L_p metrics on the Heisenberg group and the Goemans-Linial conjecture

James R. Lee* Assaf Naor †

Abstract

We prove that the function $d : \mathbb{R}^3 \times \mathbb{R}^3 \to [0, \infty)$ given by

$$d\Big((x, y, z), (t, u, v)\Big) = \left(\left[\left((t-x)^2 + (u-y)^2\right)^2 + (v - z + 2xu - 2yt)^2\right]^{\frac{1}{2}} + (t-x)^2 + (u-y)^2\right)^{\frac{1}{2}}.$$

is a metric on $\mathbb{R}^3$ such that $(\mathbb{R}^3, \sqrt{d})$ is isometric to a subset of Hilbert space, yet $(\mathbb{R}^3, d)$ does not admit a bi-Lipschitz embedding into L_1. This yields a new simple counter example to the Goemans-Linial conjecture on the integrality gap of the semidefinite relaxation of the Sparsest Cut problem. The metric above is doubling, and hence has a padded stochastic decomposition at every scale. We also study the L_p version of this problem, and obtain a counter example to a natural generalization of a classical theorem of Bretagnolle, Dacunha-Castelle and Krivine (of which the Goemans-Linial conjecture is a particular case). Our methods involve Fourier analytic techniques, and a recent breakthrough of Cheeger and Kleiner, together with classical results of Pansu on the differentiability of Lipschitz functions on the Heisenberg group.

1 Introduction

Let $G = (V, E)$ be a graph, with a *capacity* $C(e) \geq 0$ associated to every edge $e \in E$. Assume that we are given k pairs of vertices $(s_1, t_1), ..., (s_k, t_k) \in V \times V$ and $D_1, \ldots, D_k \geq 1$. We think of the s_i as *sources*, the t_i as *targets*, and the value D_i as the *demand* of the *terminal pair* (s_i, t_i) for some *commodity* κ_i. The problem is said to have *uniform demands* if *every* pair $u, v \in V$ occurs as some (s_i, t_i) pair with $D_i = 1$. Given a non-empty subset $S \subsetneq V$, we write

$$\Phi(S) = \frac{\sum_{uv \in E} C(uv) \cdot |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{i=1}^k D_i \cdot |\mathbf{1}_S(s_i) - \mathbf{1}_S(t_i)|},$$

where $\mathbf{1}_S$ is the characteristic function of S. The value $\Phi^* = \min_{\emptyset \neq S \subsetneq V} \Phi(S)$ is the minimum over all cuts (partitions) of V, of the ratio between the total capacity crossing the cut and the total demand crossing the cut. In the case of uniform demands Φ^* is simply the edge expansion of the graph G.

*School of Mathematics, Institute for Advanced Study, Princeton, NJ. This material is based on work supported by the National Science Foundation under agreement No. DMS-0111298 and NSF grant CCR-0324906. Email: jrl@math.ias.edu.

†Theory Group, Microsoft Research, Redmond WA. Email: anaor@microsoft.com.

Computing Φ^* is NP-hard [40]. Moreover, finding a cut for which Φ^* is (approximately) attained is a basic step in approximation algorithms for several NP-hard problems [34, 1, 47]. The problem of approximating Φ^* in polynomial time is known as the *Sparsest Cut problem with general demands*, and is a famous open problem in the field of approximation algorithms. The best known algorithm for this problem is based on the following classical semidefinite relaxation (see [19, 2] for the motivation for this relaxation).

SDP relaxtion for Sparsest Cut

$$
\begin{array}{rl}
\min & \sum_{uv\in E} C(uv)\,\|x_u - x_v\|_2^2 \\
\text{s.t.} & x_u \in \mathbb{R}^n \qquad\qquad \forall u \in V \\
& \sum_{u,v\in V} D(u,v)\,\|x_u - x_v\|_2^2 = 1 \\
& \|x_u - x_v\|_2^2 \le \|x_u - x_w\|_2^2 + \|x_w - x_v\|_2^2 \\
& \qquad\qquad\qquad \forall u, v, w \in V
\end{array}
$$

Goemans and Linial (see [19, 36]) observed that this SDP produces a metric space (X, d) such that $(X, \sqrt{d})$ is isometric to a subset of Hilbert space. Such metrics are known in the literature as *negative type metrics* or *squared L_2 metrics*. Moreover, following the approach of London, Linial and Rabinovich [39] (see also the work of Aumann and Rabani [5]), they noted that the cut-cone characterization of subsets of L_1 (see [17]) implies that the integrality gap of this SDP can be bounded by the least distortion with which (X, d) embeds into L_1, i.e. the smallest $L > 0$ for which there is an embedding $f : X \to L_1$ satisfying for all $x, y \in X$, $d(x, y) \le \|f(x) - f(y)\|_1 \le L\, d(x, y)$. We will denote the smallest such L by $c_1(X)$ in what follows. In fact it is known that the worst-case integrality gap of the SDP over all instances of graphs with n nodes is *precisely* the largest value of $c_1(X)$ as X ranges over all n-point metric spaces of negative type. Goemans and Linial therefore made the following conjecture (see [19, 38]), which would imply that there exists a polynomial time algorithm which approximates Φ^* to within a constant factor.

The Goemans-Linial conjecture: Every metric space of negative type embeds with $O(1)$ distortion into L_1.

In a recent remarkable paper [26] Khot and Vishnoi proved that this conjecture does not hold true. In other words, there exist arbitrarily large n-point metric spaces of negative type X_n such that $\lim_{n\to\infty} c_1(X_n) = \infty$. Their construction is motivated by considerations from complexity theory, as it is based on a hardness result that will be discussed later. In particular, the Khot-Vishnoi spaces are quite complicated to describe. The purpose of the present paper is to give a different simple counter-example to the Goemans-Linial conjecture which is based on a classical and well-understood metric space— the Heisenberg group. Moreover, our example has several additional properties which lead to a solution of related problems. We will describe the Heisenberg group geometry later, but for concreteness we first state explicitly our counter-example.

Theorem 1.1. *Define $d : \mathbb{R}^3 \times \mathbb{R}^3 \to [0, \infty)$ by*

$$d\Big((x,y,z),(t,u,v)\Big) = \left(\left[\left((t-x)^2+(u-y)^2\right)^2 + (v-z+2xu-2yt)^2\right]^{\frac12} + (t-x)^2+(u-y)^2\right)^{\frac12}.$$

Then d is a metric on $\mathbb{R}^3$ and $(\mathbb{R}^3,\sqrt{d})$ is isometric to a subset of Hilbert space (i.e. $(\mathbb{R}^3,d)$ is a metric space of negative type). But, $(\mathbb{R}^3,d)$ does not embed bi-Lipschitzly into L_1. Thus

$$\lim_{n\to\infty} c_1\left(\{0,\ldots,n\}^3,d\right)=\infty. \tag{1}$$

Our approach can be used to answer additional questions to which the Khot-Vishnoi method does not apply. In order to motivate them we recall some background which explains why the Goemans-Linial conjecture is a natural question. A classical theorem of Schoenberg [45] (see also [50]) states that a metric space (X,d) is isometric to a subset of L_2 if and only if d^2 is a negative definite kernel on X, i.e. for every $x_1,\ldots,x_n\in X$ and every $c_1,\ldots,c_n\in\mathbb{C}$ with $\sum_{j=1}^n c_j=0$ we have $\sum_{j,k=1}^n d(x_j,x_k)^2c_j\overline{c_k}\le 0$. Thus (X,d) is of negative type if and only if d is negative definite on X. More generally Schoenberg proved that for $1\le p\le 2$ the function $\|x-y\|_p^p$ is negative definite on L_p. It follows that L_1 is of negative type (this corollary of Schoenberg's theorem is easy to prove directly). It also follows that for $1\le p\le 2$ the space L_p equipped with the metric $\|x-y\|_p^{p/2}$ is isometric to a subset of L_2. Bretagnolle, Dacunha-Castelle and Krivine [7] proved the following beautiful converse to this result of Schoenberg in the case of normed spaces: If $(X,\|\cdot\|)$ is a normed space and $\|x-y\|^p$ is negative definite then X is linearly isometric to a subset of L_p. Stated differently, X equipped with the metric $\|x-y\|^{p/2}$ is isometric to a subset of L_2 if and only if $(X,\|\cdot\|)$ is isometric to a subset of L_p.

