Acta Math., 207 (2011), 291–373
DOI: 10.1007/s11511-012-0071-9

Compression bounds for Lipschitz maps from the Heisenberg group to L_1

by

Jeff Cheeger
Courant Institute
New York, NY, U.S.A.

Bruce Kleiner
Courant Institute
New York, NY, U.S.A.

Assaf Naor
Courant Institute
New York, NY, U.S.A.

Contents

1. Introduction and statement of main results

Theorem 1.1, the main result of this paper, is a quantitative bi-Lipschitz non-embedding theorem, in which the domain is a metric ball in the Heisenberg group, $\mathbb{H}$, with its Carnot–Carthéodory metric, $d^{\mathbb{H}}$, and the target is the space L_1; for the definition of $d^{\mathbb{H}}$

J.C. was supported in part by NSF grant DMS-0704404. B.K. was supported in part by NSF grant DMS-0805939. A.N. was supported in part by NSF grants CCF-0635078 and CCF-0832795, BSF grant 2006009 and the Packard Foundation.

see §2.4. This result has consequences of a purely mathematical nature, as well as for theoretical computer science.

Define $c_p(X, d^X)$, the L_p *distortion* of the metric space (X, d^X), to be the infimum of those $D>0$ for which there exists a mapping $f: X \to L_p$ satisfying

$$\frac{\|f(x)-f(y)\|_{L_p}}{d^X(x,y)} \in [1, D]$$

for all distinct $x, y \in X$. The quantitative study of bi-Lipschitz embeddings of finite metric spaces in L_p spaces goes back to [28] and [50]. The modern period begins with a result of Bourgain [12] who answered a question of [39] by showing that for every fixed p, any n-point metric space can be embedded in L_p with distortion $\lesssim \log n$.(1) By [53], Bourgain's theorem is sharp for any fixed $p<\infty$.

Since L_1, equipped with the square root of its usual distance, is well known to be isometric to a subset of L_2 (see for example [63]) it follows that if (X, d^X) isometrically embeds in L_1, then $\big(X, \sqrt{d^X}\big)$ isometrically embeds in L_2. Metrics for which $\big(X, \sqrt{d^X}\big)$ isometrically embeds in L_2 are said to be *of negative type*. Such metrics will play a fundamental role in our discussion. On the other hand, it is also well known that L_2 embeds isometrically in L_1 (see for example [64]) which implies that

$$c_1(X, d^X) \leqslant c_2(X, d^X)$$

for all (X, d^X). Recently, it was shown that for n-point metric spaces of negative type, $c_2(X, d^X) \lesssim (\log n)^{1/2+o(1)}$, and in particular $c_1(X, d^X) \lesssim (\log n)^{1/2+o(1)}$; see [5], which improves on a corresponding result in [15]. For embeddings in L_2, the result of [5] is sharp up to the term $o(1)$; see [28].

As shown in [48], the Carnot–Carathéodory metric $d^{\mathbb{H}}$ *is* bi-Lipschitz equivalent to a metric of negative type. From this and Corollary 1.2 of Theorem 1.1 below, it follows immediately that for all n, there exist n-point metric spaces of negative type with

$$c_1(X, d^X) \gtrsim (\log n)^{\delta},$$

for some explicit $\delta>0$. From the standpoint of such *non-embedding theorems*, the target L_1 presents certain challenges. Lipschitz functions $f: \mathbb{R} \to L_1$ need not be differentiable anywhere. Therefore, a tool which is useful for L_p targets with $1<p<\infty$ is not available. Moreover, the fact that L_2 embeds isometrically in L_1 implies that bi-Lipschitz embedding in L_1 is no harder than in L_2 and might be strictly easier in cases of interest.

(1) In this paper, the symbols $\lesssim$ and $\gtrsim$ denote the corresponding inequalities, up to a universal multiplicative constant, which in all cases can be explicitly estimated from the corresponding proof. Similarly, $\asymp$ denotes equivalence up to such a factor.

The sparsest cut problem is a fundamental NP-hard problem in theoretical computer science. This problem will be formally stated in Appendix A, where some additional details of the discussion which follows will be given; see also our paper [24] (and the references therein) which focuses on the computer science aspect of our work. A landmark development took place in the 1990s, when it was realized that this optimization problem for a certain functional defined on all subsets of an n-vertex weighted graph, is equivalent to an optimization problem for a corresponding functional over all functions from an n-vertex weighted graph to the space L_1; see [9], [10] and [53]. This is a consequence of the *cut cone* representation for metrics induced by maps to L_1, which also plays a fundamental role in this paper; see §2. Once this reformulation has been observed, one can relax the problem to an optimization problem for the corresponding functional over functions with values in *any* n-point metric space. The relaxed problem turns out to be a linear program, and hence is solvable in polynomial time. Define the *integrality gap* of this relaxation to be the supremum over all n-point weighted graphs of the ratio of the solution of the original problem to the relaxed one. The integrality gap measures the performance of the relaxation in the worst case. It is essentially immediate that the integrality gap is less than or equal to the supremum of $c_1(X, d^X)$ over all n-point metric spaces (X, d^X), and hence, by Bourgain's theorem, it is $\lesssim \log n$; see [53] and [9]. This upper bound relies only on the form of the functional and not on any special properties of L_1. A duality argument based on the cut cone representation shows that the integrality gap is actually *equal to* the supremum of $c_1(X, d^X)$ over all n-point metric spaces.

Subsequently, Goemans [31] and Linial [51] observed that if in relaxing the sparsest cut problem as above, one restricts to n-point metric spaces of negative type, one obtains a semidefinite programming problem, which, by the ellipsoid algorithm, can still be solved in polynomial time with arbitrarily good precision; see [34]. As above, by the duality argument, the integrality gap for the Goemans–Linial semidefinite relaxation is actually equal to the supremum of $c_1(X, d^X)$ over all n-point metric spaces of negative type. Based on certain known embedding results for particular metric spaces of negative type, the hope was that this integrality gap might actually be bounded (the "Goemans–Linial conjecture") or in any case, bounded by a very slowly growing function of n. At present, one knows the upper bound $\lesssim (\log n)^{1/2+o(1)}$, which follows from [5]. This result makes the Goemans–Linial semidefinite relaxation the most successful algorithm to date for solving the sparsest cut problem to within a definite factor. In the opposite direction, it was shown in [43] that the integrality gap for the Goemans–Linial semidefinite relaxation is $\gtrsim \log \log n$. The analysis of [43] improved upon that of the breakthrough result of [41], which was the first to show that the Goemans–Linial semidefinite relaxation cannot yield a constant factor approximation algorithm for the sparsest cut problem, and thus resolving the Goemans–Linial conjecture [31], [51], [52]. These lower bounds on the integrality gap depend on its characterization as the supremum of $c_1(X, d^X)$ over all n-point metric spaces of negative type.

Motivated by the potential relevance to the sparsest cut problem, the question of whether $(\mathbb{H}, d^{\mathbb{H}})$ bi-Lipschitz embeds in L_1 was raised in [48]. In response, it was shown in [20] that if $U \subset \mathbb{H}$ is open and $f: U \to L_1$ is Lipschitz (or more generally of bounded variation), then for almost all $x \in U$ (with respect to Haar measure) and y varying in the coset of the center of $\mathbb{H}$ containing x, one has

$$\lim_{y \to x} \frac{\|f(y) - f(x)\|_{L_1}}{d^{\mathbb{H}}(x, y)} = 0.$$

Thus, $(\mathbb{H}, d^{\mathbb{H}})$ does not admit a bi-Lipschitz embedding into L_1.

For further applications as discussed above, a quantitative version of this theorem of [20] was required; see Theorem 1.1 below.(2) It follows from Corollary 1.2 of Theorem 1.1 that there exists a sequence of n-point metric spaces, (X_n, d^{X_n}), of negative type, such that $c_1(X_n, d^{X_n}) \gtrsim (\log n)^\delta$, and hence that the integrality gap for the Goemans–Linial relaxation of sparsest cut is $\gtrsim (\log n)^\delta$ for some explicit $\delta > 0$; compare Remark 1.3. This represents an exponential improvement on the above mentioned lower bound $\gtrsim \log \log n$; compare also the upper bound $\leqslant (\log n)^{1/2 + o(1)}$.

We also give a purely mathematical application of Theorem 1.1 to the behavior of the L_1 compression rate of the discrete Heisenberg group. The L_1 compression rate is a well-studied invariant of the asymptotic geometry of a finitely generated group, which was defined by Gromov in [32]; see below for the definition.

In what follows, given $p \in \mathbb{H}$ and $r > 0$, we denote the $d^{\mathbb{H}}$-open ball of radius r centered at p by $B_r(p) = \{x \in \mathbb{H} : d^{\mathbb{H}}(x, p) < r\}$. On cosets of the center of H (viewed as the z-axis in $\mathbb{R}^3$) let $\mathcal{L}$ denote the 1-dimensional Lebesgue measure associated with the standard Euclidean metric on $\mathbb{R}^3$. Thus, $\mathcal{L} \times \mathcal{L}$ denotes the corresponding measure on pairs of points (x_1, x_2) which lie in the center.

Note that cosets of the center are parameterized by $\mathbb{R}^2$. Below, $\mathcal{L}_2$ denotes Lebesgue measure on $\mathbb{R}^2$. Let F denote a subset of pairs (x_1, x_2), such that x_1 and x_2 lie on some coset of the center $\underline{L}$. Define a measure on such subsets by stipulating that the measure of F is given by

$$\int_{\mathbb{R}^2} (\mathcal{L} \times \mathcal{L})(F \cap (\underline{L} \times \underline{L}))\, d\mathcal{L}_2(\underline{L}).$$

(2) For purposes of exposition, in Theorem 1.1, we restrict our attention to the case of Lipschitz maps, although everything we say has an analogous statement which apply to BV maps as well, sometimes with minor variations.

THEOREM 1.1. (Quantitative central collapse) *There is a universal constant* $\delta\in(0,1)$ *such that for every* $p\in\mathbb{H}$, *every* $f\colon B_1(p)\to L_1$ *with* $\mathrm{Lip}(f)\leqslant 1$ *and every* $\varepsilon\in\left(0,\frac{1}{4}\right)$, *there exists* $r\geqslant\frac{1}{2}\varepsilon$ *such that, with respect to Haar measure, for at least half*(3) *of the points* $x\in B_{1/2}(p)$, *at least half of the points* $(x_1,x_2)\in B_r(x)\times B_r(x)$ *which lie on the same coset of the center, with* $d^{\mathbb{H}}(x_1,x_2)\in\left[\frac{1}{2}\varepsilon r,\frac{3}{2}\varepsilon r\right]$, *we have*

$$\frac{\|f(x_1)-f(x_2)\|_{L_1}}{d^{\mathbb{H}}(x_1,x_2)}\leqslant\frac{1}{(\log(1/\varepsilon))^{\delta}}. \tag{1.1}$$

In particular, compression by a factor $\eta\in\left(0,\frac{1}{2}\right)$ is guaranteed to occur for a pair of points whose distance is $\gtrsim e^{-2\eta^{-c}}$, where $c=\delta^{-1}$ (and δ is as in (1.1)).

The constant δ in Theorem 1.1 can be explicitly estimated from our proof; a crude estimate is $\delta=2^{-60}$ in Theorem 1.1. At various points in the proof, we have sacrificed sharpness in order to simplify the exposition; this is most prominent in Proposition 7.3 below, which must be iterated several times, thus magnifying the non-sharpness. Obtaining the best possible δ in Theorem 1.1 remains an interesting open question, the solution of which probably requires additional ideas beyond those contained in this paper.

Before discussing the consequences of Theorem 1.1, we briefly indicate the reason for the form of the estimate (1.1); see also the discussion of §2. We will associate with f a *non-negative* quantity, the *total non-monotonicty*, which can be written as a *sum over scales*, and on each scale as an integral over locations. The assumption $\mathrm{Lip}(f)\leqslant 1$ turns out to imply an a-priori bound on the total non-monotonicity. Since for $\varepsilon\in(0,1)$, the total number of scales between 1 and ε is $\asymp\log(1/\varepsilon)$, by the pigeonhole principle, there exists a scale as in (1.1), such that at most locations, the total non-monotonicity is $\lesssim 1/\log(1/\varepsilon)$. We show using a stability theorem, Theorem 4.3, that for suitable δ, this gives (1.1). This discussion fits very well with a quantitative result (and phenomenon) due in a rather different context to Jones; see [40]. In particular, the proof of Theorem 1.1 actually gives a stronger result, namely, a Carleson measure estimate in the sense of Semmes; see his appendix in [33]. As explained in Appendix B, the argument indicated above can be viewed as a particular instance of a general argument which leads to a Carleson measure estimate.

