$. The center of $\mathbb{H}$ is the z-axis $\{0\}\times\{0\}\times\mathbb{R}$. For $g\in\mathbb{H}$ the *horizontal plane* at g is defined as $\mathbb{H}_g = g(\mathbb{R}\times\mathbb{R}\times\{0\})$. An affine line $L\subseteq\mathbb{H}$ is called a *horizontal line* if for some $g\in\mathbb{H}$ it passes through g and is contained in the affine plane $\mathbb{H}_g$. The standard scalar product $\langle\cdot,\cdot\rangle$ on $\mathbb{H}_e$ naturally induces a scalar product $\langle\cdot,\cdot\rangle_g$ on $\mathbb{H}_g$ by $\langle gx, gy\rangle_g = \langle x,y\rangle$. Consequently, we can define the Carnot-Carathéodory metric $d^{\mathbb{H}}$ on $\mathbb{H}$ by letting $d^{\mathbb{H}}(g,h)$ be the infimum of lengths of smooth curves $\gamma:[0,1]\to\mathbb{H}$ such that $\gamma(0)=g$, $\gamma(1)=h$ and for all $t\in[0,1]$ we have $\gamma'(t)\in H_{\gamma(t)}$ (and, the length of $\gamma'(t)$ is computed with respect to the scalar product $\langle\cdot,\cdot\rangle_{\gamma(t)}$). The *ball-box principle* (see [39]) implies that $d^{\mathbb{H}}\big((a,b,c),(a',b',c')\big)$ is bounded above and below by a constant multiple of $|a-a'|+|b-b'|+\sqrt{|c-c'+ab'-ba'|}$. Moreover, since the integer grid $\mathbb{H}(\mathbb{Z})$ is a discrete cocompact subgroup of $\mathbb{H}$, the word metric d_W on $\mathbb{H}(\mathbb{Z})$ is bi-Lipschitz equivalent to the restriction of $d^{\mathbb{H}}$ to $\mathbb{H}(\mathbb{Z})$ (see, e.g, [19]). For $\theta>0$ define the dilation operator $\delta_\theta:\mathbb{H}\to\mathbb{H}$ by $\delta_\theta(a,b,c)=(\theta a,\theta b,\theta^2 c)$. Then for all $g,h\in\mathbb{H}$ we have $d^{\mathbb{H}}(\delta_\theta(g),\delta_\theta(h))=\theta d^{\mathbb{H}}(g,h)$. The Lebesgue measure $\mathscr{L}_3$ on $\mathbb{R}^3$ is a Haar measure of $\mathbb{H}$, and the volume of a $d^{\mathbb{H}}$-ball of radius r is proportional to r^4.

5.1. Heisenberg metrics with isometric L_p snowflakes.

For every $(a,b,c)\in\mathbb{H}$ and $p\in[1,2)$, define

$$M_p(a,b,c) = \sqrt[4]{(a^2+b^2)^2+4c^2}\cdot\left(\cos\left(\frac{p}{2}\arccos\left(\frac{a^2+b^2}{\sqrt{(a^2+b^2)^2+4c^2}}\right)\right)\right)^{1/p}.$$

It was shown in [52] that M_p is a *group norm* on $\mathbb{H}$, i.e., for all $g,h\in\mathbb{H}$ and $\theta\geqslant 0$ we have $M_p(gh)\leqslant M_p(g)+M_p(h)$, $M_p(g^{-1})=M_p(g)$ and $M_p(\delta_\theta(g))=\theta M_p(g)$. Thus $d_p(g,h)\stackrel{\mathrm{def}}{=}M_p(g^{-1}h)$ is a left-invariant metric on $\mathbb{H}$. The metric d_p is bi-Lipschitz equivalent to $d^{\mathbb{H}}$ with distortion of order $1/\sqrt{2-p}$ (see [52]). Moreover, it was shown in [52] that $(\mathbb{H}, d_p^{p/2})$ admits an isometric embedding into L_2. Thus, in particular, the metric space $(\mathbb{H}, d_1)$, which bi-Lipschitz equivalent to $(\mathbb{H}, d^{\mathbb{H}})$, is of negative type.