Specializing the above discussion to the case $p=1$ we see that the Goemans-Linial conjecture is true for normed spaces. It is thus natural to ask if this phenomenon holds for arbitrary metric spaces. Moreover, from a computational viewpoint, since optimization problems over L_1 metrics are so important for the analysis of the cut structure of graphs, but are intractable computationally (see [17]), one might hope that the negative type property of L_1 metrics characterizes L_1 embeddability. This would reduce optimization problems over cuts to the Euclidean case, where we have efficient techniques, such as semidefinite programming, at our disposal. As we have seen, this hope fails. But as a motivation for the Goemans-Linial conjecture, the theorem of Bretagnolle, Dacunha-Castelle and Krivine is true for all $1<p<2$, and it is thus just as natural to ask if it holds for general metrics. In other words, if (X,d) is a metric space such that $(X,d^{p/2})$ is isometric to a subset of L_2, does X embed into L_p? The following theorem shows that the answer to this question is negative.

Theorem 1.2. *Fix $1\le p<2$ and define $d:\mathbb{R}^3\times\mathbb{R}^3\to[0,\infty)$ by*

$$d\Big((x,y,z),(t,u,v)\Big)=\left[\left((t-x)^2+(u-y)^2\right)^2+(v-z+2xu-2yt)^2\right]^{\frac14}$$
$$\cdot\left\{\cos\left[\frac{p}{2}\arccos\left(\frac{(t-x)^2+(u-y)^2}{\left[\left((t-x)^2+(u-y)^2\right)^2+(v-z+2xu-2yt)^2\right]^{\frac12}}\right)\right]\right\}^{1/p}.$$

Then d is a metric on $\mathbb{R}^3$ and $(\mathbb{R}^3,d^{p/2})$ is isometric to a subset of Hilbert space. But, $(\mathbb{R}^3,d)$ does not embed bi-Lipschitzly into L_p.

Note that when $p=1$ the metric in Theorem 2.2 is proportional to the metric in Theorem 1.1. Moreover, for all $1\le p<2$ the metric in Theorem 1.2 is easily seen to be bi-Lipschitz equivalent to the metric in Theorem 2.2 (we will prove this later). Thus the fact that it does not embed into

L_p follows from Theorem 1.1, since L_p is isometric to a subset of L_1. Nevertheless, the proof of the non-embeddability result in the case $1 < p < 2$ is significantly easier, and is based on more classical results.

An additional novel aspect of Theorem 1.1 is that the metric constructed there is decomposable. Recall that the modulus of padded decomposability of X is the smallest $\alpha > 0$ such that for every $\Delta > 0$ there exists a distribution over partitions of X into sets of diameter at most Δ, such that for every $x \in X$ with probability at least $\frac{1}{2}$ the entire ball of radius Δ/α centered at x is contained in the element of the partition to which x belongs (we refer to [33, 28] for a general discussion of this notion). We will see that the spaces $(\mathbb{R}^3, d)$ in Theorem 1.1 and Theorem 1.2 are doubling (see [23] for more information on the doubling condition), so that by the results of [21] they have a finite modulus of padded decomposability. It is a folklore problem, stated explicitly in [49], whether decomposable spaces embed into L_1. Theorem 1.1 shows that the answer to this question is negative. On the other hand the Khot-Vishnoi spaces are easily seen not to have a uniformly bounded modulus of padded decomposability.

Padded decomposability is a central tool in metric embeddings which is used in numerous contexts. In particular, a theorem of Klein, Plotkin and Rao [27] states that planar graphs, or more generally graphs which exclude a fixed minor, are decomposable. It is a famous conjecture (stated in [37, 22]) that such metrics embed into L_1. Our results show that if true, the proof of this conjecture will have to use more information than just the fact that such graph families are decomposable. This contrasts the fact that all known bi-Lipschitz embedding theorems for these spaces only use the fact that they are decomposable.

In the next subsections of this introduction we describe in greater detail the Heisenberg groups, and the ingredients of the proof of Theorem 1.1 and Theorem 1.2.

1.1 Heisenberg groups and the Koranyi norm

Let G be a group with identity element e. A function $N : G \to [0, \infty)$ is called a group semi-norm on G if $N(e) = 0$, every $g \in G$ satisfies $N(g^{-1}) = N(g)$, and every $g, h \in G$ satisfy $N(gh) \le N(g) + N(h)$. If, in addition, $N(g) = 0$ implies that $g = e$, then N is called a group norm on G. Observe that every group norm N on G induces a right-invariant metric ρ_N on G given by $\rho_N(g, h) = N(gh^{-1})$.

Fix an integer $n \ge 1$. For $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$ denote $|z| = \sqrt{|z_1|^2 + \ldots + |z_n|^2}$. Additionally, we will use the standard symplectic form on $\mathbb{C}^n$, defined for $z, w \in \mathbb{C}^n$ as:

$$[z, w] = \sum_{j=1}^{n} \Im(z_j \overline{w_j}),$$

where here and in what follows, $\Im(\zeta)$ and $\Re(\zeta)$ denote the imaginary part and real part of the complex number ζ, respectively.

The $2n+1$ dimensional Heisenberg group $\mathbb{H}^{2n+1}$ is defined via the following (non-commutative) group operation on $\mathbb{C}^n \times \mathbb{R}$:

$$(z, s) \cdot (w, t) = (z + w, t + s + 2[z, w]).$$

The Koranyi norm N_0 on $\mathbb{H}^{2n+1}$ is defined as

$$N_0(z, t) = \left(|z|^4 + t^2\right)^{1/4}.$$

It is a classical fact (see [15, 48]) that N_0 is indeed a group norm on $\mathbb{H}^{2n+1}$.

Remark 1.1. The Heisenberg groups are examples of the more general class of Carnot groups, which are themselves particular cases of a class of path metrics known as sub-Riemannian geometries. In particular, there is an important intrinsic geodesic metric on $\mathbb{H}^{2n+1}$, which is known as the Carnot-Carathéodory metric. We will not need to define this metric here, and it suffices to say that it is bi-Lipschitz equivalent the the metric induced by the Koranyi norm. However, understanding the proofs of the results that we will use requires familiarity with the Carnot-Carathéodory structure on $\mathbb{H}^{2n+1}$. Moreover, some of the results stated here carry over to general Carnot groups. We refer the interested reader to [20, 41] for more information about the fascinating world of sub-Riemannian geometries and Carnot groups.

Define $N : \mathbb{H}^{2n+1} \to [0, \infty)$ by

$$N(z,t) = \sqrt{\sqrt{|z|^4 + t^2} + |z|^2} = \sqrt{[N_0(z,t)]^2 + |z|^2}. \tag{2}$$

Since for any two group semi-norms N_1, N_2 on a group G, $\sqrt{N_1^2 + N_2^2}$ is also a group semi-norm on G, it follows that N is a group norm on $\mathbb{H}^{2n+1}$. Note that the metric d in Theorem 1.1 is precisely the metric induced on $\mathbb{H}^3$ by the group norm N. This proves that d is indeed a metric. The heart of Theorem 1.1 is the proof that $(\mathbb{H}^{2n+1}, \sqrt{\rho_N})$ is isometric to a subset of Hilbert space, and the non-embeddability of $\mathbb{H}^{2n+1}$ into L_1. We remark that it is easy to check that $(\mathbb{H}^{2n+1}, \rho_{N_0})$ is not of negative type. Our proof that $(\mathbb{H}^{2n+1}, \rho_N)$ is of negative type is naturally based on Schoenberg's theorem [45] quoted above, and is contained in Section 2. The fact that $\mathbb{H}^{2n+1}$ does not embed into L_1 follows from a recent result of Cheeger and Kleiner— this is explained in Section 1.3 below.