1.1. The discrete case

As we have indicated, Theorem 1.1 has implications in the context of finite metric spaces. These are based on properties of the discrete version of the Heisenberg group $\mathbb{H}(\mathbb{Z})$ (defined below) and its metric balls, which follow from Theorem 1.1.

(3) In Theorem 1.1, by suitably changing the constant δ, "at least half" can be replaced by any definite fraction.

We emphasize at the outset that our results in the discrete case are obtained (without difficulty) directly from the corresponding *statements* in the continuous case, and not by a "discretization" of their *proofs*. As in [20] and in [21], where a different proof of the non-quantitative result of [20] is given, the proof of Theorem 1.1 is carried out in the continuous case, because in that case the methods of real analysis are available. Nonetheless, in the present instance, the quantitative issues remain highly non-trivial and the proof requires new ideas beyond those of [20] and [21]; see the discussion at the beginning of §2 and in particular Remark 2.1.

We will view $\mathbb{H}$ as $\mathbb{R}^3$ equipped with the non-commutative product

$$(a,b,c)\cdot(a',b',c') = (a+a', b+b', c+c'+ab'-ba');$$

for further discussion of $\mathbb{H}$, see §2. From the multiplication formula, it follows directly that for $R>0$, the map $A_R: \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$A_R(a,b,c) = (Ra, Rb, R^2c) \tag{1.2}$$

is an automorphism of $\mathbb{H}$. It is also a homothety of the metric $d^{\mathbb{H}}$. The discrete Heisenberg group $\mathbb{H}(\mathbb{Z})$ is the integer lattice $\mathbb{Z}^3$, equipped with the above product. It is a discrete co-compact subgroup of $\mathbb{H}$. For further discussion of the Heisenberg group see §2.

Fix a finite set of generators T of a finitely generated group Γ. The *word metric* d_T on Γ is the left-invariant metric defined by stipulating that $d_T(g_1,g_2)$ is the length of the shortest word in the elements of T and their inverses which expresses $g_1^{-1}g_2$. Up to bi-Lipschitz equivalence, the metric d_T is independent of the choice of generating set. For the case of $\mathbb{H}(\mathbb{Z})$, we can take $T=\{(\pm1,0,0),(0,\pm1,0),(0,0,\pm1)\}$. For definiteness, from now on this choice will be understood. By an easy general lemma, given a free co-compact action of a finitely generated group Γ acting freely and co-compactly on a length space (X,d^X), the metric on Γ induced by the restriction to any orbit of the metric d^X is bi-Lipschitz equivalent to d_T; see [14]. For the case of $\mathbb{H}(\mathbb{Z})$, we can take $(X,d^X)=(\mathbb{H},d^{\mathbb{H}})$.

Define $\varrho: \mathbb{R}^3\times\mathbb{R}^3 \to [0,\infty)$ by

$$\begin{aligned}&\varrho((x,y,z),(t,u,v))\\ &\quad := \left(\left[\left(\frac{(t-x)^2+(u-y)^2}{2}\right)^2 + (v-z+xu-yt)^2\right]^{1/2} + \frac{(t-x)^2+(u-y)^2}{2}\right)^{1/2}.\end{aligned}$$

It was shown in [48] that $(\mathbb{H},\varrho)$ is a metric of negative type, bi-Lipschitz equivalent to $(\mathbb{H},d^{\mathbb{H}})$.

It follows from [20] that

$$\lim_{n\to\infty} c_1(\{0,...,n\}^3,\varrho) = \infty, \tag{1.3}$$

but no information can be deduced on the rate of blow up as $n\to\infty$. From Theorem 1.1 we get the following corollary, whose proof will be explained at the end of this subsection.

COROLLARY 1.2. *For the constant* $\delta>0$ *in Theorem* 1.1 *we have, for all* $n\in\mathbb{N}$, *metric spaces* $(\{0,...,n\}^3,\varrho)$ *of negative type satisfying*

$$c_1(\{0,...,n\}^3,\varrho)\gtrsim(\log n)^\delta. \tag{1.4}$$

Remark 1.3. Since the metric spaces $c_1(\{0,...,n\}^3,\varrho)$ are of negative type, relation (1.4) implies that the integrality gap of the Goemans–Linial relaxation of the sparsest cut problem is $\gtrsim(\log n)^\delta$ for $\delta>0$ as Theorem 1.1.

Remark 1.4. A metric space is said to be *doubling* if a metric ball $B_{2r}(x)$ can be covered by at most $N<\infty$ metric balls of radius r, where N is independent of x and r. For n-point metric spaces (X,d) which are doubling, the bound in Bourgain's theorem can be sharpened to

$$c_1(X,d)\leqslant c_2(X,d)\lesssim\sqrt{\log n},$$

which follows from the results of [8] and [58]; see the explanation in [35]. The metric spaces $(\mathbb{H},d^{\mathbb{H}})$ and $(\mathbb{H}(\mathbb{Z}),d_T)$ are doubling. (To see this, use for example the left invariance of $d^{\mathbb{H}}$ and the homotheties A_R.) Before the bi-Lipschitz non-embeddability of $\mathbb{H}$ into L_1 was established in [20], there was no known example of a doubling metric space which does not admit a bi-Lipschitz embedding into L_1. Corollary 1.2 shows that there is a sequence of n-point doubling metric spaces for which $c_1(X,d)\gtrsim(\log n)^\delta$.

Remark 1.5. The behavior of the L_2 distortion for n-point doubling metric spaces is much easier to understand than the L_1 distortion. Namely, for fixed doubling constant, the above mentioned bound $c_2(X,d)\lesssim\sqrt{\log n}$ cannot be improved; see [46] and, for dependence on the doubling constant, [42].

Let Γ be a finitely generated group and $f\colon(\Gamma,d_T)\to L_1$ be a 1-Lipschitz function. Gromov [32] defined the *compression rate* $\omega_f\colon[1,\infty)\to[0,\infty)$ by

$$\omega_f(t):=\inf\{\|f(x)-f(y)\|_{L_1}:d_T(x,y)\geqslant t\}.$$

Stated differently, ω_f is the largest non-decreasing function for which

$$\|f(x)-f(y)\|_{L_1}\geqslant\omega_f(d_T(x,y))\quad\text{for all } x,y\in\Gamma.$$

It follows from [20] that for any 1-Lipschitz map $f\colon\mathbb{H}(\mathbb{Z})\to L_1$ we have

$$\liminf_{t\to\infty}\frac{\omega_f(t)}{t}=0,$$

but [20] does not give any information on the rate at which $\omega_f(t)/t$ tends to zero. From Theorem 1.1 we can obtain the following bound.

Corollary 1.6. *For every* $f:\mathbb{H}(\mathbb{Z})\to L_1$ *which is* 1*-Lipschitz with respect to the word metric* d_W *we have, for arbitrarily large* t,

$$\omega_f(t)\lesssim\frac{t}{(1+\log t)^\delta}, \tag{1.5}$$

where $\delta>0$ *is the constant in Theorem* 1.1.

Remark 1.7. A general result from [62] (see Corollary 5 there) implies that if an increasing function $\omega:[1,\infty)\to[0,\infty)$ satisfies

$$\int_1^\infty \frac{\omega(t)^2}{t^3}\,dt<\infty,$$

then there exists a mapping $f:\mathbb{H}(\mathbb{Z})\to L_1$ which is Lipschitz in the metric d_T and such that $\omega_f\gtrsim\omega$. In fact, f can be chosen to take values in the smaller space L_2. By choosing

$$\omega(t)=\frac{t}{\sqrt{1+\log t}\,\log\log(2+t)},$$

and bringing in Corollary 1.6, it follows that in the terminology of [7] the discrete Heisenberg group $(\mathbb{H}(\mathbb{Z}),d_T)$ has L_1 *compression gap*

$$\left(\frac{t}{\sqrt{1+\log t}\,\log\log(2+t)},\frac{t}{(1+\log t)^\delta}\right).$$

It would be of interest to evaluate the supremum of those $\delta>0$ for which every Lipschitz function $f:\mathbb{H}(\mathbb{Z})\to L_1$ satisfies (1.5).

We close this subsection by explaining how Corollaries 1.6 and 1.2 are deduced from Theorem 1.1. A key point is to pass from the discrete settings of these corollaries to the continuous setting of Theorem 1.1 via a Lipschitz extension theorem. The basic idea is simple and general (see [48, Remark 1.6]). Additionally, the homotheties A_R defined in (1.2) are used to convert information from Theorem 1.1 concerning small scales, into information concerning large scales. The existence of these homotheties is, of course, a special property of $(\mathbb{H},d^{\mathbb{H}})$.

We will give the details for Corollary 1.6; the case of Corollary 1.2 is entirely similar.

Proof of Corollary 1.6. Fix a map $f:\mathbb{H}(\mathbb{Z})\to L_1$, which is 1-Lipschitz with respect to the word metric d_T. The map f can be extended to a map $\tilde f:\mathbb{H}\to L_1$ whose Lipschitz constant with respect to $d^{\mathbb{H}}$ satisfies $\mathrm{Lip}(\tilde f)\lesssim 1$. This fact follows from the general result of [47] which states that such an extension is possible for any Banach-space-valued mapping from any doubling subset of a metric space to the entire metric space, but in the present simpler setting it also follows from a straightforward partition-of-unity argument.

Fix $R>1$ and define $g_R: B_1(0,0,0)\to L_1$ by $g_R(x)=\tilde{f}(\delta_R(x))/R$. Since

$$d^{\mathbb{H}}(A_R(x),A_R(y))=Rd^{\mathbb{H}}(x,y),$$

we have $\mathrm{Lip}(g_R)\lesssim 1$. For $\varepsilon\in\left(0,\frac{1}{4}\right)$, an application of Theorem 1.1 shows that there exist $x,y\in\mathbb{H}$ such that $R\gtrsim d^{\mathbb{H}}(x,y)\gtrsim \varepsilon R$, and

$$\|\tilde{f}(x)-\tilde{f}(y)\|_{L_1}\lesssim \frac{d^{\mathbb{H}}(x,y)}{(\log(1/\varepsilon))^{\delta}}. \tag{1.6}$$

Choose $\varepsilon=1/\sqrt{R}$. Since there exist $a,b\in\mathbb{Z}^3$ such that $d^{\mathbb{H}}(a,x)\lesssim 1$ and $d^{\mathbb{H}}(b,y)\lesssim 1$, and since $\mathrm{Lip}(\tilde{f})\lesssim 1$, it follows from (1.6) that provided R is large enough,

$$\omega_f(d_W(a,b))\leqslant \|f(a)-f(b)\|_{L_1}\lesssim \frac{d^{\mathbb{H}}(a,b)}{(\log R)^{\delta}}\lesssim \frac{d^{\mathbb{H}}(a,b)}{(\log d^{\mathbb{H}}(a,b))^{\delta}},$$

implying (1.5). □

2. Proof of Theorem 1.1: overview and background

We begin with an informal overview of the proof of Theorem 1.1. We then proceed to a more detailed discussion, including relevant background material.

Many known bi-Lipschitz non-embedding results are based in essence on a differentiation argument. Roughly, one shows that at almost all points, in the infinitesimal limit, a Lipschitz map converges to a map with a special structure. One then shows (which typically is not difficult) that maps with this special structure cannot be bi-Lipschitz. The term "special structure" means different things in different settings. When the domain and range are Carnot groups as in Pansu's differentiation theorem [57], "special structure" means a group homomorphism. As observed in [18] and [48], Pansu's theorem extends to the case where the domain is $(\mathbb{H},d^{\mathbb{H}})$ and the target is an infinite-dimensional Banach space with the Radon–Nikodym property; in particular, the target can be L_p, $1<p<\infty$.

The above approach fails for embeddings into L_1, since even when the domain is $\mathbb{R}$, Lipschitz maps need not be differentiable anywhere. A simple example is provided by the map $t\mapsto\chi_{[0,t]}$, where $\chi_{[0,t]}$ denotes the characteristic function of $[0,t]$; see [4]. Nevertheless, the result on central collapse proved in [20] can be viewed as following from a differentiation theorem, provided one interprets this statement via a novel notion of "infinitesimal regularity" of mappings introduced in [20].