The fact that $(\mathbb{H}, d^{\mathbb{H}})$ does not admit a bi-Lipschitz embedding into L_p for any $1 \leqslant p < \infty$ will show that the generalized Goemans-Linial conjecture (see Section 4.3.4) is false. In particular, $(\mathbb{H}, d_1)$, and hence by a standard rescaling argument also $(\mathbb{H}(\mathbb{Z}), d_1)$, is a counter-example to the Goemans-Linial conjecture. Note that it is crucial here that we are dealing with the function space L_p rather than the sequence space ℓ_p, in order to use a compactness argument to deduce from this statement that there exist arbitrarily large n-point metric spaces $(\mathscr{M}_n, d)$ such that $(\mathscr{M}_n, d^{p/2})$ is isometric to a subset of L_2, yet $\lim_{n\to\infty} c_p(\mathscr{M}_n) = \infty$. The fact that this statement follows from non-embeddability into L_p is a consequence of a well known ultrapower argument (see [42]), yet for ℓ_p this statement is false (e.g., ℓ_2 does not admit a bi-Lipschitz embedding into ℓ_p, but all finite subsets of ℓ_2 embed isometrically into ℓ_p). Unfortunately, this issue creates substantial difficulties in the case of primary interest $p = 1$. In the reflexive range $p > 1$, or for a separable dual space such as ℓ_1 $(= c_0^*)$, the non-embeddability of $\mathbb{H}$ follows from a natural extension of a classical result of Pansu [61], as we explain in Section 5.2. This approach fails badly when it comes to embeddings into L_1: for this purpose a novel method of Cheeger and Kleiner [25] is needed, as described in Section 5.3.

5.2. Pansu differentiability. Let X be a Banach space and $f : \mathbb{H} \to X$. Following [61], f is said to have a Pansu derivative at $x \in \mathbb{H}$ if for every $y \in \mathbb{H}$ the limit $D_f^x(y) \stackrel{\mathrm{def}}{=} \lim_{\theta\to 0} \big(f(x\delta_\theta(y)) - f(x)\big)/\theta$ exists, and $D_f^x : \mathbb{H} \to X$ is a group homomorphism, i.e., for all $y_1, y_2 \in \mathbb{H}$ we have $D_f^x(y_1 y_2^{-1}) = D_f^x(y_1) - D_f^y(y_2)$. Pansu proved [61] that every $f : \mathbb{H} \to \mathbb{R}^n$ which is Lipschitz in the metric $d^{\mathbb{H}}$ is Pansu differentiable almost everywhere. It was observed in [52, 27] that this result holds true if the target space $\mathbb{R}^n$ is replaced by any Banach space with the Radon-Nikodým property, in particular X can be any reflexive Banach space such as L_p for $p \in (1, \infty)$, or a separable dual Banach space such as ℓ_1. As noted by Semmes [66], this implies that $\mathbb{H}$ does not admit a bi-Lipschitz embedding into any Banach space X with the Radon-Nikodým property: a bi-Lipschitz condition for f implies that at a point $x \in \mathbb{H}$ of Pansu differentiability, D_f^x is also bi-Lipschitz, and in particular a group isomorphism. But that's impossible since $\mathbb{H}$ is non-commutative, unlike the additive group of X.

5.3. Cheeger-Kleiner differentiability. Differentiability theorems fail badly when the target space is L_1, even for functions defined on $\mathbb{R}$; consider Aronszajn's example [3] of the "moving indicator function" $t \mapsto \mathbf{1}_{[0,t]} \in L_1$. For L_1-valued Lipschitz functions on $\mathbb{H}$, Cheeger and Kleiner [25, 28] developed an alternative differentiation theory, which is sufficiently strong to show that $\mathbb{H}$ does not admit a bi-Lipschitz embedding into L_1. Roughly speaking, a differentiation theorem states that in the infinitesimal limit, a Lipschitz mapping converges to a mapping that belongs to a certain "structured" subclass of mappings (e.g., linear mappings or group homomorphisms). The Cheeger-Kleiner theory shows that, in a sense that will be made precise below, L_1-valued Lipschitz functions on $\mathbb{H}$ are in the infinitesimal limit similar to Aronszajn's moving indicator.

For an open subset $U \subseteq \mathbb{H}$ let $\mathrm{Cut}(U)$ denote the space of (equivalences classes up to measure zero) of measurable subsets of U. Let $f : U \to L_1$ be a Lipschitz function. An infinitary variant of the cut-cone decomposition of Corollary 3.2 (see [25]) asserts that there exists a measure Σ_f on $\mathrm{Cut}(U)$, such that for all $x, y \in U$ we have $\|f(x) - f(y)\|_1 = \int_{\mathrm{Cut}(U)} |\mathbf{1}_E(x) - \mathbf{1}_E(y)| d\Sigma_f(E)$. The measure Σ_f is called the *cut measure* of f. The idea of Cheeger and Kleiner is to detect the "infinitesimal regularity" of f in terms of the infinitesimal behavior of the measure Σ_f; more precisely, in terms of the shape of the sets E in the support of Σ_f, after passing to an infinitesimal limit.