In order to prove Theorem 1.2 fix $p \in [1, 2)$ and define $M : \mathbb{H}^{2n+1} \to [0, \infty)$ by

$$M(z,t) = \left(|z|^4 + t^2\right)^{1/4} \left\{ \cos\left[\frac{p}{2} \arccos\left(\frac{|z|^2}{\sqrt{|z|^4 + t^2}}\right)\right]\right\}^{1/p}. \tag{3}$$

The following lemma implies that the metric d in Theorem 1.2 is indeed a metric, since it is precisely the metric on $\mathbb{H}^3$ induced by the group norm M.

Lemma 1.3. *M is a group norm on $\mathbb{H}^{2n+1}$ which satisfies for all $g \in \mathbb{H}^{2n+1}$,*

$$\sqrt{1 - \frac{p}{2}} \cdot N_0(g) \le \left[\cos\left(\frac{\pi p}{4}\right)\right]^{1/p} \cdot N_0(g) \le M(g) \le N_0(g). \tag{4}$$

Lemma 1.3 is proved in Section 3, as well as the fact that $(\mathbb{H}^{2n+1}, M^{p/2})$ is isometric to a subset of L_2 (the proof of which uses Schoenberg's theorem once more). The non-embeddability of $\mathbb{H}^{2n+1}$ into L_p when $p > 1$ follows from a simple extension of a differentiability theorem of Pansu [42] to the case of functions with values in Banach spaces with the Radon-Nikodým property, and an elegant observation of Semmes [46]. This is described in Section 1.2 below.

Remark 1.2. It is well known (see [23]) that $(\mathbb{H}^{2n+1}, N_0)$ is doubling with constant $O(1)^n$. By a theorem of Assouad [4, 23] we know that for every $\varepsilon \in (0, 1)$ the metric space $(\mathbb{H}^{2n+1}, N_0^{1-\varepsilon})$ is bi-Lipschitz equivalent to a subset of L_2 (with distortion depending on n and ε). Thus the main issue in Theorem 1.2 is to pass to an equivalent metric on $\mathbb{H}^{2n+1}$ for which the embedding is isometric.

As a by product we see from the bounds in (3) that $(\mathbb{H}^{2n+1}, N_0^{1-\varepsilon})$ embeds into Hilbert space with distortion *independent of* n. This distortion is $O\left(\frac{1}{\sqrt{\varepsilon}}\right)$, which coincides with the general bound for doubling spaces proved in [31].

1.2 Pansu differentiability and the Radon-Nikodým property

For every $\theta \in \mathbb{R}$ define the dilation operator $\delta_\theta : \mathbb{H}^{2n+1} \to \mathbb{H}^{2n+1}$ by $\delta_\theta(z,t) = (\theta z, \theta^2 t)$. Observe that for every $g, h \in \mathbb{H}^{2n+1}$, $\delta_\theta(gh) = (\delta_\theta g)(\delta_\theta h)$ and $\rho_{N_0}(\delta_\theta(g), \delta_\theta(h)) = |\theta| \rho_{N_0}(g,h)$. Let X be a Banach space and $F : \mathbb{H}^{2n+1} \to X$ be a Lipschitz mapping. We shall say that F has a Pansu derivative at $g \in \mathbb{H}^{2n+1}$ if for every $h \in \mathbb{H}^{2n+1}$ the limit

$$D_F^g(h) \equiv \lim_{\theta \to 0} \frac{F(\delta_\theta(h)g) - F(g)}{\theta}$$

exists, and D_F^g is a homomorphism, i.e. for all $h_1, h_2 \in \mathbb{H}^{2n+1}$, $D_F^g(h_1 h_2^{-1}) = D_F^g(h_1) - D_F^g(h_2)$.

Pansu [42] proved that if X is finite dimensional then every Lipschitz mapping $F : \mathbb{H}^{2n+1} \to X$ is Pansu-differentiable (Lebesgue) almost everywhere. We remark that Pansu's proof of this theorem extends almost verbatim to the case when X has the Radon-Nikodým property. A Banach space X is said to have the Radon-Nikodým property (RNP) if every Lipschitz function $f : \mathbb{R} \to X$ is differentiable almost everywhere. This is not the original definition of the Radon-Nikodým property, but it is equivalent to it and is most convenient for our purposes. We refer to Chapter 3 in [6] for more details. Examples of spaces without the RNP are L_1, c_0 and $C(0,1)$. On the other hand, separable conjugate Banach spaces and reflexive Banach spaces are known to have the RNP [6]. For example, since $\ell_1 = c_0^*$ and ℓ_1 is separable, it has the RNP.

Since Pansu's proof of his differentiability theorem into finite dimensional spaces uses only the differentiability of Lipschitz functions along geodesics, it extends to the case when X has the RNP (the modification of Pansu's argument is straightforward, and will not be included here). In fact, a very simple proof which is particularly convenient to extend to the Heisenberg case is the Preiss-Zajíček theorem [43] on the differentiability outside a σ-porous set of Lipschitz functions on $\mathbb{R}^n$ with values in RNP spaces (see also the second proof of Proposition 6.41 in [6]). Recall that for a metric space (Z,d) a set $A \subseteq M$ is called porous if there exists a number $0 < \lambda < 1$ such that for every $x \in A$ and every $\varepsilon > 0$ there is $y \in Z$ such that $0 < d(x,y) < \varepsilon$ and $A \cap B(y, \lambda d(x,y)) = \emptyset$. A countable union of porous sets is called σ-porous. Clearly porous sets are nowhere dense, so that σ-porous sets are of the first category. The Lebesgue density theorem implies that σ-porous sets on $\mathbb{H}^{2n+1}$ also have measure zero. Thus a slight strengthening of Pansu's differentiability theorem is that if X has the RNP then every Lipschitz function $F : \mathbb{H}^{2n+1} \to X$ is Pansu-differentiable outside a σ-porous set.

As noted by Semmes in [46], these observations imply that $\mathbb{H}^{2n+1}$ does not embed bi-Lipschitzly into any Banach space with the RNP. Indeed, if $F : \mathbb{H}^{2n+1} \to X$ were such an embedding then let $g \in \mathbb{H}^{2n+1}$ be a point of Pansu differentiability of F. For every $h, k \in \mathbb{H}^{2n+1}$ we have $D_F^g(hk) = D_F^g(h) + D_F^g(k) = D_F^g(kh)$. Thus

$$\begin{aligned} 0 = \lim_{\theta \to 0} \left\| \frac{F(\delta_\theta(hk)g) - F(g)}{\theta} - \frac{F(\delta_\theta(kh)g) - F(g)}{\theta} \right\|_X \\ \ge \frac{1}{\|F^{-1}\|_{\mathrm{Lip}}} \cdot \liminf_{\theta \to 0} \frac{\rho_{N_0}(\delta_\theta(hk)g, \delta_\theta(kh)g)}{\theta} \ge \frac{\rho_{N_0}(hk, kh)}{\|F^{-1}\|_{\mathrm{Lip}}}, \end{aligned}$$

and this is a contradiction since $\mathbb{H}^{2n+1}$ is not commutative.

Since for $p > 1$, L_p has the RNP, it follows that $\mathbb{H}^{2n+1}$ does not embed bi-Lipschitzly in L_p, proving the last part of Theorem 1.2.

Remark 1.3. Since ℓ_1 has the RNP, it follows from the above discussion that $\mathbb{H}^{2n+1}$ does not embed into ℓ_1. But this does not suffice to deduce the discretization statement in (1), since it is not true that if a metric space does not embed into ℓ_1 then its finite subsets cannot embed into ℓ_1 with uniformly bounded distortion. Indeed, ℓ_2 does not embed into ℓ_1 (see [35]), but all its finite subsets embed isometrically into ℓ_1. Another such example is L_1 itself. For the Goemans-Linial conjecture (and hence also for the integrality gap of the Sparsest Cut SDP), which originally deals with finite metrics, it is crucial, and sufficient, to prove non-embeddability into L_1, since L_1 has the property that if all the finite subsets of a separable metric space embed into it with uniformly bounded distortion, then so does the entire space (see [24]).