The approach of [20] starts out with the cut-cone representation of L_1 metrics (see [20] and [27]), which asserts that for every $f:\mathbb{H}\to L_1$ we can write

$$\|f(x)-f(y)\|_{L_1} = \int_{2^{\mathbb{H}}} |\chi_E(x)-\chi_E(y)|\, d\Sigma_f(E)$$

for all x and y, where Σ_f is a canonically defined measure on $2^{\mathbb{H}}$ (see the discussion following (2.5) for precise formulations). The differentiation result of [20] can be viewed as a description of the infinitesimal structure of the measure Σ_f. It asserts that at most locations, in the infinitesimal limit, Σ_f is supported on vertical half spaces. This is achieved by first showing that the Lipschitz condition on f implies that Σ_f is supported on special subsets, those with finite perimeter. One then uses results on the local structure of sets of finite perimeter in the Heisenberg group to complete the proof; the argument is described in greater detail later in this section.

An alternative approach to this result on the infinitesimal behavior of Σ_f, which does not require the introduction of sets of finite perimeter, was obtained in [21]. This approach is based on the classification of monotone sets. A subset $E\subset\mathbb{H}$ is called *monotone* if for every *horizontal* line L, up to a set of measure zero, $E\cap L$ is a subray of the line. If we view $\mathbb{H}$ as $\mathbb{R}^3$, the horizontal lines are a certain codimension-1 subset of all the lines; for details, see §2.4. The proof of [21] proceeds by showing that infinitesimally the measure Σ_f is supported on monotone subsets, and a non-trivial classification theorem which asserts that monotone subsets are half spaces. (The proof is recalled in §8.)

Our proof of Theorem 1.1 combines the methods of [20] and [21], with several significant new ingredients. Inspired by the proof in [21], our argument is based on an appropriately defined notion of almost monotonicity of subsets. But, unlike [21], we require the use of perimeter bounds as well. In our situation, perimeter bounds are used in finding a controlled scale such that at most locations, apart from a certain collection of cuts, the mass of Σ_f is supported on subsets which are sufficiently close to being monotone. In actuality, the excluded cuts may have infinite measure with respect to Σ_f. Nonetheless, using perimeter bounds and the isoperimetric inequality in $\mathbb{H}$, we show that their contribution to the metric is negligibly small.

A crucial, and by far the most complicated, new ingredient in this paper is a stability version of the classification of monotone sets of [21], which asserts that sets which are almost monotone are quantitatively close to half spaces. One of the inherent reasons for the difficulty of proving such a stability result arises from the need to work locally, i.e., to consider almost monotone subsets inside a metric ball in $\mathbb{H}$ of finite radius. Here the situation is fundamentally different from the corresponding classification result in [21]: there are even precisely monotone subsets in such a ball which are not half spaces(4);

(4) In $\mathbb{R}^n$, monotone subsets of a ball are necessarily the intersection of the ball with a half space.

see Example 9.1. Inevitably, our classification result must take this complication into account. We can only assert that on a (controllably) smaller ball, the given almost-monotone subset is close to a half space.

In order to make the above informal description precise, we will require some additional preliminaries, which are explained in the remainder of this introduction. For a discussion, in a wider context, of estimates whose form is equivalent to that of (1.1), and which arise from our general mechanism, see Appendix B.

Remark 2.1. As we explained above, the present paper belongs to the general topic of controlling the scale at which the infinitesimal "special structure" of a mapping (arising from a differentiation theorem) appears in approximate form. However, it is important to realize that *it is impossible in general to obtain a quantitative estimate for the rate at which the limiting situation is approached.* To see this point, consider a model problem associated with the sequence of functions

$$f_n(x) = \frac{\sin nx}{n}.$$

Although $|f_n'(x)| \leqslant 1$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$, it is not possible to control, independently of n, the difference between f_n and its first-order Taylor polynomial; note that f_n'', which controls the remainder term, is not bounded independently of n. Nonetheless, there is an explicit uniform estimate for the scale above which any f_n is approximated by *some* linear function, which might not be equal to its first-order Taylor series. As far as we are aware, the first instance of a result of this type is in [40]; see also for example [11], and for additional information, the appendix by Semmes in [33].

2.1. PI spaces

A natural setting for a large part of our work is that of PI spaces: a certain class of metric measure spaces, of which the Heisenberg group is a member. Together with the key concept of "upper gradient" on which their definition is based, these spaces were introduced in [38].

Let (X, d^X) denote a metric space. If $f: X \to \mathbb{R}$, then a Borel measurable function $g: X \to [0, \infty]$ is called an *upper gradient* of f if for all rectifiable curves $c: [0, \ell] \to X$, parameterized by arc length,

$$|f(c(\ell)) - f(c(0))| \leqslant \int_0^\ell g(c(s))\, ds. \tag{2.1}$$

A complete metric measure space (X, d^x, μ) is called a *PI space* if a *doubling condition* and *Poincaré inequality* hold. That is, there exists $p \geqslant 1$ and for all $R > 0$ there exist

$\beta = \beta(R)$, $\tau = \tau(R)$ and $\Lambda = \Lambda(R)$ such that for all $x \in X$ and $r \leqslant R$ we have

$$\mu(B_{2r}(x)) \leqslant \beta \mu(B_r(x)), \tag{2.2}$$

$$⨍_{B_r(x)\times B_r(x)} |f(x_1)-f(x_2)|\, d(\mu\times\mu)(x_1,x_2) \leqslant \tau r \left(⨍_{B_{\Lambda r}(x)} g^p\, d\mu \right)^{1/p} \tag{2.3}$$

for all measurable f and upper gradients g of f, where we use the notation

$$⨍_U h\, d\mu := \frac{1}{\mu(U)} \int_U h\, d\mu.$$

As shown in [36], (2.2) and (2.3) imply a strengthening of (2.3), which we will need. Namely, for some $\chi = \chi(\beta,\tau) > 1$, $\tau' = \tau'(\beta,\tau)$, one has the *Poincaré–Sobolev inequality*

$$\left(⨍_{B_r(x)\times B_r(x)} |f(x_1)-f(x_2)|^{\chi p}\, d(\mu\times\mu)(x_1,x_2) \right)^{1/\chi p} \leqslant \tau' r \left(⨍_{B_{\Lambda r}(x)} g^p\, d\mu \right)^{1/p} \tag{2.4}$$

for all measurable f and upper gradients g of f.

It is important that for the case of the Heisenberg group of dimension $2n+1$, relation (2.4) holds with $p=1$ and $\chi = (2n+2)/(2n+1)$; in particular, if $2n+1=3$, then $\chi = \frac{4}{3}$.

Observe that (2.3), for fixed p, implies (2.3) for any $p' > p$. In this paper, without further explicit mention, we will always assume that $p=1$, which is necessary for the ensuing results on finite perimeter. Solely in order to have one fewer constant to list, we will also make the innocuous assumption that $\Lambda = 1$. For the cases of primary interest here, $X = \mathbb{R}^n$ or $X = \mathbb{H}$, relation (2.4) does hold with $\Lambda = 1$ and, as noted above, with $p=1$.

2.2. Maps to L_1: cut metrics and cut measures

Let X denote a set and let $E \subset X$ denote a subset. In place of "subset", we will also use the term "cut". Associated with E there is the so-called *elementary cut metric* on X, defined by

$$d_E(x_1,x_2) = \begin{cases} 0, & \text{if either } x_1,x_2 \in E \text{ or } x_1,x_2 \in E^c, \\ 1, & \text{otherwise} \end{cases}$$

(here, and in what follows, we set $E^c = X \setminus E$).

Remark 2.2. Clearly, $d_E = d_{E^c}$. For this reason, it is common to define the space of cuts of X as the quotient of the power set 2^X by the involution $E \mapsto E^c$, although we do not do this here.

By definition, a *cut metric* d_Σ is an integral of elementary cut metrics with respect to some measure Σ on 2^X. Thus,

$$d_\Sigma(x_1,x_2)=\int_{2^X} d_E(x_1,x_2)\,d\Sigma(E). \tag{2.5}$$

For $f\colon U\to L_1$, the pull-back metric induced by f is defined as

$$d_f(x_1,x_2)=\|f(x_1)-f(x_2)\|_1.$$

It follows from the "cut-cone characterization" of L_1 metrics (see [20] and [27]) that there is a canonically defined measure Σ_f on 2^X (on an associated σ-algebra) such that (2.5) holds with $\Sigma=\Sigma_f$.

Now let (X,μ) denote a σ-finite measure space. Let $U\subset X$ denote a measurable subset and $f\colon U\to L_1(Y,\nu)$ be a map which satisfies

$$\int_U \|f(x)\|_1\,d\mu<\infty.$$

There is a variant of the description of d_f in terms of Σ_f in the L_1 framework; see [20]. In this context, 2^U is replaced by a measure-theoretic version of the space of cuts. One should regard a *cut* as an equivalence class of measurable sets $E\subset U$ of finite μ-measure, where two sets are considered equivalent if their symmetric difference has measure zero. The collection of such cuts may be identified with the subset of $L_1(U)$ consisting of characteristic functions. Hence it inherits a topology and a Borel structure from $L_1(U)$. For our purposes, there is no harm in blurring the distinction between measurable sets and their equivalence classes, which, for purposes of exposition, is done below.

Let $\mathrm{Cut}(U)$ denote the space of cuts. As above, we have

$$d_f(x_1,x_2)=\int_{\mathrm{Cut}(U)} d_E(x_1,x_2)\,d\Sigma_f(E), \tag{2.6}$$

where Σ_f is a suitable σ-finite Borel measure on $\mathrm{Cut}(U)$. Here we view d_f as an element of $L_1^{\mathrm{loc}}(U\times U)$, whose restriction to any subset $V\subset U$, with finite measure, lies in $L_1(V\times V)$. Associated with two such L_1-metrics $d_1,d_2\in L_1(V\times V)$, there is a well-defined L_1-distance, given by

$$\|d_1-d_2\|_{L_1(V\times V)}=\int_{V\times V} |d_1(x_1,x_2)-d_2(x_1,x_2)|\,d(\mu\times\mu)(x_1,x_2). \tag{2.7}$$

It is shown in [20] that if X is a PI space, $U\subset B_1(p)\subset X$ is an open set and f is 1-Lipschitz, or more generally 1-BV, then the cut measure Σ_f has finite total perimeter:

$$\int_{\mathrm{Cut}(U)} \mathrm{Per}(E)(U)\,d\Sigma_f(E)=\int_{\mathrm{Cut}(U)} \mathrm{PER}(E,U)\,d\Sigma_f(E)<c(\beta,\tau). \tag{2.8}$$

Intuitively, for U open, the *perimeter* $\mathrm{PER}(E,U)$ is the codimension-1 measure of the (measure-theoretic) boundary of E inside U. In actuality, $U\mapsto\mathrm{PER}(E,U)$ defines a Borel measure $\mathrm{Per}(E)$. The restriction of this measure to Borel subsets of U is denoted $\mathrm{Per}(E,U)$. The mass of $\mathrm{Per}(E,U)$ is denoted $\mathrm{PER}(E,U)$ and is equal to $\mathrm{Per}(E)(U)$; for the definition, see (3.5). Moreover, there is a *total perimeter measure* λ_f, which is a Radon measure on U, such that

$$\lambda_f=\int_{\mathrm{Cut}(U)}\mathrm{Per}(E,U)\,d\Sigma_f(E), \tag{2.9}$$

$$\mathrm{Mass}(\lambda_f)=\lambda_f(U)=\int_{\mathrm{Cut}(U)}\mathrm{PER}(E,U)\,d\Sigma_f(E). \tag{2.10}$$

In particular, if f has *bounded variation* (in short, f is BV), then Σ_f is supported on cuts with *finite perimeter*. For the precise definitions and relevant properties of cut measures, BV maps and perimeter measures, see [20].

2.3. The Heisenberg group as a Lie group

Recall that the 3-dimensional Heisenberg group $\mathbb{H}$ can be viewed as $\mathbb{R}^3$ equipped with the group structure

$$(a,b,c)\cdot(a',b',c')=(a+a',b+b',c+c'+ab'-ba'). \tag{2.11}$$

Note that the inverse of (a,b,c) is $(-a,-b,-c)$. The center of $\mathbb{H}$ consists of the 1-dimensional subgroup $\{0\}\times\{0\}\times\mathbb{R}$. There is a natural projection

$$\pi\colon\mathbb{H}\longrightarrow\frac{\mathbb{H}}{\mathrm{Center}(\mathbb{H})}=\mathbb{R}^2$$

and the *cosets of the center* are the *vertical lines* in $\mathbb{R}^3$, i.e., lines parallel to the z-axis.

Since the correction term $ab'-a'b$ in (2.11), which measures the failure of the multiplication to be commutative, can be viewed as a determinant, we get the following very useful geometric interpretation:

$(*)$ *The correction term* $ab'-a'b$ *is the signed area of the parallelogram spanned by the vectors* $\pi(a,b,c)$ *and* $\pi(a',b',c')$.