Theorem 5.1 (Cheeger-Kleiner differentiability theorem [25, 28]). *For almost every $x \in U$ there exists a measure Σ_f^x on* $\mathrm{Cut}(\mathbb{H})$ *such that for all $y, z \in \mathbb{H}$ we have*

$$\lim_{\theta \to 0} \frac{\|f(x\delta_\theta(y)) - f(x\delta_\theta(z))\|_1}{\theta} = \int_{\mathrm{Cut}(\mathbb{H})} |\mathbf{1}_E(y) - \mathbf{1}_E(z)| d\Sigma_f^x(E). \tag{19}$$

Moreover, the measure Σ_f^x is supported on affine half-spaces whose boundary is a vertical plane, i.e., a plane which isn't of the form $\mathbb{H}_g$ for some $g \in \mathbb{H}$ (equivalently, an inverse image, with respect to the orthogonal projection from $\mathbb{R}^3$ onto $\mathbb{R}\times\mathbb{R}\times\{0\}$, of a line in $\mathbb{R} \times \mathbb{R} \times \{0\}$).

Theorem 5.1 is incompatible with f being bi-Lipschitz, since the right hand side of (19) vanishes when y, z lie on the same coset of the center of $\mathbb{H}$, while if f is bi-Lipschitz the left hand side of (19) is at least a constant multiple of $d^{\mathbb{H}}(y, z)$.

5.4. Compression bounds for L_1 embeddings of the Heisenberg group.

Theorem 1.1 and Theorem 1.2 are both a consequence of the following result from [29]:

Theorem 5.2 (Quantitative central collapse [29]). *There exists a universal constant $c \in (0, 1)$ such that for every $p \in \mathbb{H}$, every 1-Lipschitz $f : B(p, 1) \to L_1$, and every $\varepsilon \in \left(0, \frac{1}{4}\right)$, there exists $r \geqslant \varepsilon$ such that with respect to Haar measure, for at least half of the points $x \in B(p, 1/2)$, at least half of the pairs of points $(x_1, x_2) \in B(x, r) \times B(x, r)$ which lie on the same coset of the center with $d^{\mathbb{H}}(x_1, x_2) \in [\varepsilon r/2, 3\varepsilon r/2]$ satisfy:*

$$\|f(x_1) - f(x_2)\|_1 \leqslant \frac{d^{\mathbb{H}}(x_1, x_2)}{(\log(1/\varepsilon))^c}.$$

It isn't difficult to see that Theorem 5.2 implies Theorem 1.1 and Theorem 1.2. For example, in the setting of Theorem 1.1 we are given a bi-Lipschitz embedding $f : \{1, \ldots, n\}^3 \to L_1$, and using either the general extension theorem of [51] or a partition of unity argument, we can extend f to a Lipschitz (with respect to $d^{\mathbb{H}}$) mapping $\bar{f} : [1, n]^3 \to L_1$, whose Lipschitz constant is at most a constant multiple of the Lipschitz constant of f. Theorem 5.2 (after rescaling by n) produces a pair of points $y, z \in [1,n]^3$ of distance $\gtrsim \sqrt{n}$, whose distance is contracted under $\bar{f}$ by $\gtrsim (\log n)^c$. By rounding y, z to their nearest integer points in $\{1,\ldots,n\}^3$, we conclude that f itself must have bi-Lipschitz distortion $\gtrsim (\log n)^c$. The deduction of Theorem 1.2 from Theorem 5.2 is just as simple; see [29].