Remark 1.4. In [13] Cheeger and Kleiner extend Pansu's theorem in a different direction by generalizing Cheeger's theory of metric differentiation [10] to the case of mappings from certain metric measure spaces to a separable dual Banach space. The fact that $\mathbb{H}^{2n+1}$ does not embed into L_p, $p > 1$, follows from their results as well. In [12] they use this theory to give an example of a doubling metric space which embeds into L_1 but does not embed into ℓ_1. This result can be viewed as an extreme infinite version of the results [8, 32, 31] that there is no dimension reduction in L_1 in the spirit of the Johnson-Lindenstrauss dimension reduction lemma [25].

1.3 The Cheeger-Kleiner theorem on collapse towards the center

The crucial non-embeddability result in Theorem 1.1 does not follow from a differentiation statement, since L_1 does not have the RNP, and thus even Lipschitz mappings from the real line into L_1 might not have a point of differentiability (consider for example the mapping $t \mapsto \mathbf{1}_{[0,t]}$). Moreover, differentiability theorems cannot hold for mappings into general Banach spaces since $\mathbb{H}^{2n+1}$ is isometric to a subset of L_∞, and this isometry cannot have a derivative in any reasonable sense. This problem is a long standing major obstacle to proving L_1 non-embeddability. In a beautiful tour-de-force paper [11] Cheeger and Kleiner proved that in a certain sense there is a weak notion of differentiability for mappings from $\mathbb{H}^{2n+1}$ into L_1. This notion is strong enough to prove that $\mathbb{H}^{2n+1}$ does not embed into L_1. We will not state here the exact (somewhat complicated) formulation of the Cheeger-Kleiner L_1 differentiation result— instead we will state its main corollary. They show that if $U \subseteq \mathbb{H}^{2n+1}$ is an open subset, and $F : U \to L_1$ is a Lipschitz function, then for almost every $(z,t) \in U$ we have

$$\lim_{\varepsilon \to 0^+} \frac{\|F(z,t+\varepsilon) - F(z,t)\|_1}{\sqrt{\varepsilon}} = 0. \tag{5}$$

This implies that F is not bi-Lipschitz. Indeed, otherwise we would have for every $(z,t) \in U$ and $\varepsilon > 0$,

$$\|F(z,t+\varepsilon) - F(z,t)\|_1 \geq \frac{1}{\|F^{-1}\|_{\mathrm{Lip}}} \cdot \rho_{N_0}((z,t+\varepsilon),(z,t)) = \frac{1}{\|F^{-1}\|_{\mathrm{Lip}}} \cdot \sqrt{\varepsilon}.$$

The statement in (5) says that Lipschitz maps on $\mathbb{H}^{2n+1}$ "collapse" in the direction of the center of $\mathbb{H}^{2n+1}$ at almost every point, and thus they cannot be bi-Lipschitz. This in itself can be viewed as a weak differentiation result. The proof of (5) is quite remarkable. Cheeger and Kleiner start by using the cut decomposition on the image of F to induce a family of cuts on $\mathbb{H}^{2n+1}$. They observe that in a certain sense most of these cuts are given by subsets of $\mathbb{H}^{2n+1}$ with finite perimeter. They then apply recent results in geometric measure theory on the fine structure of subsets of $\mathbb{H}^{2n+1}$ with finite perimeter, and use this structural information to prove (5). This geometric investigation of the type of cuts that can appear in the cut decomposition of certain spaces is a novel approach to L_1 non-embeddability and L_1 differentiation, and we expect that it will have additional applications in the future.

Remark 1.5. A natural question that occurs is what is the rate with which the distortion tends to infinity in (1). It is not difficult to see that explicit rates would follow from a version of (5) with an explicit rate of convergence to 0. We discuss this below in Remark 1.6. This rate at which (5) tends to 0 is the topic of a work in progress of Cheeger, Kleiner and the second named author [14]. Heuristic considerations suggest that the rate in (1) might be $c_1\left(\{0,\ldots,n\}^3, d\right) \geq (\log n)^{\Omega(1)}$. This should follow from a technical (albeit tedious) "quantification" of the arguments in [11]. It is clearly of interest to find a simpler and shorter argument— this is intended to be the main focus of [14]. We do not report a specific exponent here since the quantitative version of (5) is quite long and involved, and it is not clear at this point what are the precise rates that are obtained. One of the reasons for this complication is that a quantitative version of (5) involves proving effective versions of the results from geometric measure theory that were used in [11]. Such effective versions were not previously known, and they are of independent interest.

We remark that the Khot-Vishnoi example in [26] gives an n-point metric space X_n of negative type such that $c_1(X_n) \geq (\log\log n)^{\frac{1}{6}-o(1)}$. This was improved recently by Krauthgamer and Rabani [29] to a lower bound of $\Omega(\log\log n)$. The best known upper bounds for the Euclidean distortion (and, hence, L_1 distortion) of n-point negative type metrics are due to Arora, Lee, and Naor [2], who proved that such metrics embed into L_2 with distortion $O\left(\sqrt{\log n}\cdot\log\log n\right)$. A recent paper of the first named author yields a small improvement to $O\left(\sqrt{\log n \log\log n}\right)$ [30]. The result of Khot and Vishnoi is based on the analysis of a hardness result which they proved (in the same paper) for the Sparsest Cut problem with general demands, assuming the unique games conjecture (such a hardness result was also obtained in [9]). Improving this (conditional) hardness lower bound to $(\log n)^{\Omega(1)}$ remains an important challenge.

The best known integrality gap for the Goemans-Linial semidefinite relaxation of Sparsest Cut with uniform demands, due to Arora, Rao and Vazirani [3], is $O\left(\sqrt{\log n}\right)$. Very recently Devanur, Khot, Saket and Vishnoi [16] showed that this integrality gap also tends to infinity with the number of vertices of the graph. We remark that doubling metrics, and more generally decomposable metrics, cannot yield an integrality gap for the SDP relaxation of uniform Sparsest Cut due to the results of Rabinovich [44].

Remark 1.6. We now show the relationship between the rate of convergence in (5) and quantitative lower bounds on $c_1(X)$ for finite subsets $X \subseteq \mathbb{H}^3$. Let $B(0,r)$ denote the open ball of radius r centered at 0 in $(\mathbb{H}^3, \rho_{N_0})$. Fix $\delta \in (0,\frac{1}{2})$ and let $X \subseteq B(0,1)$ be a δ-net in $B(0,1)$. It is easily checked that $|X| \leq (4/\delta)^4$.

Assume that for some non-decreasing function $R : (0,\frac{1}{4}) \to [0,1]$ with $1 \geq R(\varepsilon) \geq \sqrt{\varepsilon}$ for every $\varepsilon \in (0,\frac{1}{4})$, the following assertion holds true: For every 1-Lipschitz map $F : B(0,1) \to L_1$ there exists a point $(z,t) \in B(0,\frac{1}{2})$ such that

$$\|F(z,t+\varepsilon) - F(z,t)\|_1 \leq R(\varepsilon)\sqrt{\varepsilon}. \tag{6}$$

We claim that by choosing $\delta = \frac{1}{4}\sqrt{\varepsilon}\cdot R(\varepsilon)$ one has

$$c_1(X) \geq \frac{c}{R\left(16\,|X|^{-1/4}\right)}, \tag{7}$$

where $c > 0$ is a universal constant.