$(**)$ *Equivalently, if we regard* $\mathbb{R}^3$ *as* $\mathbb{R}^2\times\mathbb{R}$, *then* $ab'-a'b$ *is the standard symplectic form* ω *on* $\mathbb{R}^2$.

Let $K\colon\mathbb{R}^2\to\mathbb{R}^2$ be an invertible linear transformation and set

$$A_K(a,b,c):=(K(a,b),c\det K). \tag{2.12}$$

It is easily checked that A_K is an *automorphism* of $\mathbb{H}$.

Let H_g denote the affine plane passing through $g\in\mathbb{H}$, which is the image under left multiplication by $g\in\mathbb{H}$, of the subspace $\mathbb{R}^2\times\{0\}\subset\mathbb{H}$. We call H_g the *horizontal* 2-*plane* at $g\in\mathbb{H}$. Since the automorphism A_K preserves the subspace $\mathbb{R}^2\times\{0\}\subset\mathbb{H}_e$, where $\mathbb{H}_e$ denotes the tangent space at the identity $e=(0,0,0)$, it follows that A_K maps horizontal subspaces to horizontal subspaces. The collection $\{H_g\}_{g\in\mathbb{H}}$ defines a left-invariant connection on the principle bundle $\mathbb{R}\to\mathbb{H}\to\mathbb{R}^2$, which in coordinates has the following explicit description.

The plane $H_{(0,0,0)}$ is given by $(u,v,0)$, where u and v take arbitrary real values. In general,

$$H_{(a,b,c)}=(a,b,c)\cdot(u',v',0)=(a+u',b+v',c-bu'+av'),$$

so putting $a+u'=u$ and $b+v'=v$, we get

$$H_{(a,b,c)}=(u,v,c+av-bu). \tag{2.13}$$

The affine 2-planes in $\mathbb{R}^3$ whose projections to $\mathbb{R}^2$ are surjective are just those which admit a parameterization $(u,v,c+av-bu)$. It follows that *every* such 2-plane arises as the horizontal 2-plane associated with a unique point (a,b,c), and conversely that every horizontal 2-plane associated with some point in $\mathbb{R}^3$ projects surjectively onto $\mathbb{R}^2$.

A line in $\mathbb{R}^3$ which passes through the point (a,b,c) and lies in the plane $H_{(a,b,c)}$ can be written as

$$L=(a,b,c)+t(u,v,-bu+av),\quad t\in\mathbb{R}, \tag{2.14}$$

where u and v are fixed and t varies. In discussing the Heisenberg case, unless otherwise indicated, the term *line* will refer exclusively to such a *horizontal* line. The collection of all such lines is denoted by lines($\mathbb{H}$).

Definition 2.3. A *half-space* $\mathcal{P}\subset\mathbb{H}=\mathbb{R}^3$ is the set of points lying on one side of some 2-plane P, including those points of the plane itself.

The half-space $\mathcal{H}$ is called *horizontal* if its associated 2-plane H is horizontal. Otherwise it is called *vertical.* Thus, a vertical half-space is the inverse image under π of an ordinary half-plane in $\mathbb{R}^2$.

2.4. The Heisenberg group as a PI space

Consider the left-invariant Riemannian metric $\langle\cdot,\cdot\rangle$ on $\mathbb{H}$ which corresponds to the standard Euclidean metric at the tangent space to the identity. The *Carnot–Carathéodory distance* $d^{\mathbb{H}}(x_1,x_2)$ is defined to be the infimum of lengths of curves $c\colon[0,\ell]\to\mathbb{H}$, from

x_1 to x_2, such that for all s the tangent vector $c'(s)$ is *horizontal*, i.e., $c'(s)$ lies in the 2-dimensional subspace of the tangent space corresponding to the affine plane $H_{c(s)}$. The length of $c'(s)$ is calculated with respect to $\langle \cdot, \cdot \rangle$. In particular, for all $g_1, g_2 \in \mathbb{H}$ one has

$$d^{\mathbb{H}}(g_1, g_2) = d^{\mathbb{H}}(g_1^{-1}, g_2^{-1}). \tag{2.15}$$

A well-known consequence (see [56]) of the definition of $d^{\mathbb{H}}$, in combination with $(*)$ and $(**)$ above, is that if $p, q \in \mathbb{H}$, with $\pi(p) = \pi(q)$, and $c: [0, \ell] \to \mathbb{R}^2$ is a curve parameterized by arc length such that its horizontal lift $\tilde{c}$ starts at p and ends at q, then the vertical separation of p and q (in coordinates) is the signed area enclosed by the curve c.

Geodesics of d^H can be characterized as those smooth horizontal curves which project (locally) either to a circular arc or to a line segment in $\mathbb{R}^2$. Thus, the lines $L \in \mathrm{lines}(\mathbb{H})$ are precisely those geodesics of $d^{\mathbb{H}}$ which are affine lines in $\mathbb{H} = \mathbb{R}^3$. Any two points of $\mathbb{H}$ can be joined by a minimal geodesic of $d^{\mathbb{H}}$, although typically *not* by a horizonal line $L \in \mathrm{lines}(\mathbb{H})$. Note that, when viewed as a curve in $(\mathbb{H}, d^{\mathbb{H}})$, any affine line which is not in $\mathrm{lines}(\mathbb{H})$ has the property that any of its finite subsegments has infinite length.

The distance $d^{\mathbb{H}}((a,b,c),(a',b',c'))$, is bounded above and below by a constant multiple of

$$\sqrt{(a-a')^2 + (b-b')^2} + |c-c'+ab'-ba'|^2. \tag{2.16}$$

When restricted to any vertical line, it is just the *square root of the coordinate distance*. Thus, there is a constant $C > 0$ such that, for the metric $d^{\mathbb{H}}$, the metric ball centered at $e = (0,0,0)$ satisfies the *box-ball principle*:

$$B_{C^{-1}r}(e) \subset \{(a,b,c) : |a| < r, |b| < r \text{ and } |c| < \sqrt{r}\} \subset B_{Cr}(e). \tag{2.17}$$

Thus, in coordinates, $B_r(e)$ looks roughly like a cylinder whose base has radius r and whose height is $2r^2$.

In this paper we will consider the PI space $(\mathbb{H}, d^{\mathbb{H}}, \mathcal{L}_3)$, where $\mathcal{L}_3$ denotes the Haar measure on $\mathbb{H}$, which coincides with the Lebesgue measure on $\mathbb{R}^3$.

Let r_θ denote the rotation in $\mathbb{R}^2$ by an angle θ and let I denote the identity on $\mathbb{R}^2$. From now on, we write O_θ for A_{r_θ}. Since O_θ preserves horizontal subspaces and induces an isometry on $H_e = \mathbb{R}^2 \times \{0\}$, it follows that O_θ is an isometry of $(\mathbb{H}, d^{\mathbb{H}})$. Clearly, O_θ preserves the measure $\mathcal{L}_3$ as well. Similarly, for $\psi \in \mathbb{R}$, the automorphism $A_{\psi I}$ scales the metric by a factor ψ and the measure $\mathcal{L}_3$ by a factor ψ^4; see (2.12).

2.5. Behavior under blow up of finite-perimeter cuts of $\mathbb{H}$

Let $E \subset B_1(e) \subset \mathbb{H}$ be a finite-perimeter (FP) cut. Then at $\mathrm{Per}(E)$-a.e. $p \in E$, asymptotically under blow up, the measure of the symmetric difference of E and some vertical half-space $\mathcal{V}$, goes to zero; see [29] and [30]. Equivalently, the indicator χ_E converges to $\chi_{\mathcal{V}}$ in the L_1^{loc} sense. The corresponding theorem for $\mathbb{R}^n$ is due to De Giorgi [25], [26]. The results of [29] and [30] depend essentially on those of [1] and [2], in which an asymptotic doubling property for the perimeter measure is proved for arbitrary PI spaces.

If $d_{\mathcal{V}}$ is an elementary cut metric associated with a vertical half-space then the restriction of $d_{\mathcal{V}}$ to a coset of the center is trivial. Thus, the results of [29] and [30], together with (2.5), suggest that under blow up, at almost all points, a Lipschitz map $f:\mathbb{H}\to L_1$ becomes degenerate in the direction of (cosets of) the center. In particular, there exists no bi-Lipschitz embedding of $\mathbb{H}$ in L_1. This is the heuristic argument behind the main result of [20].

In order to prove Theorem 1.1, we will take a different approach, leading to a *quantitative version* of a somewhat crude form of the blow-up results of [29] and [30] (and the corresponding earlier results of [25] and [26]). Here, "crude" means that our argument does not give *uniqueness* of the blow up, nor in the Heisenberg case, does it show that only *vertical* half-spaces arise [30]. For our purposes, neither of these properties is needed. Our approach is based on the notion of *monotone sets* as introduced in [21]; the discussion there, while not quantitative, does recover the results on verticality and uniqueness of blow ups.

2.6. Monotone sets and half-spaces

In the following definition and elsewhere in the paper, E^c denotes the complement of E.

Definition 2.4. Fix an open set $U\subseteq\mathbb{H}$. We denote by lines(U) the space of unparameterized oriented horizontal lines whose intersection with U is non-empty. Let $\mathcal{N}_U$ denote the unique left-invariant measure on lines$(\mathbb{H})$, normalized so that $\mathcal{N}_U(\mathrm{lines}(U))=1$. A subset $E\subseteq U$ is *monotone with respect to* U if for $\mathcal{N}_U$-a.e. line L, both $E\cap L$ and $(U\setminus E)\cap L$ are *essentially connected*, in the sense that there exist connected subsets $F_L=F_L(E), F_L^c=F_L^c(E)\subseteq L$ (i.e., each of F_L and F_L^c is either empty, equals L, or is an interval, or a ray in L) such that the symmetric differences $(E\cap L)\triangle F_L$ and $((U\setminus E)\cap L)\triangle F_L^c$ have 1-dimensional Hausdorff measure zero.

Monotone subsets of $\mathbb{H}$ were introduced in [21], where they were used to give a relatively short proof of the non-embedding theorem of [20], which does not require the introduction of FP sets, and hence does not depend on [29] and [30]. Instead, a blow-up argument is used to directly reduce the non-embedding theorem to the special case in which the cut measure Σ_f is supported on monotone cuts; compare the discussion in the next subsection. For the case $U=\mathbb{H}$, a non-trivial classification result asserts that if E is monotone, then $\mathcal{L}_3(E\triangle\mathcal{P})=0$ for some half-space $\mathcal{P}$; see [21].

Remark 2.5. If we regard $\mathbb{H}$ as $\mathbb{R}^3$, then horizontal lines in $\mathbb{H}$ are a particular codimension-1 subset of the set of all affine lines in $\mathbb{R}^3$. A typical pair of points lies on no horizontal line. However, the classification of monotone subsets of $\mathbb{H}$ is precisely the same as for $\mathbb{R}^3$ with its standard metric. In the latter case the proof is trivial, while in the former case it is not.

2.7. Degeneracy of cut metrics which are supported on half-spaces

Once monotone subsets are known to be half-spaces it follows (a posteriori) that the connectedness condition in Definition 2.4 holds for *almost every* affine line $\underline{L}$, i.e., not just for horizontal ones. Thus, if a cut measure Σ is supported on monotone cuts, d_Σ has the property that if $x_1, x_2, x_3 \in \underline{L}$ and x_2 lies between x_1 and x_3, then

$$d_\Sigma(x_1, x_3) = d_\Sigma(x_1, x_2) + d_\Sigma(x_2, x_3).$$

But if $\underline{L}$ is *not horizontal*, i.e., not a coset of the center, then $d^{\mathbb{H}}|_{\underline{L}}$ is comparable to the square root of the coordinate distance, and it is trivial to verify that this metric is not bi-Lipschitz equivalent to one with the property mentioned above; see (4.1).

In proving Theorem 1.1, we will show that on a definite scale, using Theorem 4.3, we can reduce modulo a controlled error, to the case in which the cut measure is supported on cuts which are close to half-spaces. The error term, though controlled, is larger than the term which corresponds to the model case of monotone cuts.

2.8. δ-monotone sets

Theorem 4.3, which asserts that an approximately monotone set is close to some half-space (in the sense that the symmetric difference has small measure) plays a key role in the proof of Theorem 1.1. But unlike in [21], we cannot dispense with consideration of FP sets. We use (2.8), the bound on the total perimeter, to obtain a bound on the total non-monotonicity, and hence to get the estimate for a scale on which the total non-monotonicity is so small that Theorem 4.3 applies; see the discussion after Theorem 1.1 and compare with Remark 2.6 below.