Theorem 5.2 is a quantitative version of Theorem 5.1, in the sense that it gives a definite lower bound on the macroscopic scale at which a given amount of collapse of cosets of the center, as exhibited by the differentiation result (19), occurs. As explained in [29, Rem. 2.1], one cannot hope in general to obtain rate bounds in differentiation results such as (19). Nevertheless, there are situations where "quantitative differentiation results" have been successfully proved; important precursors of Theorem 5.2 include the work of Bourgain [17], Jones [43], Matoušek [57], and Bates, Johnson, Lindenstrauss, Preiss, Schechtman [13]. Specifically, we should mention that Bourgain [17] obtained a lower bound on $\varepsilon > 0$ such that any embedding of an ε-net in a unit ball of an n-dimensional normed space X into a normed space Y has roughly the same distortion as the distortion required to embed all of X into Y, and Matoušek [57], in his study of embeddings of trees into uniformly convex spaces, obtained quantitative bounds on the scale at which "metric differentiation" is almost achieved, i.e., a scale at which discrete geodesics are mapped by a Lipschitz function to "almost geodesics". These earlier results are in the spirit of Theorem 5.2, though the proof of Theorem 5.2 in [29] is substantially more involved.

We shall now say a few words on the proof of Theorem 5.2; for lack of space this will have to be a rough sketch, so we refer to [29] for more details, as well as to the somewhat different presentation in [30]. Cheeger and Kleiner obtained two different proofs of Theorem 5.1. The first proof [25] started with the important observation that the fact that f is Lipschitz forces the cut measure Σ_f to be supported on sets with additional regularity, namely sets of *finite perimeter*. Moreover, there is a definite bound on the *total perimeter*: $\int_{\mathrm{Cut}(U)} \mathrm{PER}(E, B(p,1))d\Sigma_f(E) \lesssim 1$, where $\mathrm{PER}(E, B(p,1))$ denotes the perimeter of E in the ball $B(p,1)$ (we refer to the book [2], and the detailed explanation in [25, 29] for more information on these notions). Theorem 5.2 is then proved in [25] via an appeal to results [35, 36] on the infinitesimal structure of sets of finite perimeter in $\mathbb{H}$. A different proof of Theorem 5.2 was found in [28]. It is based on the notion of *metric differentiation*, which is used in [28] to reduce the problem to mappings $f : \mathbb{H} \to L_1$ for which the cut measure is supported on *monotone sets*, i.e., sets $E \subseteq \mathbb{H}$ such that for every horizontal line L, up to a set of measure zero, both $L \cap E$ and $L \cap (\mathbb{H} \setminus E)$ are either empty or subrays of L. A non-trivial classification of monotone sets is then proved in [28]: such sets are up to measure zero half-spaces.

This second proof of Theorem 5.2 avoids completely the use of perimeter bounds. Nevertheless, the starting point of the proof of Theorem 5.2 can be viewed as a hybrid argument, which incorporates both perimeter bounds, and a new classification of *almost* monotone sets. The quantitative setting of Theorem 5.2 leads to issues that do not have analogues in the non-quantitative proofs (e.g., the approximate classification results of "almost" monotone sets in balls cannot be simply that such sets are close to half-spaces in the entire ball; see [29, Example 9.1]).

In order to proceed we need to quantify the extent to which a set $E \subseteq B(x,r)$ is monotone. For a horizontal line $L \subseteq \mathbb{H}$ define the non-convexity $\mathrm{NC}_{B(x,r)}(E,L)$ of (E,L) on $B(x,r)$ as the infimum of $\int_{L\cap B(x,r)} \left|\mathbf{1}_I - \mathbf{1}_{E\cap L\cap B(x,r)}\right| d\mathcal{H}^1_L$ over all sub-intervals $I \subseteq L \cap B(x,r)$. Here $\mathcal{H}^1_L$ is the 1-dimensional Hausdorff measure on L (induced from the metric $d^{\mathbb{H}}$). The non-monotonicity of (E,L) on $B(x,r)$ is defined to be $\mathrm{NM}_{B(x,r)}(E,L) \stackrel{\mathrm{def}}{=} \mathrm{NC}_{B(x,r)}(E,L) + \mathrm{NC}_{B(x,r)}(\mathbb{H}\setminus E, L)$. The total non-monotonicity of E on $B(x,r)$ is defined as:

$$\mathrm{NM}_{B(x,r)}(E) \stackrel{\mathrm{def}}{=} \frac{1}{r^4}\int_{\mathrm{lines}(B(x,r))} \mathrm{NM}_{B(x,r)}(E,L)d\mathcal{N}(L),$$

where $\mathrm{lines}(U)$ denotes the set of horizontal lines in $\mathbb{H}$ which intersect U, and $\mathcal{N}$ is the left invariant measure on $\mathrm{lines}(\mathbb{H})$, normalized so that the measure of $\mathrm{lines}(B(e,1))$ is 1.