Indeed, let $f : X \to L_1$ be a bi-Lipschitz map with $\|f\|_{\mathrm{Lip}} \leq 1$. Since $\mathbb{H}^3$ is doubling, by [33, Th. 1.6] there exists a universal constant $K \geq 1$ (independent of f) and a map $\widetilde{f} : \mathbb{H}^3 \to L_1$ for which $\|\widetilde{f}\|_{\mathrm{Lip}} \leq K$ and $\tilde{f}|_X = f$. Let $F = \frac{1}{K}\tilde{f}$, so that $\|F\|_{\mathrm{Lip}} \leq 1$. Let $(z,t) \in B(0,\frac{1}{2})$ be the point which results from applying (6) to F, and let $(z_1,t_1),(z_2,t_2) \in X$ be such that $\rho_{N_0}((z_1,t_1),(z,t)) \leq \delta$ and $\rho_{N_0}((z_2,t_2),(z,t+\varepsilon)) \leq \delta$. Then

$$\begin{aligned} R(\varepsilon)\sqrt{\varepsilon} \geq \|F(z,t+\varepsilon)-F(z,t)\|_1 &\geq \|F(z_1,t_1)-F(z_2,t_2)\|_1 - 2\delta \\ &\geq \frac{\rho_{N_0}((z_1,t_1),(z_2,t_2))}{K\cdot\|f^{-1}\|_{\mathrm{Lip}}} - 2\delta \geq \frac{\sqrt{\varepsilon}-2\delta}{K\cdot\|f^{-1}\|_{\mathrm{Lip}}} - 2\delta. \end{aligned}$$

Equivalently,

$$\|f^{-1}\|_{\mathrm{Lip}} \geq \frac{1}{K}\cdot\frac{\sqrt{\varepsilon}-2\delta}{R(\varepsilon)\sqrt{\varepsilon}+2\delta} \geq \frac{1}{3K\cdot R(\varepsilon)}.$$

Since $|X| \leq (4/\delta)^4 \leq (16/\varepsilon)^4$, we conclude that (7) holds. Thus one obtains a bound of the form $c_1(X) \geq (\log|X|)^{\Omega(1)}$ as long as $R(\varepsilon)^{-1} \geq (\log(\frac{1}{\varepsilon}))^{\Omega(1)}$ holds in (6).

2 A metric of negative type on the Heisenberg group

Let G be a group with identity element e. A a complex valued function $K : G\times G \to \mathbb{C}$ is called a Hermitian kernel on G if every $g,h \in G$ satisfy $K(g,h) = \overline{K(h,g)}$. A Hermitian kernel K on G is said to be positive definite if

$$\sum_{\ell,m=1}^{n} K(g_\ell,g_m)c_\ell\overline{c_m} \geq 0$$

for all $g_1,\ldots,g_n \in G$ and for all complex scalars $c_1,\ldots,c_n \in \mathbb{C}$. A Hermitian kernel K on G is called negative definite if

$$\sum_{\ell,m=1}^{n} K(g_\ell,g_m)c_\ell\overline{c_m} \leq 0$$

for all $g_1,\ldots,g_n \in G$ and for all complex scalars $c_1,\ldots,c_n \in \mathbb{C}$ satisfying $\sum_{j=1}^n c_j = 0$.

Let $F : G \to \mathbb{C}$ be a function such that every $g \in G$ satisfies $F(g^{-1}) = \overline{F(g)}$. The Hermitian kernel on G induced by F, denoted K_F, is defined by $K_F(g,h) = F(gh^{-1})$. F is said to be positive definite (resp. negative definite) if K_F is positive definite (resp. negative definite).

We will use the following classical fact, due to Scoenberg [45] (see also the book [50] and Proposition 8.5 in [6]).

Proposition 2.1. *Let $K : G\times G \to \mathbb{R}$ be a real-valued kernel on G satisfying $K(g,g) = 0$ for every $g \in G$. Then K is negative definite if and only if there exists a Hilbert space H and a function $T : G \to H$ such that for all $g,h \in G$,*

$$K(g,h) = \|T(g)-T(h)\|^2.$$

The main result of this section is:

Theorem 2.2. *Let N be as in (2). Then $(\mathbb{H}^{2n+1}, \rho_N)$ is a metric space of negative type, i.e. $(\mathbb{H}^{2n+1}, \rho_N)$ is a metric space and $(\mathbb{H}^{2n+1}, \sqrt{\rho_N})$ embeds isometrically in Hilbert space.*

By Proposition 2.1 all that remains is to show that N is a negative definite function on $\mathbb{H}^{2n+1}$, and the remainder of this section is devoted to the proof of this fact. We remark that it is easy to verify that the Koranyi norm itself in not negative definite on $\mathbb{H}^{2n+1}$.

Lemma 2.3. *For every $\lambda \in \mathbb{C}$ define $\Phi_\lambda : \mathbb{H}^{2n+1} \to \mathbb{R}$ by*

$$\Phi_\lambda(z,t) = e^{-|\lambda|\cdot|z|^2 + i\lambda t}.$$

Then Φ_λ is a positive definite function on $\mathbb{H}^{2n+1}$.

Proof. We have

$$\begin{aligned} K_{\Phi_\lambda}((z,s),(w,t)) &= \Phi_\lambda((z,s)\cdot(-w,-t)) \\ &= \exp\left[-|\lambda|\sum_{j=1}^n |z_j - w_j|^2 + i\lambda(s - t - 2\sum_{j-1}^n \Im(z_j\overline{w_j}))\right] \\ &= \left(\prod_{j=1}^n e^{-|\lambda|(|z_j|^2+|w_j|^2)}\right)\cdot e^{i\lambda(s-t)}\cdot\left(\prod_{j=1}^n e^{2|\lambda|(\Re(z_j\overline{w_j}) - i\cdot\mathrm{sign}(\lambda)\Im(z_j\overline{w_j}))}\right). \end{aligned}$$

Since the point-wise product of positive definite kernels is positive definite (see [6], Proposition 8.2), it suffices to show that each term in the above product is a positive definite kernel on $\mathbb{H}^{2n+1}$. The fact that $e^{i\lambda(s-t)}$ and $e^{-|\lambda|(|z_j|^2+|w_j|^2)}$ are positive definite follows the fact that for all complex scalars $c_1, \ldots, c_k \in \mathbb{C}$, the matrix $(c_\ell\overline{c_m})_{\ell,m}$ is positive semidefinite. It remains to check that $e^{2|\lambda|(\Re(z_j\overline{w_j}) - i\cdot\mathrm{sign}(\lambda)\Im(z_j\overline{w_j}))}$ is positive definite. Since, for any positive definite kernel K, the kernel e^K is also positive definite (see [6], Proposition 8.2), it is enough to show that

$$\Re(z_j\overline{w_j}) - i\cdot\mathrm{sign}(\lambda)\Im(z_j\overline{w_j})$$

is positive definite. This equals $z_j\overline{w_j}$ if $\lambda < 0$ and $\overline{z_j\overline{w_j}} = \overline{z_j}w_j$ if $\lambda \geq 0$. In both cases, the kernel is positive definite. □

Proof of Theorem 2.2. Our goal is to show that N is a negative definite function on $\mathbb{H}^{2n+1}$. In what follows, we use some notions from Fourier analysis. For a function $f : \mathbb{R} \to \mathbb{R}$, we denote its Fourier transform by $\widehat{f}(t) = \int_{\mathbb{R}} e^{itx} f(x)dx$. The convolution of two functions $f, g : \mathbb{R} \to \mathbb{R}$ is defined as $(f * g)(t) = \int_{\mathbb{R}} f(t-x)g(x)dx$, so that $\widehat{f * g} = \widehat{f}\cdot\widehat{g}$.

Fix $\varepsilon > 0$. The existence of symmetric $\frac{1}{2}$-stable distributions (see [18]) implies that there exits a non-negative integrable function $\varphi_\varepsilon : \mathbb{R} \to [0,\infty)$ such that, for all $t \in \mathbb{R}$, $\widehat{\varphi}_\varepsilon(t) = e^{-\varepsilon\sqrt{|t|}}$. Lemma 2.3 shows that the function $F_\varepsilon : \mathbb{H}^{2n+1} \to \mathbb{C}$ given by

$$F_\varepsilon(z,t) = \int_{\mathbb{R}} e^{-|\lambda|\cdot|z|^2 + i\lambda t}\varphi_\varepsilon(\lambda)\, d\lambda$$

is positive definite on $\mathbb{H}^{2n+1}$. For every $a > 0$, denote

$$h_a(x) = \frac{a}{\pi} \cdot \frac{1}{a^2 + x^2}.$$

Then $\int_{\mathbb{R}} \widehat{h}_a(x)dx = 1$ and $\widehat{h}_a(t) = e^{-a|t|}$ for all $t \in \mathbb{R}$ (see, e.g. [18]). Denoting $f_\varepsilon(t) = e^{-\varepsilon\sqrt{|t|}}$, the inversion formula for the Fourier transform implies that $\varphi_\varepsilon = \frac{1}{2\pi}\widehat{f}_\varepsilon$. Another application of the inversion formula gives

$$F_\varepsilon(z,t) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\lambda t}\widehat{h}_{|z|^2}(\lambda)\widehat{f}_\varepsilon(\lambda)d\lambda = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\lambda t}(\widehat{h_{|z|^2} * f_\varepsilon})(\lambda)d\lambda = (h_{|z|^2} * f_\varepsilon)(t).$$