Remark 2.6. If one restricts the attention to 1-Lipschitz maps $f: B \to L_1$, rather than more general BV maps, then using the fact that the 1-Lipschitz condition is preserved under restriction to horizontal lines, it is possible to derive the above mentioned bound on the total non-monotonicity without reference to FP sets. However, even for Lipschitz maps, the introduction of FP sets cannot be avoided; see Lemma 4.1 which concerns an issue arising from the fact that the mass of the cut measure can be infinite; see also Proposition 5.1.

We will need to consider subsets $E\subset B_R(p)\subset\mathbb{H}$, which are δ-monotone on $B_R(p)$; see Definition 4.2. Here, small δ means approximately monotone, and $\delta=0$ corresponds to the case of monotone sets as in Definition 2.4. After rescaling, Theorem 4.3 states that there exists a constant $a<\infty$, such that if $E\subset B_R(p)\subset\mathbb{H}$ is ε^a-monotone, and $R\geqslant\varepsilon^{-3}$, then $\mathcal{L}_3((E\cap B_1(p))\triangle(\mathcal{P}\cap B_1(p)))\lesssim\varepsilon$ for some half-space $\mathcal{P}$.

Remark 2.7. Absent the assumption $R\geqslant\varepsilon^{-3}$, the conclusion of Theorem 4.3 can fail, even if "ε^a-monotone" is replaced by "monotone"; see Example 9.1. For the analogous result in $\mathbb{R}^n$, ε^{-3} can indeed be replaced by 1.

Remark 2.8. The discussion of monotone sets can be formulated for arbitrary PI spaces; see [23]. But in general, monotone subsets *need not be rigid.* In [22], the flexibility of monotone subsets of Laakso spaces [45] is used to *construct* bi-Lipschitz embeddings of Laakso spaces into L_1. This is of interest since these spaces do not admit a bi-Lipschitz embedding into any Banach space with the Radon–Nikodym property, e.g., separable dual spaces, such as L_p, $1<p<\infty$, or ℓ_1; see [19].

2.9. The kinematic formula; perimeter and non-monotonicity

To bound the total non-monotonicity in terms of the total perimeter, we use a *kinematic formula* for the perimeter of an FP set. From now on, we will just refer to *the* kinematic formula. In $\mathbb{R}^n$, the kinematic formula expresses the perimeter of an FP set E, as an integral with respect to the natural measure on the space of lines L, of the 1-dimensional perimeter function $\mathrm{PER}(E\cap L)$. In [55], a suitable kinematic formula has been proved for Carnot groups; see (6.1). In that context, a *line* means a horizontal line $L\in\mathrm{lines}(\mathbb{H})$.

Up to a set of measure zero, an FP subset of $\mathbb{R}$ is a finite union of disjoint closed intervals for any open interval I, the perimeter $\mathrm{PER}(E,I)$ is the number of endpoints of these intervals which are contained in I; see [3]. On the other hand, the condition that E is monotone can be reformulated as the requirement that for almost every line L, the 1-dimensional perimeter $\mathrm{PER}(E\cap L)$ is either 0 or 1. Since our initial quantitative data provides an integral bound on the mass of the total perimeter measure (see (2.8) and (3.13)), it is not surprising that the kinematic formula plays a key role in our discussion.

3. Preliminaries

In this section, X denotes a PI space. In particular, it could be $\mathbb{R}^n$ or $\mathbb{H}$. Fix $p\in X$ and let $f\colon B_r(p)\to L_1$ denote a Lipschitz map. In this paper we will often rescale this ball to unit size and correspondingly rescale the map f, and hence the induced metric d_{Σ_f}, or equivalently the cut measure. Finally, we rescale the measure μ, so that the rescaled ball has unit measure.

We set

$$\breve{f} = r^{-1} f, \tag{3.1}$$

$$\Sigma_{\breve{f}} = r^{-1}\Sigma_f, \tag{3.2}$$

$$\breve{d}^X = r^{-1} d^X, \tag{3.3}$$

$$\breve{\mu}(\,\cdot\,) = \frac{\mu(\,\cdot\,)}{\mu(B_r(p))}. \tag{3.4}$$

For the case of Lipschitz maps, we always use the full set of rescalings (3.1)–(3.4).

Note that, for U open, $\operatorname{Per}(E)(U)$ is defined by

$$\operatorname{Per}(E)(U) = \inf_{\{h_i\}} \liminf_{i\to\infty} \int_U \operatorname{Lip}(h_i)\, d\mu, \tag{3.5}$$

where the infimum is taken over all sequences of Lipschitz functions $\{h_i\}_{i=1}^\infty$, with

$$h_i \xrightarrow{L_1^{\mathrm{loc}}} \chi_E.$$

Since a rescaling as in (3.3) has the effect

$$\operatorname{Lip}(h_i) \longmapsto r \operatorname{Lip}(h_i), \tag{3.6}$$

it follows that (3.3) and (3.4) imply that

$$\operatorname{Per}(E) \longmapsto r \frac{1}{\mu(B_r(p))} \operatorname{Per}(E). \tag{3.7}$$

Remark 3.1. When generalizing our considerations to BV functions, or in particular sets of finite perimeter, it is of interest to consider as well the effect of a rescaling as in (3.3) and (3.4), with (3.1) and (3.2) omitted. If (3.1) and (3.2) are omitted, then (3.7) is a relevant rescaling. For the case of FP sets in particular, it is the relevant rescaling. It can be used to give a quantitative analog of Theorem 1.1 for a single FP set E, in which the set of points at which, on a controlled scale, E is not close to a half-space, has small codimension-1 Hausdorff content, as measured with respect to coverings by balls of small radius.

If (3.1) and (3.2) are not omitted, then, by (3.2) and (3.7), the corresponding rescaling factor for the total perimeter measure λ_f is

$$r\frac{1}{\mu(B_r(p))}r^{-1},$$

i.e.,

$$\lambda_f \longmapsto \frac{1}{\mu(B_r(p))}\lambda_f. \tag{3.8}$$

3.1. Normalization

After rescaling f, Σ_f, d^X and μ as in (3.1)–(3.4), we will often denote the rescaling of the ball $B_r(x)$ as $\breve{B}_r(x)$. Thus,

$$\breve{B}_r(x) \subset X, \tag{3.9}$$

$$\breve{\mu}(\breve{B}_r(x)) = 1, \tag{3.10}$$

$$\breve{f}\colon \breve{B}_r(x) \longrightarrow L_1, \tag{3.11}$$

$$\mathrm{Lip}(\breve{f}) = 1, \tag{3.12}$$

or, more generally,

$$\int_{\mathrm{Cut}(\breve{B}_r(x))} \mathrm{Per}(E)(\breve{B}_r(x))\, d\Sigma_{\breve{f}}(E) \leqslant 1. \tag{3.13}$$

3.2. Standard inequalities

We will make repeated use of the trivial inequality (often called Markov's inequality), which states that for a measure space (Y,ν) and $f \in L_1(Y,\nu)$, one has

$$\nu(\{x : f(x) \geqslant t\}) \leqslant \frac{1}{t}\int_Y |f|\, d\nu. \tag{3.14}$$

We also use the weak-type-$(1,1)$ inequality for the maximal function, which is a consequence of the doubling property of the measure; see, e.g., [61, Chapter 1]. We now recall this basic estimate.

Let β denote the doubling constant of μ. Let ζ denote a Radon measure. Fix an open subset U and let $\mathcal{C}$ denote the collection of closed balls $\overline{B_r(y)}$ such that $\overline{B_{5r}(y)} \subset U$. Given $k, j \in \mathbb{N}$, define $\mathcal{B}_{j,k} = \mathcal{B}_{j,k}(\zeta) \subset \mathcal{C}$ by

$$\mathcal{B}_{j,k} := \{\overline{B_r(y)} \in \mathcal{C} : \zeta(\overline{B_r(y)}) \geqslant kr^{-j}\mu(\overline{B_r(y)})\}, \tag{3.15}$$

and set

$$B_{j,k} = \bigcup_{\overline{B_r(q)} \in \mathcal{B}_{j,k}} \overline{B_r(q)}. \tag{3.16}$$

By a standard covering argument (see [37] and [61]), there is a disjoint subcollection $\{B_{r_i}(q_i)\}_{i=1}^{\infty} \subset \mathcal{B}_{j,k}$ such that

$$B_{j,k} \subset \bigcup_{i=1}^{\infty} \overline{B_{5r_i}(q_i)}, \tag{3.17}$$

$$\sum_{i=1}^{\infty} (5r_i)^{-j} \mu(\overline{B_{5r_i}(q_i)}) \leqslant 5^{-j} k^{-1} \beta^3 \zeta(U). \tag{3.18}$$

Relations (3.17) and (3.18) imply that the codimension-j Hausdorff content of $B_{j,k}$ (as defined with respect to the measure μ) is bounded by $5^{-j} k^{-1} \beta^3 \zeta(U)$. The codimension-0 Hausdorff content ($j=0$) is just the measure μ itself.

4. Reduction to the stability of individual monotone sets

In this section we reduce the proof of the main degeneration theorem (Theorem 1.1) to Theorem 4.3, a stability theorem for individual monotone sets, which states roughly that a set which is almost monotone is almost a half-space. The complete proof of Theorem 4.3 will occupy §§7–11.

The next few paragraphs contain an overview of this section. For definiteness, we will restrict our attention to the Heisenberg group. Everything we say applies, mutadis mutandis, to the simpler case of $\mathbb{R}^n$ as well.

We begin by considering an ideal case, the proof of which is given prior to the more general Proposition 4.4. Then we explain how to reduce to the ideal case up to an error which is controlled.

Let $d_{\mathcal{P}}$ denote a cut metric for which the cut measure is supported on cuts $\mathcal{P}$ which are half-spaces. Let $\varepsilon > 0$. We will see by an easy argument that for every line $\underline{L}$ through the identity $e \in \mathbb{H}$, which makes an angle $\geqslant \theta > 0$ with the horizontal plane, for half([5]) of the pairs of points $x_1, x_2 \in \underline{L}$, with $\frac{1}{2}\varepsilon \leqslant d^{\mathbb{H}}(x_1, x_2) \leqslant \frac{3}{2}\varepsilon$, we have

$$d_{\mathcal{P}}(x_1, x_2) \lesssim_\theta \varepsilon d^{\mathbb{H}}(x_1, x_2), \tag{4.1}$$

where the implied constant in (4.1) depends only on θ.

([5]) The measure on pairs (x_1, x_2), with $\frac{1}{2}\varepsilon \leqslant d^{\mathbb{H}}(x_1, x_2) \leqslant \frac{3}{2}\varepsilon$, is $\mathcal{L} \times \mathcal{L}$, where $\mathcal{L}$ denotes the Lebesgue measure associated with the Euclidean metric on $\underline{L}$. Of course, in the statement, "half" can be replaced by any definite fraction <1, provided $\lesssim_\theta$ in (4.1) is allowed to depend on that fraction.

In proving Theorem 1.1, we first find a scale such that at a typical location, the support of the cut measure consists almost entirely of cuts which are almost half-spaces, i.e., cuts which differ from half-spaces by sets of small measure. As explained below, for this, it suffices to find a scale on which the total non-monotonicity is small. By the pigeonhole principle, it is easy to find a scale on which the total perimeter is small. On such a scale we apply Proposition 4.5, which asserts that the total non-monotonicity of a cut metric can be bounded in terms of the total perimeter. Proposition 4.5 is proved in §6.

On a scale on which the total non-monotonicity is small, the effect of cuts which are not almost half-spaces can be absorbed into the error term, after which they can be ignored. However, to get to a situation in which our conclusion can be obtained by appying Theorem 4.3, we must further reduce to one in which there is a suitable bound on the mass of the cut measure. Otherwise, the total effect of the small deviations of the remaining individual cuts from being half-spaces could carry us uncontrollably far from the ideal case considered in (4.1). This point is addressed in Lemma 4.1, which states that the cuts can be decomposed into a subset which makes a contribution which can be absorbed into the error term for the cut metric, and one for which the mass has a definite bound. Lemma 4.1 is proved in §5.

4.1. Controlling the cut measure

The following lemma will be applied to rescaled balls, on a scale on which the total non-monotonicity is sufficiently small. For f Lipschitz, relation (4.2) in the hypothesis of Lemma 4.1 holds for all such balls. If more generally f is BV, then for most such balls, it holds after suitably controlled rescaling.

Given metrics d and d', define $\|d-d'\|_{L^1}$ as in (2.7).