The following stability result for monotone sets constitutes the bulk of [29]:

Theorem 5.3. *There exists a universal constant $a > 0$ such that if a measurable set $E \subseteq B(x,r)$ satisfies $\mathrm{NM}_{B(x,r)}(E) \leqslant \varepsilon^a$ then there exists a half-space $\mathcal{P}$ such that*

$$\frac{\mathscr{L}_3\left((E\cap B(x,\varepsilon r))\triangle\mathcal{P}\right)}{\mathscr{L}_3(B(x,\varepsilon r))} < \varepsilon^{1/3}.$$

Perimeter bounds are used in [29, 30] for two purposes. The first is finding a controlled scale r such that at most locations, apart from a certain collection of cuts, the mass of Σ_f is supported on subsets which satisfy the assumption of Theorem 5.3 (see [30, Sec. 9]). But, the excluded cuts may have infinite measure with respect to Σ_f. Nonetheless, using perimeter bounds once more, together with the isoperimetric inequality in $\mathbb{H}$ (see [60, 21]), it is shown that their contribution to the metric is negligibly small (see [30, Sec. 8]).

By Theorem 5.3, it remains to deal with the situation where all the cuts in the support of Σ_f are close to half-spaces: note that we are not claiming in Theorem 5.3 that the half-space is vertical. Nevertheless, a simple geometric argument shows that even in the case of cut measures that are supported on general (almost) half-spaces, the mapping f must significantly distort some distances. The key point here is that if the cut measure is actually supported on half spaces, then it follows (after the fact) that for *every affine* line L, if $x_1, x_2, x_3 \in L$ and x_2 lies between x_1 and x_3 then $\|f(x_1) - f(x_3)\|_1 = \|f(x_1) - f(x_2)\|_1 + \|f(x_2) - f(x_3)\|_1$. But if L is vertical then $d^{\mathbb{H}}|_L$ is bi-Lipschitz to the *square root* of the difference of the z-coordinates, and it is trivial to verify that this metric on L is not bi-Lipschitz equivalent to a metric on L satisfying this additivity condition. For the details of (a quantitative version of) this final step of the argument see [30, Sec. 10].

References

[1] A. Agrawal, P. Klein, R. Ravi, and S. Rao. Approximation through multicommodity flow. In *31st Annual Symposium on Foundations of Computer Science*, pages 726–737. IEEE Computer Soc., Los Alamitos, CA, 1990.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[3] N. Aronszajn. Differentiability of Lipschitzian mappings between Banach spaces. *Studia Math.*, 57(2):147–190, 1976.

[4] S. Arora, J. R. Lee, and A. Naor. Fréchet embeddings of negative type metrics. *Discrete Comput. Geom.*, 38(4):726–739, 2007.

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

[6] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231 (electronic), New York, 2004. ACM.

[7] G. Arzhantseva, C. Drutu, and M. Sapir. Compression functions of uniform embeddings of groups into Hilbert and Banach spaces. *J. Reine Angew. Math.*, 633:213–235, 2009.

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

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

[10] T. Austin, A. Naor, and A. Valette. The Euclidean distortion of the lamplighter group. Preprint, 2007. To appear in Discrete Comput. Geom.

[11] D. Avis and M. Deza. The cut cone, L^1 embeddability, complexity, and multicommodity flows. *Networks*, 21(6):595–617, 1991.

[12] K. Ball. Isometric embedding in l_p-spaces. *European J. Combin.*, 11(4):305–311, 1990.

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

[14] 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.

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

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

[17] 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.

[18] J. Bretagnolle, D. Dacunha-Castelle, and J.-L. Krivine. Lois stables et espaces L^p. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 2:231–259, 1965/1966.

[19] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[20] S. Buyalo and V. Schroeder. Embedding of hyperbolic spaces in the product of trees. *Geom. Dedicata*, 113:75–93, 2005.

[21] L. Capogna, D. Danielli, and N. Garofalo. The geometric Sobolev embedding for vector fields and the isoperimetric inequality. *Comm. Anal. Geom.*, 2(2):203–215, 1994.

[22] S. Chawla. Sparsest cut. In M.-Y. Kao, editor, *Encyclopedia of Algorithms*. Springer, 2008.

[23] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. *ACM Trans. Algorithms*, 4(2):Art. 22, 18, 2008.

[24] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Comput. Complexity*, 15(2):94–114, 2006.