Since F_ε is positive definite on $\mathbb{H}^{2n+1}$, the function $\frac{1-F_\varepsilon}{\varepsilon}$ is negative definite on $\mathbb{H}^{2n+1}$, and

$$\begin{aligned} \lim_{\varepsilon\to 0}\frac{1-F_\varepsilon(z,t)}{\varepsilon} &= \lim_{\varepsilon\to 0}\left[h_{|z|^2} * \left(\frac{1-f_\varepsilon}{\varepsilon}\right)\right](t) \\ &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}}\frac{1-e^{-\varepsilon\sqrt{|x|}}}{\varepsilon}h_{|z|^2}(t-x)\,dx \\ &= \frac{|z|^2}{\pi}\int_{\mathbb{R}}\frac{\sqrt{|x|}}{|z|^4+(t-x)^2}\,dx. \end{aligned}$$

The next lemma shows that the latter expression is equal to $N(z,t)$, completing the proof. □

Lemma 2.4. *For every $r, t \in \mathbb{R}$*

$$\frac{r^2}{\pi}\int_{\mathbb{R}}\frac{\sqrt{|x|}}{r^4+(t-x)^2}dx = \sqrt{\sqrt{r^4+t^2}+r^2}.$$

Proof. Making the change of variable $x = r^2 v$ and $s = t/r^2$, our goal is to prove that

$$\int_0^\infty\left(\frac{1}{1+(s-v)^2}+\frac{1}{1+(s+v)^2}\right)\sqrt{v}\,dv = \pi\sqrt{\sqrt{1+s^2}+1}. \tag{8}$$

By continuity, we may assume that $s \neq 0$. Consider the function $\psi : \{\zeta :\ \Im(\zeta) > 0\} \to \mathbb{C}$ given by $\psi(\zeta) = \left(\frac{1}{1+(s-\zeta)^2}+\frac{1}{1+(s+\zeta)^2}\right)\sqrt{\zeta}$. We take here the principle branch of the square root, i.e. if $\zeta = \rho e^{i\theta}$ where $\rho > 0$ and $\theta \in (0,\pi)$ then $\sqrt{\zeta} = \sqrt{\rho}e^{i\theta/2}$. The poles of ψ are at $i \pm s$, so that, by the residue theorem, the left-hand side of (8) equals $\Re\left[2\pi i \mathrm{Res}_{i+s}(\psi) + 2\pi i \mathrm{Res}_{i-s}(\psi)\right]$. Since the poles of ψ are simple, a direct computation gives that the required integral equals

$$\Re\left[2\pi i\left(\frac{\sqrt{i+s}}{2i}+\frac{\sqrt{i-s}}{2i}\right)\right] = \pi\left(\sqrt{\frac{\sqrt{s^2+1}+s}{2}}+\sqrt{\frac{\sqrt{s^2+1}-s}{2}}\right) = \pi\sqrt{\sqrt{1+s^2}+1},$$

which is the required identity. □

3 The L_p case

We now generalize the argument of Section 2.2 to prove Theorem 1.2. The idea of the proof is the same, and we will therefore use the same notation and be sketchy at some places. In what follows we fix $1 \le p < 2$.

Fix $\varepsilon > 0$ and let $\varphi_\varepsilon : \mathbb{R} \to [0, \infty)$ satisfy for all $t \in \mathbb{R}$, $\widehat{\varphi}_\varepsilon(t) = e^{-\varepsilon|t|^{p/2}}$. The existence of φ_ε follows from the existence of a symmetric $\frac{p}{2}$-stable distribution [18]. Define $F_\varepsilon : \mathbb{H}^{2n+1} \to \mathbb{C}$ by

$$F_\varepsilon(z,t) = \int_{\mathbb{R}} e^{-|\lambda|\cdot|z|^2 + i\lambda t} \varphi_\varepsilon(\lambda)\, d\lambda,$$

which is a positive definite function on $\mathbb{H}^{2n+1}$ by Lemma 2.3. As before, we write $h_a(x) = \frac{a}{\pi} \cdot \frac{1}{a^2+x^2}$ and $f_\varepsilon(t) = e^{-\varepsilon|t|^{p/2}}$. Arguing as in the proof of Theorem 2.2 we obtain the identity

$$\lim_{\varepsilon \to 0} \frac{1 - F_\varepsilon(z,t)}{\varepsilon} = \frac{|z|^2}{\pi} \int_{\mathbb{R}} \frac{|x|^{p/2}}{|z|^4 + (t-x)^2}\, dx.$$

Thus the mapping

$$(z,t) \mapsto \frac{|z|^2}{\pi} \int_{\mathbb{R}} \frac{|x|^{p/2}}{|z|^4 + (t-x)^2}\, dx \tag{9}$$

is negative definite on $\mathbb{H}^{2n+1}$. This integral is calculated in the following lemma.

Lemma 3.1. *For every $r, t \in \mathbb{R}$*

$$\frac{r^2}{\pi} \int_{\mathbb{R}} \frac{|x|^{p/2}}{r^4 + (t-x)^2} dx = 2 \cos\left(\frac{p\pi}{4}\right) \cdot (r^4 + t^2)^{p/4} \cdot \cos\left[\frac{p}{2} \arccos\left(\frac{r^2}{\sqrt{r^4 + t^2}}\right)\right].$$

Proof. Making the change of variable $x = r^2 v$ and $s = t/r^2$ we find that:

$$\frac{r^2}{\pi} \int_{\mathbb{R}} \frac{|x|^{p/2}}{r^4 + (t-x)^2} dx = \frac{r^p}{\pi} \int_0^\infty \left(\frac{1}{1 + (s-v)^2} + \frac{1}{1 + (s+v)^2}\right) v^{p/2} dv.$$

As in Lemma 2.4 we define $\psi : \{\zeta :\ \Im(\zeta) > 0\} \to \mathbb{C}$ by $\psi(\zeta) = \left(\frac{1}{1+(s-\zeta)^2} + \frac{1}{1+(s+\zeta)^2}\right) \zeta^{p/2}$, where if $\zeta = \rho e^{i\theta}$ for $\rho > 0$ and $\theta \in (0, \pi)$ then $\zeta^{p/2} = \rho^{p/2} e^{ip\theta/2}$. By the residue theorem,

$$\begin{aligned}\int_0^\infty \left(\frac{1}{1 + (s-v)^2} + \frac{1}{1 + (s+v)^2}\right) v^{p/2} dv &= \Re\left[2\pi i \left(\frac{(i+s)^{p/2}}{2i} + \frac{(i-s)^{p/2}}{2i}\right)\right] \\ &= 2\pi \cos\left(\frac{p\pi}{4}\right) \cdot \left(1 + s^2\right)^{p/4} \cdot \cos\left[\frac{p}{2} \arccos\left(\frac{1}{\sqrt{1+s^2}}\right)\right].\end{aligned}$$

This implies the required identity. □

We have thus shown that if M is as in (3) then $(\mathbb{H}^{2n+1}, M^{p/2})$ is isometric to a subset of L_p (since the integral in (9) equals M^p). In order to prove Theorem 1.2 it remains to prove Lemma 1.3. We begin with the following standard lemma, whose simple proof we include for the sake of completeness.