LEMMA 4.1. *Let $f\colon B_1(p)\to L_1$ satisfy*

$$\lambda_f(B_1(p))\leqslant 1. \tag{4.2}$$

Given $\eta>0$, the support of Σ_f can be written as a disjoint union $D_1\cup D_2$, such that if $d_f=d_1+d_2$ denotes the corresponding decomposition of the cut metric d_f, then

$$\Sigma_f(D_1)\leqslant \frac{1}{\eta^3}, \tag{4.3}$$

$$\|d_f-d_1\|_{L_1}\lesssim \eta. \tag{4.4}$$

If f is 1-Lipschitz, then so are d_1 and d_2.

Lemma 4.1 is a consequence of the more general Proposition 5.1, which will be proved in §5.

4.2. Stability of monotone sets

Let lines($\mathbb{H}$) denote the space of unparameterized oriented horizontal lines, and let lines(U) denote the collection of horizontal lines whose intersection with U is non-empty. Let $\mathcal{N}$ denote the unique left-invariant measure on lines($\mathbb{H}$), normalized so that

$$\mathcal{N}(\mathrm{lines}(B_1(e)))=1. \tag{4.5}$$

Let $\mathcal{H}^1_L$ denote 1-dimensional Hausdorff measure on $L\in\mathrm{lines}(\mathbb{H})$ with respect to the metric induced from $d^{\mathbb{H}}$. We note that if $E\subset\mathbb{H}$ is measurable, then $L\cap E$ is measurable for $\mathcal{N}$-a.e. $L\in\mathrm{lines}(\mathbb{H})$.

Given a ball $B_r(x)\subset\mathbb{H}$ and $L\in\mathrm{lines}(B_r(x))$, we define the *non-convexity* of (E,L) on $B_r(x)$, denoted $\mathrm{NC}_{B_r(x)}(E,L)$, by

$$\mathrm{NC}_{B_r(x)}(E,L):=\inf_{\substack{I\subset L\cap B_r(x)\\ I \text{ subinterval}}}\int_{L\cap B_r(x)}|\chi_I-\chi_{E\cap L\cap B_r(x)}|\,d\mathcal{H}^1_L, \tag{4.6}$$

where we allow the subinterval I to be empty in the infimum above. Similarly, we define the *non-monotonicity* of (E,L) on $B_r(x)$ by

$$\mathrm{NM}_{B_r(x)}(E,L):=\mathrm{NC}_{B_r(x)}(E,L)+\mathrm{NC}_{B_r(x)}(E^c,L). \tag{4.7}$$

The *total non-convexity* and *total non-monotonicity* of E on $B_r(X)$ are defined to be the following (scale-invariant) quantities:

$$\mathrm{NC}_{B_r(x)}(E):=\frac{1}{r^4}\int_{\mathrm{lines}(B_r(x))}\mathrm{NC}_{B_r(x)}(E,L)\,d\mathcal{N}(L), \tag{4.8}$$

$$\mathrm{NM}_{B_r(x)}(E):=\frac{1}{r^4}\int_{\mathrm{lines}(B_r(x))}\mathrm{NM}_{B_r(x)}(E,L)\,d\mathcal{N}(L). \tag{4.9}$$

Note that $\mathrm{NM}_{B_r(x)}(E)=0$ if E is monotone, or more generally if the the symmetric difference of E and some monotone subset has measure zero.

Definition 4.2. A cut $E\subset B_r(p)\subset\mathbb{H}$ is said to be δ-*monotone* on $B_r(x)$ if

$$\mathrm{NM}_{B_r(x)}(E)<\delta. \tag{4.10}$$

We have the following stability theorem.

THEOREM 4.3. *There exists $a>0$ (e.g., $a=2^{52}$ works here), such that if the cut $E\subset B_1(x)$ is ε^a-monotone on $B_1(x)$, then there exists a half-space $\mathcal{P}\subseteq\mathbb{H}$ such that*

$$\frac{\mathcal{L}_3((E\cap B_{\varepsilon^3}(x))\triangle\mathcal{P})}{\mathcal{L}_3(B_{\varepsilon^3}(x))}\lesssim\varepsilon.$$

4.3. Cut measures supported on cuts which are almost half-spaces

In this subsection, we begin by verifying (4.1), which concerns the case of a cut measure which is supported on half-spaces. Then, in Proposition 4.4, we consider the more general case of a cut measure which is supported on cuts which are almost half-spaces, i.e., cuts that satisfy the conclusion of Theorem 4.3. Of course, there is an error term in the general case, which leads to a weaker estimate than in the ideal case.

Proof of (4.1). For simplicity, we consider pairs of points (x_1, x_2) with $d^{\mathbb{H}}(x_1, x_2)=\varepsilon$. For $d^{\mathbb{H}}(x_1, x_2)=t$, with $t\in[\frac{1}{2}\varepsilon, \frac{3}{2}\varepsilon]$, the proof is the same.

Consider first an elementary cut metric d_E on the real line, associated with a subset E such that E and E^c are connected. Clearly, if x_3 lies between x_1 and x_2, then

$$d_E(x_1, x_3)+d_E(x_3, x_2) = d_E(x_1, x_2).$$

By linearity, this holds more generally for cut metrics on the line whose cut measures are supported on such cuts.

Let $\Sigma_{\mathcal{P}}$ denote a cut measure on $B_1(p)\subset\mathbb{H}$ such that every cut in the support of $\Sigma_{\mathcal{P}}$ is of the form $\mathcal{P}\cap B_1(p)$, for some half-space $\mathcal{P}$. Then for almost every affine (not necessarily horizontal) line $\underline{L}$, the restriction $d_{\mathcal{P}}|_{\underline{L}}$ is a cut metric as in the previous paragraph.

Consider a subinterval $I\subset\underline{L}\cap B_1(p)$ with endpoints y and z, such that $d^{\mathbb{H}}(y, z)=\frac{1}{2}$, say, and hence $\mathcal{L}(I)\geqslant\frac{1}{4}$. Let ε be as in (4.1) and let $\psi>0$ be arbitrary. Let further α index those intervals $J_\alpha\subset I$ whose endpoints $x_{1,\alpha}$ and $x_{2,\alpha}$ satisfy $d^{\mathbb{H}}(x_{1,\alpha}, x_{2,\alpha})=\varepsilon$ and $d_{\mathcal{P}}(x_{1,\alpha}, x_{2,\alpha})\geqslant\psi d^{\mathbb{H}}(x_{1,\alpha}, x_{2,\alpha})$. Note that if $\underline{L}$ is as in (4.1) (i.e., it makes a definite angle θ with the horizontal plane) then $\mathcal{L}(J_\alpha)\asymp_\theta \varepsilon d^{\mathbb{H}}(x_{1,\alpha}, x_{2,\alpha})\asymp_\theta\varepsilon^2$.

By a standard covering argument, there is a disjoint subcollection $\{J_{\alpha_1}, ..., J_{\alpha_N}\}$ of $\{J_\alpha\}_\alpha$, such that $\bigcup_\alpha J_\alpha\subset 5J_{\alpha_1}\cup...\cup 5J_{\alpha_N}$, and by the length-space property of $d_{\mathcal{P}}$, we have

$$N\psi\varepsilon \leqslant d_{\mathcal{P}}(x_{1,1}, x_{2,1})+...+d_{\mathcal{P}}(x_{1,N}, x_{2,N}) \leqslant d_{\mathcal{P}}(y, z) = \tfrac{1}{2}.$$

Thus, $N\leqslant 1/2\psi\varepsilon$, and since $\mathcal{L}(J_\alpha)\asymp_\theta\varepsilon^2$,

$$\mathcal{L}\biggl(\bigcup_\alpha J_\alpha\biggr) \lesssim_\theta \frac{5\varepsilon^2}{2\psi\varepsilon} = \frac{5\varepsilon}{2\psi}.$$

By taking ψ to be a suitably large multiple (depending on θ) of ε, and incorporating this multiple into the symbol $\lesssim_\theta$ in (4.1), we get $\mathcal{L}(\bigcup_\alpha J_\alpha)\leqslant\frac{1}{2}\mathcal{L}(I)$, which suffices to prove (4.1). □

To simplify the statement of Proposition 4.4, we state it for unit balls. In the application (to Lipschitz maps), we will set $\varepsilon_1=\varepsilon^{40}$ and consider the rescaling as in (3.1)–(3.4) of $B_{\varepsilon_1^3}(x)$ to unit size. Prior to stating the proposition, we give a few preliminaries.

Below, without loss of generality we may (and will) assume that p is the indentity element of $\mathbb{H}$. We consider only affine lines $\underline{L}$ such that $\underline{L}\cap B_{1/2}(p)\neq\varnothing$. Let $\underline{x}\in\underline{L}$ be such that $d^{\mathbb{H}}(\underline{x},p)$ is minimal. Using left translation by $\underline{x}^{-1}$, we identify $\underline{L}$ with a line through the identify and define the angle θ with the horizontal as in (4.1). If θ_0 is sufficiently small and $\underline{L}$ makes *Euclidean angle* $\leqslant\theta_0$ with the vertical, then the angle θ with the horizontal will be $\geqslant\frac{1}{3}\pi$.(6)

Put

$$S=\{\underline{L}:\underline{L} \text{ makes a Euclidean angle} \leqslant\theta_0 \text{ with the vertical and } \underline{L}\cap B_{1/2}(p)\neq\varnothing\}.$$

Let $\mathcal{L}$ denote the *Euclidean* 1-dimensional Lebesgue measure on $\underline{L}$. Note that it differs by a bounded factor from the 1-dimensional Lebesgue measure on $\underline{L}$ induced by the above identification via $\underline{x}^{-1}$, so the distinction between these measures plays no essential role and will be supressed below.

Recall that the space of all affine lines in $\mathbb{H}=\mathbb{R}^3$ is a 4-dimensional homogeneous manifold which carries a natural measure $\mathcal{A}$. (The precise normalization plays no role below.)

PROPOSITION 4.4. *Fix $\psi\in(0,1)$. Assume that $h\colon B_1(p)\to L_1$ is such that Σ_h satisfies the total perimeter bound* (4.2), *and assume that Σ_h is supported on cuts $E\subset B_1(p)$, such that for some half-space $\mathcal{P}_E$,*

$$\mathcal{L}_3((\mathcal{P}_E\cap B_1(p))\triangle E)\lesssim(\psi\varepsilon^4)^4. \tag{4.11}$$

Let $S_0\subset S$ denote the subset of the affine lines $\underline{L}$ such that at least $\frac{2}{3}$ of the pairs $(x_1,x_2)\in(\underline{L}\cap B_1(p))\times(\underline{L}\cap B_1(p))$ satisfy

$$\tfrac{2}{3}\varepsilon\leqslant d^{\mathbb{H}}(x_1,x_2)\leqslant\tfrac{5}{4}\varepsilon, \tag{4.12}$$

$$d_h(x_1,x_2)\leqslant\varepsilon d^{\mathbb{H}}(x_1,x_2). \tag{4.13}$$

Then

$$\mathcal{A}(S_0)\geqslant(1-\psi)\mathcal{A}(S). \tag{4.14}$$

If h is 1-Lipschitz and for all E in the support of Σ_h,

$$\mathcal{L}_3((\mathcal{P}_E\cap B_1(p))\triangle E)\lesssim\varepsilon^{80} \tag{4.15}$$

(6) For the purposes of Proposition 4.4, $\frac{1}{3}\pi$ can actually be replaced by any number $<\frac{1}{2}\pi$.

for some small enough constant $c>0$, then for all $\underline{L}\in S$ and for at least $\frac{1}{2}$ of the pairs $(x_1,x_2)\in\underline{L}\times\underline{L}$ satisfying

$$\tfrac{2}{3}\varepsilon \leqslant d^{\mathbb{H}}(x_1,x_2) \leqslant \tfrac{3}{2}\varepsilon, \tag{4.16}$$

there holds

$$d_h(x_1,x_2) \lesssim \varepsilon d^{\mathbb{H}}(x_1,x_2). \tag{4.17}$$

Proof. Assume first that (4.11) holds. Let D_1, D_2 and $d_h=d_1+d_2$ be as in Lemma 4.1, with $\eta=\psi\varepsilon^4$. Hence, $\Sigma_h(D_1)\leqslant(\psi\varepsilon^4)^{-3}$ and $\|d_h-d_1\|_{L_1}\lesssim\psi\varepsilon^4$.