[25] J. Cheeger and B. Kleiner. Differentiating maps into L^1 and the geometry of BV functions. To appear in Ann. Math., preprint available at http://arxiv.org/abs/math/0611954, 2006.

[26] J. Cheeger and B. Kleiner. Generalized differentiation and bi-Lipschitz nonembedding in L^1. *C. R. Math. Acad. Sci. Paris*, 343(5):297–301, 2006.

[27] J. Cheeger and B. Kleiner. On the differentiability of Lipschitz maps from metric measure spaces to Banach spaces. In *Inspired by S. S. Chern*, volume 11 of *Nankai Tracts Math.*, pages 129–152. World Sci. Publ., Hackensack, NJ, 2006.

[28] J. Cheeger and B. Kleiner. Metric differentiation, monotonicity and maps to L^1. Preprint available at `http://arxiv.org/abs/0907.3295`, 2009.

[29] J. Cheeger, B. Kleiner, and A. Naor. Compression bounds for Lipschitz maps from the Heisenberg group to L_1. Preprint, 2009. `http://arxiv.org/abs/0910.2026`.

[30] J. Cheeger, B. Kleiner, and A. Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the Sparsest Cut SDP. In *Proceedings of 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, pages 555–564, 2009.

[31] J. Chuzhoy and S. Khanna. Polynomial flow-cut gaps and hardness of directed cut problems [extended abstract]. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 179–188. ACM, New York, 2007.

[32] 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.

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

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

[35] B. Franchi, R. Serapioni, and F. Serra Cassano. Rectifiability and perimeter in the Heisenberg group. *Math. Ann.*, 321(3):479–531, 2001.

[36] B. Franchi, R. Serapioni, and F. Serra Cassano. On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.*, 13(3):421–466, 2003.

[37] M. X. Goemans. Semidefinite programming in combinatorial optimization. *Math. Programming*, 79(1-3, Ser. B):143–161, 1997. Lectures on mathematical programming (ismp97) (Lausanne, 1997).

[38] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 of *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.

[39] M. Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-riemannian geometry*, Progr. in Math., pages 79–323. Birkhäuser, Basel, 1996.

[40] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[41] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS*, pages 534–543. IEEE Computer Society, 2003.

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

[43] P. W. Jones. Lipschitz and bi-Lipschitz functions. *Rev. Mat. Iberoamericana*, 4(1):115–121, 1988.

[44] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 767–775 (electronic), New York, 2002. ACM.

[45] S. Khot. Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. To appear in *Proceedings of the International Congress of Mathematicians, (Hyderabad, 2010)*, 2010.

[46] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357 (electronic), 2007.

[47] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Conference on Foundations of Computer Science (FOCS 2005)*, pages 53–62, 2005.

[48] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[49] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1010–1017, New York, 2006. ACM.

[50] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 92–101 (electronic), New York, 2005. ACM.

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

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

[53] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

[54] N. Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press.

[55] N. Linial. Squared ℓ_2 metrics into ℓ_1. In *Open problems on embeddings of finite metric spaces, edited by J. Matoušek*, page 5. 2002.

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

[57] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.

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

[59] 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.

[60] P. Pansu. Une inégalité isopérimétrique sur le groupe de Heisenberg. *C. R. Acad. Sci. Paris Sér. I Math.*, 295(2):127–130, 1982.

[61] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[62] S. D. Pauls. The large scale geometry of nilpotent Lie groups. *Comm. Anal. Geom.*, 9(5):951–982, 2001.

[63] Y. Rabinovich. On average distortion of embedding metrics into the line. *Discrete Comput. Geom.*, 39(4):720–733, 2008.

[64] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry (Miami Beach, FL, 1999)*, pages 300–306 (electronic), New York, 1999. ACM.

[65] I. J. Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.

[66] S. Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

[67] F. Shahrokhi and D. W. Matula. The maximum concurrent flow problem. *J. Assoc. Comput. Mach.*, 37(2):318–334, 1990.

[68] D. B. Shmoys. Cut problems and their application to divide-and-conquer. In *Approximation Algorithms for NP-hard Problems, (D.S. Hochbaum, ed.)*, pages 192–235. PWS, 1997.

[69] R. Tessera. Quantitative property A, Poincaré inequalities, L^p-compression and L^p-distortion for metric measure spaces. *Geom. Dedicata*, 136:203–220, 2008.

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

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

New York University, Courant Institute of Mathematical Sciences, 251 Mercer Street,
New York, NY 10012, USA
E-mail: naor@cims.nyu.edu