Lemma 3.2. *Denote $C = \{(a,b) \in \mathbb{R}^2 \setminus \{0\} : \ a \ge b \ge 0\}$. Assume that $1 < \beta$ and $\gamma : [1, \beta] \to [0, \infty]$ is continuously differentiable, concave, $\gamma(1) = 1$, $\gamma(\beta) = 0$, and $\gamma'(1) < 0$. Then for every $(a,b) \in C$ there is a unique $\mu = \mu(a,b) \in [a/\beta, a]$ such that $\frac{b}{\mu} = \gamma\left(\frac{a}{\mu}\right)$. Moreover, $\mu : C \to [0, \infty)$ satisfies for every $x, y \in C$, $\mu(x+y) \le \mu(x) + \mu(y)$, and if $x \le y$ coordinate-wise then $\mu(x) \le \mu(y)$.*

Proof. The assumptions imply that γ is strictly decreasing. To prove the existence of μ take $(a,b) \in C$ and for $\mu \in [a/\beta, a]$ let $f(\mu) = \mu\gamma\left(\frac{a}{\mu}\right)$. Direct differentiation shows that γ is strictly increasing. Moreover $f\left(\frac{a}{\beta}\right) = \frac{a}{\beta}\gamma(\beta) = 0 \le b$ and $f(a) = a\gamma(1) = a$. Thus there exists a unique μ for which $f(\mu) = b$.

If $x, y \in C$ write $x = (x_1, x_2)$, $y = (y_1, y_2)$ and denote $A = \mu(x) + \mu(y)$. Then by the concavity of γ,

$$\gamma\left(\frac{x_1 + y_1}{\mu(x) + \mu(y)}\right) \ge \frac{\mu(x)}{\mu(x) + \mu(y)}\gamma\left(\frac{x_1}{\mu(x)}\right) + \frac{\mu(y)}{\mu(x) + \mu(y)}\gamma\left(\frac{y_1}{\mu(y)}\right) = \frac{x_2 + y_2}{\mu(x) + \mu(y)},$$

and from the above reasoning we deduce that $\mu(x) + \mu(y) \ge \mu(x+y)$. Assume now that $x_1 \le y_1$ and $x_2 \le y_2$. Then denoting $g(t) = t\gamma\left(\frac{x_1}{t}\right)$ for $t \in [x_1/\beta, x_1]$, we know that g is increasing and $g\left(\mu(x_1, x_2)\right) = x_2 \le y_2 = g\left(\mu(x_1, y_2)\right)$. Thus $\mu(x_1, x_2) \le \mu(x_1, y_2)$. Assume for the sake of contradiction that $\mu(y_1, y_2) < \mu(x_1, y_2)$. Then $\gamma\left(\frac{x_1}{\mu(x_1,y_2)}\right) = \frac{y_2}{\mu(x_1,y_2)} < \frac{y_2}{\mu(y_1,y_2)} = \gamma\left(\frac{y_1}{\mu(y_1,y_2)}\right)$. Since γ is decreasing we deduce that $\frac{y_1}{\mu(x_1,y_2)} \ge \frac{x_1}{\mu(x_1,y_2)} > \frac{y_1}{\mu(y_1,y_2)}$, which is a contradiction. □

The following lemma is the crucial step in proving that M is indeed a group norm on $\mathbb{H}^{2n+1}$.

Lemma 3.3. *Define $\gamma : \left[1, \left[\cos\left(\frac{\pi p}{4}\right)\right]^{-1/p}\right] \to [0, \infty)$ by*

$$\gamma(a) = a\sqrt{\cos\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right)}.$$

Then γ satisfies the conditions of Lemma 3.2.

Before proving Lemma 3.3 we use it to prove Lemma 1.3.

Proof of Lemma 1.3. Let $\mu : C \to [0, \infty)$ be the function from Lemma 3.2 which corresponds to the function γ in Lemma 3.3. Then for $(a,b) \in C$ we have

$$\frac{b}{\mu(a,b)} = \gamma\left(\frac{a}{\mu(a,b)}\right) = \frac{a}{\mu(a,b)}\sqrt{\cos\left(\frac{2}{p}\arccos\left(\frac{\mu(a,b)^p}{a^p}\right)\right)}.$$

Solving this equation we see that

$$\mu(a,b) = |a|\left\{\cos\left[\frac{p}{2}\arccos\left(\frac{b^2}{a^2}\right)\right]\right\}^{1/p}.$$

Thus $M(z,t) = \mu(N_0(z,t), |z|)$. It follows that for every $(z,t), (\zeta,\tau) \in \mathbb{H}^{2n+1}$ we have

$$\begin{aligned} M\left((z,t)\cdot(\zeta,\tau)\right) &= \mu\left(N_0\left((z,t)\cdot(\zeta,\tau)\right), |z+\zeta|\right) \\ &\le \mu\left(N_0\left(z,t\right) + N_0(\zeta,\tau), |z| + |\zeta|\right) \\ &\le \mu(N_0(z,t), |z|) + \mu(N_0(\zeta,\tau), |\zeta|) \\ &= M(z,t) + M(\zeta,\tau). \end{aligned}$$

The right-hand inequality in (4) follows from the fact that $\arccos\left(\frac{|z|^2}{\sqrt{|z|^4+t^2}}\right) \in [0,\pi/2]$, and the left-hand inequality in (4) is an easy elementary numerical inequality which we do not prove here (actually we just care about the asymptotic behavior as p tend to 2, which is obvious). $\square$

Before passing to the proof of Lemma 3.3 we record some elementary numerical inequalities that will be used in the ensuing argument.

Lemma 3.4 (Auxiliary numerical inequalities that will be used later). *The following numerical inequalities hold true in the specified ranges*

1. $\tan(\lambda t) \geq \lambda \tan t$ *for* $\lambda \in [1,\infty)$ *and* $t \in \left[0, \frac{\pi}{2\lambda}\right)$.
2. $\frac{4x}{\pi} \leq \max\left\{1, 3\sin^2 x\right\}$ *for* $x \in [0,1]$.
3. $2\sin u - u\cos u - u \geq 0$ *for* $u \in [0,\pi]$.
4. $\sin(4x) - 4x + 12x\sin^2 x \geq 0$ *for* $0 \leq x \leq \arcsin\left(\frac{1}{\sqrt{3}}\right)$.

Proof. The first inequality is simply a consequence of the convexity of the function $\tan(\cdot)$ on $[0,\pi/2)$. To prove the second inequality note that if $\frac{4x}{\pi} \leq 1$ then there is nothing to prove, so assume that $x \in [\pi/4, 1]$. We must show that in this range $3\sin^2 x \geq \frac{4x}{\pi}$. Using the elementary inequality $\sin x \geq x - \frac{x^3}{6}$ we see that it is enough to prove the inequality $x\left(1 - \frac{x^2}{6}\right)^2 \geq \frac{4\pi}{3}$, which is valid for $x \in [\pi/4, 1]$. To prove the third inequality write $\psi(u) = 2\sin u - u\cos u - u$. Then $\psi''(u) = u\cos u$. So ψ' is increasing on $[0,\pi/2]$ and decreasing on $[\pi/2,\pi]$. But $\psi'(0) = 0$, so that ψ' is either always non-negative or first non-negative and then negative. Thus ψ is either increasing or first increasing and then decreasing. In both cases the minimum of ψ is attained at one of the endpoints $\{0,\pi\}$, where its value is 0.