Using (4.11), it is straightforward to construct a half-space $\mathcal{P}_E$, which varies measurably with respect to Σ_h (alternatively, one can use a measurable selection theorem as in [44]), such that $\mathcal{L}_3((\mathcal{P}_E\cap B_1(p))\triangle E)\lesssim(\psi\varepsilon^4)^4$. Put

$$d_{\mathcal{P},1}(x_1,x_2) = \int_{D_1} d_{\mathcal{P}_E}(x_1,x_2)\, d\Sigma_1(E). \tag{4.18}$$

From $\Sigma_h(D_1)\leqslant(\psi\varepsilon^4)^{-3}$ and $\mathcal{L}_3((\mathcal{P}_E\cap B_1(p))\triangle E)\lesssim(\psi\varepsilon^4)^4$, we obtain

$$\|d_1-d_{\mathcal{P},1}\|_{L_1} \lesssim \psi\varepsilon^4. \tag{4.19}$$

This, together with $\|d_h-d_1\|_{L_1}\lesssim\psi\varepsilon^4$, implies that

$$\|d_h-d_{\mathcal{P},1}\|_{L^1} \lesssim \psi\varepsilon^4. \tag{4.20}$$

Write the integral in the definition of the L_1 norm in (4.20) as an iterated integral, integrating first over pairs of points lying on a given affine line $\underline{L}\in S$, and then over the set of lines $\underline{L}\in S$, with respect to the measure $\mathcal{A}$. It follows from Markov's inequality (3.14) that, for a fraction $\geqslant 1-\psi$ of affine lines $\underline{L}$ in S,

$$\|d_h-d_{\mathcal{P},1}\|_{L_1(\underline{L}\times\underline{L},\mathcal{L}\times\mathcal{L})} \lesssim \varepsilon^4. \tag{4.21}$$

On such a line $\underline{L}$, consider the set of pairs satisfying (4.12) for which (4.1) holds for the metric $d_{\mathcal{P},1}$. By noting that the measure of the space of pairs of points on $\underline{L}$ satisfying (4.12) is $\gtrsim\varepsilon^2$, and applying Markov's inequality once more, on $\underline{L}\times\underline{L}$ in (4.21), we find that for at least $\frac{1}{2}$ of the pairs at which (4.1) holds for $d_{\mathcal{P},1}$, the error term arising from (4.21) is $\lesssim\varepsilon^2$. This gives (4.13).

Finally, assume that h is Lipschitz. Let $\mathrm{Haus}^{\mathbb{R}^3}$ denote Hausdorff distance associated with the Euclidean distance $d^{\mathbb{R}^3}$. We define a metric $\varrho^{\mathbb{R}^3}$ on S by setting, for $\underline{L}_1,\underline{L}_2\in S$,

$$\varrho^{\mathbb{R}^3}(\underline{L}_1,\underline{L}_2) = \mathrm{Haus}^{\mathbb{R}^3}(\underline{L}_1\cap B_1(p),\underline{L}_2\cap B_1(p)).$$

Clearly, with respect to the distance $\varrho^{\mathbb{R}^3}$, the restriction of the measure $\mathcal{A}$ to S is Ahlfors 4-regular. Thus, it follows from (4.14) that if we take $\psi=\eta\varepsilon^{16}=\eta(\varepsilon^4)^4$, for a sufficiently small universal constant η, then S_0 is $\left(\frac{1}{100}\varepsilon\right)^4$-dense in S with respect to $\varrho^{\mathbb{R}^3}$. So, given $\underline{L}\in S$, we can choose $\underline{L}_1\in S_0$ such that $\varrho^{\mathbb{R}^3}(\underline{L},\underline{L}_1)\leqslant\left(\frac{1}{100}\varepsilon\right)^4$.

Let $P:\underline{L}_1\cap B_1(p)\to\underline{L}$ assign to each each $x\in\underline{L}_1\cap B_1(p)$ the closest point on $\underline{L}$. Thus, $d^{\mathbb{R}^3}(x,P(x))\leqslant\left(\frac{1}{100}\varepsilon\right)^4$. Also, at all points, the difference between the differential dP and a Euclidean isometry is bounded by ε^4. Finally, since on $B_1(p)$ we have $d^{\mathbb{H}}\leqslant(10d^{\mathbb{R}^3})^{1/2}$, for $x\in\underline{L}\cap B_1(p)$ we have

$$d^{\mathbb{H}}(x,P(x))\leqslant\varepsilon^2. \tag{4.22}$$

We may assume without loss of genenerality that ε is so small that for at least $\frac{1}{2}$ of the pairs $(x_1,x_2)\in\underline{L}\times\underline{L}$ satisfying (4.12) and (4.13), we will have $(P(x_1),P(x_2))\in B_1(p)\times B_1(p)$ and (4.16) will hold. From (4.22) and the assumption that h is 1-Lipschitz, it follows (using the triangle inequality) that for $\varepsilon<\frac{1}{4}$, say, such a pair satisfies (4.17). □

4.4. Scale estimate; total perimeter and total non-monotonicity

Next, we show how to estimate from below a scale on which, apart from a collection of cuts which contributes negligibly to the cut metric d_f, the hypothesis of Proposition 4.4 will be satisfied at most locations; see Proposition 4.6. It is from this estimate that the logarithmic behavior in (1.1) of Theorem 1.1 arises.

Recall that the cut measure Σ_f is supported on cuts E with finite perimeter. Fix $\delta>0$. Using the structure of finite-perimeter subsets of $\mathbb{H}$ and the kinematic formula, in §6 we will decompose the total perimeter measure λ_f as a sum

$$\lambda_f=\sum_{j=0}^{\infty}\widehat{w}_j, \tag{4.23}$$

in such a way that the measure $\widehat{w}_j$ controls the *total non-monotonicity* on the scale δ^j in the following sense.

For $j>0$ let, as usual, $\breve{B}_{\delta^j/4}(x)$ denote the standard rescaling of the ball $B_{\delta^j/4}(x)\subset B_1(p)$ as in (3.9)–(3.13).

Proposition 4.5. *If for all $r>0$,*

$$\lambda_f(B_r(x))\lesssim\mathcal{L}_3(B_r(x)), \tag{4.24}$$

then

$$\int_{\mathrm{Cut}(\breve{B}_{\delta^j/4}(x))}\mathrm{NM}_{\breve{B}_{\delta^j/4}(x)}(E)\,d\Sigma_f(E)\lesssim\frac{\widehat{w}_j(B_{\delta^j/4}(x))}{\mathcal{L}_3(B_{\delta^j/4}(x))}+\delta. \tag{4.25}$$

In order to apply Proposition 4.5, we need to find j such that $\widehat{w}_j(B_1(p))\leqslant\delta$, and hence, at most locations x, the term

$$\frac{\widehat{w}_j(B_{\delta^j/4}(x))}{\mathcal{L}_3(B_{\delta^j/4}(x))}$$

is $\lesssim\delta$.

PROPOSITION 4.6. *There exists $j\leqslant\delta^{-1}$ for which*

$$\operatorname{Mass}(\widehat{w}_j)\lesssim\delta. \tag{4.26}$$

For such j and at least $\frac{1}{2}$ of the points x in $B_1(p)$,

$$\int_{\operatorname{Cut}(\breve{B}_{\delta^j/4}(x))}\operatorname{NM}_{\breve{B}_{\delta^j/4}(x)}(E)\,d\Sigma_f(E)\lesssim\delta. \tag{4.27}$$

Proof. We have

$$1\gtrsim\operatorname{Mass}(\lambda_f)=\sum_{j=0}^{\infty}\operatorname{Mass}(\widehat{w}_j). \tag{4.28}$$

Thus the number of terms in the sum above for which $\operatorname{Mass}(\widehat{w}_j)\geqslant\delta$ is bounded by $\lesssim\delta^{-1}$, and the claim follows. The conclusion follows from the weak-type-$(1,1)$ inequality for the maximal function applied to the measure $\widehat{w}_j$, i.e., (3.18), together with Proposition 4.5. □

4.5. Proof of Theorem 1.1

Let a be as in Theorem 4.3. Fix $\varepsilon>0$ and let

$$\delta=c\varepsilon^{40(6+a)+2} \tag{4.29}$$

for an appropriately small enough constant $c>0$. By Proposition 4.6, we can choose $j\lesssim\delta^{-1}$, such that (4.27) holds for at least $\frac{1}{2}$ of the points x in $B_1(p)$. Below we restrict our attention to such a point x and rescale $B_{\delta^j/4}(x)$ to standard size, denoting this ball as usual by $\breve{B}_{\delta^j/4}(x)$.

By (4.25), (4.27) and Markov's inequality (3.14), we can write $\operatorname{Cut}(\breve{B}_{\delta^j/4}(x))$ as a disjoint union

$$\operatorname{Cut}(\breve{B}_{\delta^j/4}(x))=D_3\cup D_4,$$

with corresponding metric decomposition $d_f=d_3+d_4$, where

$$\Sigma_f(D_3)\leqslant c\varepsilon^{40\cdot 6+2}, \tag{4.30}$$

$$\operatorname{NM}_{\breve{B}_{\delta^j/4}(x)}(E)\leqslant\varepsilon^{40a}\qquad\text{for all } E\in D_4. \tag{4.31}$$

By (4.30), we have $d_3(x_1, x_2) < c\varepsilon^{40\cdot 6+2}$ for all x_1 and x_2. Therefore, if d_f is restricted to $B_{\varepsilon^{40\cdot 6}}(x) = B_{(\varepsilon^{80})^3}(x)$ and this ball is rescaled to unit size, we get

$$|d_f - d_4| \leqslant \varepsilon^2, \tag{4.32}$$

while, by (4.31) and Theorem 4.3, d_4 is supported on cuts which satisfy the hypothesis of Proposition 4.4. From Proposition 4.4, together with (4.32), Theorem 1.1 follows.

4.6. Preview of the proof of Theorem 4.3

After two preliminary technical results on δ-monotone subsets of $\mathbb{H}$ have been stated in §7, and the classification of monontone subsets of $\mathbb{H}$ has been reviewed in §8, Theorem 4.3 is proved in §9. However, there is a technical step in the argument (the non-degeneracy of the initial configuration) the proof of which, for reasons of exposition, is deferred until §10. The proofs of the technical results, Lemma 7.2 and Proposition 7.3, are deferred to §11 and §12. The proof of Proposition 7.3 is (by far) the most involved part of this paper.

5. Cuts with small perimeter

In this section we prove Lemma 4.1, which shows that cuts with very small perimeter make a negligible contribution to the cut metric d_f. It is most natural to argue here in the context of general PI spaces. Thus, let X denote a PI space, for example $\mathbb{R}^n$ or $\mathbb{H}$. Fix $p \in X$ and let $f: B_1(p) \to L_1$ satisfy $\operatorname{Lip}(f) \leqslant 1$.

Put

$$D_1 = \{E : \operatorname{Per}(E)(B_1(p)) \leqslant \theta\} \quad \text{and} \quad D_2 = \{E : \operatorname{Per}(E)(B_1(p)) > \theta\}, \tag{5.1}$$

and let $d_f = d_1 + d_2$ denote the corresponding decomposition of the metric d_f. Let β denote the doubling constant of X, and $\tau', \chi > 1$ be the constants in the Poincaré–Sobolev inequality (2.4).

PROPOSITION 5.1. *If $f: B_1(p) \to L_1$ satisfies*

$$\lambda_f(B_1(p)) \leqslant K, \tag{5.2}$$

then, for all $\theta > 0$,

$$\Sigma_f(D_2) \leqslant \frac{K}{\theta}, \tag{5.3}$$

$$\|d_f - d_2\|_{L_1} \leqslant \frac{2\tau' K}{1 - 2^{\chi-1}} \theta^{\chi-1}. \tag{5.4}$$

Proof. From (5.2) and Markov's inequality (3.14), we get (5.3). For $n\in\mathbb{N}$ put

$$D_{1,n}=\{E:\theta 2^{-(n+1)}\leqslant \mathrm{Per}(E)(B_1(p))\leqslant \theta 2^{-n}\}.$$

Then $D_1=\bigcup_{n\in\mathbb{N}} D_{1,n}$. By (5.2) and Markov's inequality once more, we have

$$\Sigma_f(D_{1,n})\leqslant \frac{K2^{n+1}}{\theta}.$$

Moreover, if $E\in D_{1,n}$ then, by (2.4),(7)

$$\fint_{B_1(p)\times B_1(p)} |\chi_E(x_1)-\chi_E(x_2)|\,d\mu\times d\mu\leqslant \tau'(\theta 2^{-n})^{\chi}. \tag{5.5}$$

By summing over n, we get (5.4). □

Note that when $\mathrm{Lip}(f)\leqslant 1$, by virtue of (2.8), (5.2) holds with $K=c(\beta,\tau')$. To get (4.3) and (4.4) from (5.3) and (5.4), we take θ to be a suitable multiple of η^3 (noting that for $X=\mathbb{H}$, we have $\beta=16$ and $\chi=\frac{4}{3}$).