It remains to prove the fourth inequality. To this end let $s = \sin x$, so that $0 \leq s \leq \frac{1}{\sqrt{3}}$. Note that $\sin(4x) = 4\sin x\cos x\left(1 - 2\sin^2 x\right) = 4s(1-2s^2)\sqrt{1-s^2}$. Thus the required inequality becomes $4s(1-2s^2)\sqrt{1-s^2} - 4x\left(1-3s^2\right) \geq 0$, or

$$\arcsin s \leq \frac{s(1-2s^2)\sqrt{1-s^2}}{1-3s^2}. \tag{10}$$

To prove (10) define $\theta(s) = \frac{s(1-2s^2)\sqrt{1-s^2}}{1-3s^2} - \arcsin s$. Direct (but tedious) differentiation gives

$$\theta'(s) = \frac{s^2(1+5s^2-12s^4)}{(1-3s^2)^2\sqrt{1-s^2}}.$$

Since $1 + 5s^2 - 12s^4 \geq 0$ when $0 \leq s \leq \frac{1}{\sqrt{3}}$ it follows that $\theta'(s) \geq 0$. Thus $\theta(s) \geq \theta(0) = 0$, which proves (10). $\square$

Proof of Lemma 3.3. If $g = \gamma^2$ satisfies the conditions of Lemma 3.3 then so does γ. One checks that $g'(1) = 2 - \frac{4}{p} < 0$, so that it is enough to show that $g'' \leq 0$. Direct differentiation yields

$$g''(a) = \frac{2(a^{2p}-3)}{a^{2p}-1}\cos\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) + \frac{2[a^{2p}(p-3)+3]}{(a^{2p}-1)^{3/2}}\sin\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right).$$

Note that in the range $1 \le a \le \left[\cos\left(\frac{\pi p}{4}\right)\right]^{-1/p}$ we have $\cos\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) \ge 0$. Thus the inequality $g''(a) \le 0$ is equivalent to

$$\left[a^{2p}(3-p)-3\right]\tan\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) \ge (a^{2p}-3)\sqrt{a^{2p}-1}. \tag{11}$$

We distinguish between two cases. If $a^{2p}(3-p)-3>0$ then we must show that

$$\tan\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) \ge \frac{(a^{2p}-3)\sqrt{a^{2p}-1}}{a^{2p}(3-p)-3}. \tag{12}$$

Observe that we are assuming that $\frac{2}{p}\arccos\left(\frac{1}{a^p}\right) \le \frac{\pi}{2}$. Hence by the first inequality in Lemma 3.4 (with $\lambda = \frac{2}{p} \ge 1$) we see that

$$\tan\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) \ge \frac{2}{p}\tan\left(\arccos\left(\frac{1}{a^p}\right)\right) = \frac{2}{p}\cdot\frac{\sqrt{1-a^{-2p}}}{a^{-p}} = \frac{2}{p}\cdot\sqrt{a^{2p}-1}.$$

Thus (12) is equivalent to $a^{2p} \ge 1$, as required.

It remains to deal with the case $a^{2p}(3-p)-3<0$, which is equivalent to $a^{2p} < \frac{3}{3-p}$. In this case we need to show that

$$\tan\left(\frac{2}{p}\arccos\left(\frac{1}{a^p}\right)\right) \le \frac{(3-a^{2p})\sqrt{a^{2p}-1}}{3-a^{2p}(3-p)}. \tag{13}$$

Write $x = \arccos\left(\frac{1}{a^p}\right)$, so that by our assumption $x \le \frac{\pi p}{4}$. Then the required inequality becomes

$$\tan\left(\frac{2x}{p}\right) \le \frac{2-3\sin^2 x}{p-3\sin^2 x}\cdot\tan x, \tag{14}$$

where the condition $a^{2p} < \frac{3}{3-p}$ translates to $p > 3\sin^2 x$. For fixed x the range of p for which (14) should hold is $2 \ge p \ge \max\left\{1, \frac{4x}{\pi}, 3\sin^2 x\right\}$. If this range is non-empty then $3\sin^2 x \le 2$, i.e. $x \le \arcsin\left(\sqrt{2/3}\right) < 1$. We will therefore assume from now on that this upper bound on x is satisfied, in which case the second inequality in Lemma 3.4 implies that (14) should hold for $2 \ge p \ge \max\left\{1, 3\sin^2 x\right\}$.

Denote $A(p) = (p-3\sin^2 x)\tan\left(\frac{2x}{p}\right)$. We want to show that $A(p) \le A(2)$. Now,

$$A'(p) = \tan\left(\frac{2x}{p}\right) - \frac{2x(p-3\sin^2 x)}{p^2\cos^2\left(\frac{2x}{p}\right)}.$$

It is enough to show that $A'(p) \ge 0$. Clearing the denominator and simplifying we see that it is enough to prove that

$$B(p) = \frac{p^2}{2}\sin\left(\frac{4x}{p}\right) - 2x(p-3\sin^2 x) \ge 0. \tag{15}$$

Now,

$$B'(p) = p\sin\left(\frac{4x}{p}\right) - 2x\cos\left(\frac{4x}{p}\right) - 2x.$$

We claim that $B'(p) \geq 0$. Denoting $u = \frac{4x}{p} \leq \pi$ we see that this reduces to the third inequality in Lemma 3.4. So $B'(p) \geq 0$ and hence $B(p) \geq B(p_0)$, where $p_0 = \max\{1, 3\sin^2 x\}$ and we are assured that $2 \geq p_0 \geq \frac{4x}{\pi}$. If $p_0 = 3\sin^2 x$ then $B(p_0) = \frac{p_0^2}{2}\sin\left(\frac{4x}{p_0}\right) \geq 0$, since $\frac{4x}{p_0} \leq \pi$. We therefore assume that $p_0 = 1$, in which case we know that $0 \leq x \leq \arcsin\left(\frac{1}{\sqrt{3}}\right)$. But $B(p_0) = B(1) = \frac{1}{2}\sin(4x) - 2x + 6x\sin^2 x \geq 0$, by the fourth inequality in Lemma 3.4. This concludes the proof of (15), and completes the proof of Lemma 3.3. □

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] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut [extended abstract]. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 553–562, New York, 2005. ACM.

[3] 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.

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

[5] 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, 1998.

[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. 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.

[8] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. *J. ACM*, 52(5):766–788 (electronic), 2005.

[9] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *20th Annual IEEE Conference on Computational Complexity*, pages 144–153, 2005.

[10] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.

[11] J. Cheeger and B. Kleiner. Differentiating maps into L^1 and the geometry of BV functions. Preprint, 2006.

[12] J. Cheeger and B. Kleiner. Embedding Laakso spaces into L^1. Preprint, 2006.

[13] J. Cheeger and B. Kleiner. On the differentiation of Lipschitz maps from metric measure spaces to Banach spaces. Preprint, 2006.

[14] J. Cheeger, B. Kleiner, and A. Naor. Quantitative bounds on the rate of central collapse of Lipschitz maps on the Heisenberg group. In preparation, 2006.

[15] J. Cygan. Subadditivity of homogeneous norms on certain nilpotent Lie groups. *Proc. Amer. Math. Soc.*, 83(1):69–70, 1981.

[16] N. Devanur, S. Khot, R. Saket, and N. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 537–546, 2006.

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

[18] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.

[19] M. X. Goemans. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79:143–161, 1997.

[20] M. Gromov. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian geometry*, volume 144 of *Progr. Math.*, pages 79–323. Birkhäuser, Basel, 1996.

[21] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[22] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and l_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

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

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

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

[26] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *46th Annual Symposium on Foundations of Computer Science*, pages 53–62. IEEE Computer Soc., Los Alamitos, CA, 2005.

[27] P. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In *25th Annual ACM Symposium on Theory of Computing*, pages 682–690, May 1993.

[28] 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.

[29] R. Krauthgamer and Y. Rabani. Improved lower bounds for embeddings into L_1. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1010–1017, New York, NY, USA, 2006. ACM Press.

[30] J. R. Lee. Improved distortion estimates for negative type metrics. Preprint, 2006.

[31] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: dimension, snowflakes, and average distortion. *European J. Combin.*, 26(8):1180–1190, 2005.

[32] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

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

[34] 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.

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

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

[37] N. Linial. Planar graph metrics into ℓ_1. In J. Matoušek, editor, *Open Problems, Workshop on Discrete Metric Spaces and their Algorithmic Applications*. Haifa, March 2002.

[38] N. Linial. Squared ℓ_2 metrics into ℓ_1. In J. Matoušek, editor, *Open Problems, Workshop on Discrete Metric Spaces and their Algorithmic Applications*. Haifa, March 2002.

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

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

[41] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

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

[43] D. Preiss and L. Zajíček. Sigma-porous sets in products of metric spaces and sigma-directionally porous sets in Banach spaces. *Real Anal. Exchange*, 24(1):295–313, 1998/99.

[44] Y. Rabinovich. On average distortion of embedding metrics into the line and into l_1. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 456–462 (electronic), New York, 2003. ACM.

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

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

[47] 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.

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

[49] K. Talwar. Decomposability of metrics. In J. Matoušek, editor, *Open Problems, Workshop on Discrete Metric Spaces and their Algorithmic Applications*. Haifa, March 2002.

[50] 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.