6. The kinematic formula and δ-monotone sets

In this section, using the kinematic formula, we decompose the total perimeter measure as a sum of measures $\lambda_f=w_1+w_2+\dots$. Then we prove Proposition 4.5, which states that w_j controls the total non-monotonicity on the scale δ^j.

We rely on the simple structure of sets of finite perimeter in dimension 1 and on the kinematic formula, which expresses the perimeter of an FP subset E of $\mathbb{R}^n$ or a Carnot group, as an integral over the space of lines L of the perimeters of the 1-dimensional FP sets $E\cap L$.

6.1. FP sets in dimension 1

Let V denote an open subset of $\mathbb{R}$ and let $F\subset V$ have finite perimeter, i.e., $\mathrm{Per}(F)(V)<\infty$. Then there exists a unique collection of finitely many disjoint intervals, $I_1(F),\dots,I_N(F)$, which are relatively closed in V, such that the symmetric difference of F and

$$\mathcal{I}(F)=\bigcup_{i=1}^{N} I_i(F)$$

(7) It follows immediately from (3.5), the definition of perimeter, that if (2.4) holds for Lipschitz functions, then it holds for FP sets (in which case it becomes an isoperimetric inequality). A corresponding statement is valid for more general BV functions.

has measure zero. Moreover, the perimeter measure $\mathrm{Per}(F)$ is a sum of delta functions concentrated at the endpoints of these intervals, and the perimeter $\mathrm{Per}(F)(V)$ is equal to the number of endpoints; see [3, Proposition 3.52]. In what follows, we will often assume without explicit mention that the set $F\subset\mathbb{R}$ has been replaced by its precise representative $\mathcal{I}(F)$. Note that $\mathcal{I}(F^c\cap V)=\mathcal{I}(F)^c$ and $\mathrm{Per}(F^c)=\mathrm{Per}(F)$.

6.2. The kinematic formula

A kinematic formula exists for the PI space $(\mathbb{H}, d^X, \mathcal{L})$, and more generally for any Carnot group, an in particular also for $\mathbb{R}^n$; see [55, Proposition 3.13]. Below, the notation is as introduced prior to (4.5).

Let $U\subset\mathbb{H}$ denote an open subset such that $\mathrm{Per}(E)(U)<\infty$. The kinematic formula states that the function $L\mapsto\mathrm{Per}(E\cap L)(U\cap L)$ lies in $L_1(\mathrm{lines}(U),\mathcal{N})$, and in addition (for some constant $c=c(\mathbb{H})>0$)

$$\mathrm{Per}(E)(U)=c\int_{\mathrm{lines}(U)}\mathrm{Per}(E\cap L)(U\cap L)\,d\mathcal{N}(L). \tag{6.1}$$

6.3. Perimeter bounds non-monotonicity

Fix $0<\delta<1$. For all E and L as above and all $j\geqslant 0$, let $C_j(E,L)$ denote the collection of intervals $I(E,L)$ occuring in $\mathcal{I}(E\cap L\cap B_1(x))$ such that

$$\delta^{j+1}\leqslant\mathrm{length}(I(E,L))<\delta^j. \tag{6.2}$$

Let $\mathcal{E}_j(E,L)$ denote the collection of all endpoints of intervals in $C_j(E,L)$. Let $\mathrm{card}(S)$ denote the cardinality of the set S. For $c(\mathbb{H})$ as in (6.1) and $A\subset B_1(x)$, put

$$w_j(E,L)(A)=c(\mathbb{H})\,\mathrm{card}(\{e\in\mathcal{E}_j(E,L):e\in A\}), \tag{6.3}$$

$$w_j(E)(A)=\int_{\mathrm{lines}(B_1(x))}w_j(E,L)(A)\,d\mathcal{N}(L), \tag{6.4}$$

$$\widehat{w}_j(E)(A)=\widehat{w}_j(E^c)(A)=\frac{w_j(E)(A)+w_j(E^c)(A)}{2}. \tag{6.5}$$

Finally, set

$$\widehat{w}_j(A)=\int_{\mathrm{Cut}(B_1(x))}\widehat{w}_j(E)(A)\,d\Sigma_f(E). \tag{6.6}$$

By the kinematic formula (6.1), we get the decompositions of the perimeter measure and of the total perimeter measure

$$\operatorname{Per}(E) = \sum_{j=0}^{\infty} \widehat{w}_j(E), \tag{6.7}$$

$$\lambda_f = \sum_{j=0}^{\infty} \widehat{w}_j, \tag{6.8}$$

which is a restatement of (4.23).

Proof of Proposition 4.5. Fix E with finite perimeter. For $\mathcal{N}$-a.e.

$$L \in \operatorname{lines}(B_{\delta^j/4}(x)),$$

we may assume that $L \cap E \cap B_{\delta^j/4}(x)$ and $L \cap E^c \cap B_{\delta^j/4}(x)$ consist of finitely many intervals $I_1, ..., I_N$, in the natural consecutive ordering. Note that all the intervals $I_2, ..., I_{N-1}$, including both endpoints, are contained in $B_{\delta^j/4}(x)$. In particular, these intervals lie in $C_k(E, L)$ for various $k \geqslant j$.

Since the non-monotonicity $\mathrm{NM}_{B_{\delta^j/4}(x)}(E, L)$ is defined as an infimum over intervals (see (4.7)), it can be bounded from above by employing any specific interval I. For definiteness, assume that $I_1 \subset E$, where we allow $I_1 = \varnothing$. Then put

$$I = \begin{cases} L \cap B_{\delta^j/4}(x), & \text{if } I_N \subset E, \\ I_1^c \cap B_{\delta^j/4}(x), & \text{if } I_N \subset E^c. \end{cases}$$

Rescale the ball $B_{\delta^j/4}(x)$ to unit size as in (3.9)–(3.13). Then, by employing the above chosen interval I, we get

$$\mathrm{NM}_{B_{\delta^j/4}(x)}(E, L) \lesssim \sum_{k \geqslant j} \sum_{I \in C_k(E,L)} \operatorname{length}(I). \tag{6.9}$$

Since an interval is determined by its endpoints, (6.9) implies that

$$\mathrm{NM}_{B_{\delta^j/4}(x)}(E, L) \lesssim \sum_{k \geqslant j} \delta^{k-j} \operatorname{Mass}(\widehat{w}_k(E, L)), \tag{6.10}$$

and by integrating over the space of lines, we get

$$\mathrm{NM}_{B_{\delta^j/4}(x)}(E) \lesssim \operatorname{Mass}(\widehat{w}_j) + \delta \operatorname{Mass}(\operatorname{Per}(E)), \tag{6.11}$$

which completes the proof of Proposition 4.5. $\square$

7. The quantitative interior and boundary

The proof (though not the statement) of Theorem 4.3 utilizes quantitative notions of the interior and boundary of a measurable set E. In the present section, after defining these notions, we state two key properties for δ-convex and δ-monotone subsets of $\mathbb{H}$; see Proposition 7.3 and Lemma 7.2, respectively. We also deduce two particular consequences of Proposition 7.3 for the structure of the quantitative boundary of δ-monotone sets. These are used in the proof of Theorem 4.3. Assuming Lemma 7.2 and Proposition 7.3, the proof of Theorem 4.3 is given in §9. The proofs of Lemma 7.2 and Proposition 7.3, which are the most technical parts of our discussion, are postponed until §11 and §12, respectively.

Definition 7.1. Let (X, d^X, μ) be a metric measure space and $E\subset X$ be a measurable subset. For $\alpha\in(0,1)$ and $u>0$ define

$$\operatorname{int}_{\alpha,u}(E):=\left\{x: \frac{\mu(E\cap B_u(x))}{\mu(B_u(x))}\geqslant 1-\alpha\right\}, \tag{7.1}$$

$$\partial_{\alpha,u}(E):=\left\{x: \alpha<\frac{\mu(E\cap B_u(x))}{\mu(B_u(x))}<1-\alpha\right\}. \tag{7.2}$$

Observe that $\partial_{\alpha,u}(E)=\varnothing$ for $\alpha\geqslant\frac{1}{2}$. We note the following properties which are trivial consequences of the definitions.

If $\beta\leqslant\alpha$, then

$$\operatorname{int}_{\beta,u}(E)\subset\operatorname{int}_{\alpha,u}(E) \quad\text{and}\quad \partial_{\alpha,u}(E)\subset\partial_{\beta,u}(E). \tag{7.3}$$

For $u_2\leqslant u_1$ and $x\in X$ set

$$c(u_1,u_2,x)=\frac{\mu(B_{u_1}(x))}{\mu(B_{u_2}(x))}.$$

Then

$$\begin{aligned} x\in\operatorname{int}_{\alpha,u_1}(E) \quad&\Longrightarrow\quad x\in\operatorname{int}_{c(u_1,u_2,x)\alpha,u_2}(E),\\ x\in\partial_{c(u_1,u_2,x)\alpha,u_2}(E) \quad&\Longrightarrow\quad x\in\partial_{\alpha,u_1}(E). \end{aligned} \tag{7.4}$$

If $x\in\partial_{\alpha,u}(E)$, then

$$x\in\operatorname{int}_{1-\alpha,u}(E)\cap\operatorname{int}_{1-\alpha,u}(E^c). \tag{7.5}$$

Finally,

$$(\partial_{\alpha,u}(E))^c=\operatorname{int}_{\alpha,u}(E)\cup\operatorname{int}_{\alpha,u}(E^c). \tag{7.6}$$

7.1. Quantitative interior of δ-monotone subsets

If a subset E of a ball $B_1(p)$ is close to a half-space, then it follows that there exists a sub-ball of a definite size which is almost entirely contained in either E or E^c. Conversely, proving that δ-monotone subsets have this property constitutes an important step in the proof that such a set is close to a half-space. (It is the remainder of the argument, based on Proposition 7.3 below, which requires that we shrink the size of the ball on which the conclusion is obtained.) This is the content of Lemma 7.2, which constitutes both Step A of the proof in the precisely monotone case, treated in §8, and more generally Step A′ of the case $\delta>0$, treated in §9.

LEMMA 7.2. *There exists $0<c<\frac{1}{2}$ such that if $E\subset B_r(p)$ is ε^2-monotone on $B_r(p)$ then there exists $q\in B_{r/2}(p)$ such that $q\notin\partial_{\varepsilon,cr}(E)$, i.e., $q\in\text{int}_{\varepsilon,cr}(E)\cup\text{int}_{\varepsilon,cr}(E^c)$.*

7.2. Quantitative convexity of δ-convex sets

If $E\subset\mathbb{R}^n$ is convex and L is a line passing through the points $p\in E$ and $q\in\text{int}(E)$, then the segment of L lying between p and q also consists of interior points of E. Below, for subsets of $\mathbb{H}$ which are almost convex, we give a quantitative version of this statement.

We say that $E\subset B_r(p)$ is *δ-convex* on $B_r(p)$ if, for $\text{NC}_{B_r(p)}(E,L)$ as in (4.6),

$$\text{NC}_{B_r(p)}(E)=\int_{\text{lines}(\breve{B}_r(x))}\text{NC}_{B_r(p)}(E,L)\,d\mathcal{N}(L)<\delta. \tag{7.7}$$

PROPOSITION 7.3. *There exist universal constants $c\in(0,1)$ and $C\in(1,\infty)$ with the following properties. Fix $\varkappa,\eta,\xi,r,\varrho\in(0,1)$ such that*

$$\varrho\leqslant\min\left\{\tfrac{1}{2}\varkappa r^2,c\right\}. \tag{7.8}$$

Set

$$\delta_1=\frac{c\varkappa^3\eta^2\xi\varrho^3}{r}, \tag{7.9}$$

$$\delta_2=c\varkappa^6\eta^3\xi^2\varrho^3r^6. \tag{7.10}$$

Let $L\in\text{lines}(\mathbb{H})$ be parameterized by arc length and assume that

$$L(0)\in\text{int}_{1-\eta,\varrho}(E), \tag{7.11}$$

$$L(1)\in\text{int}_{\delta_1,Cr}(E) \tag{7.12}$$

and $E\subset\mathbb{H}$ is δ_2-convex on $B_{2C}(L(0))$. Then, for all $s\in[\varkappa,1]$, we have

$$L(s)\in\text{int}_{\xi,csr}(E). \tag{7.13}$$

Remark 7.4. The importance of the proposition is the following: even if $L(0)$ is only in the very weak quantitative interior of E (η small), we can ensure that $L(s)$ is in the very strong quantitative interior of E (i.e., ξ small), provided that $L(1)$ is in the sufficiently strong quantitative interior of E (δ_1 sufficiently small) and E is sufficiently convex (δ_2 sufficiently small).

Remark 7.5. The application to the proof of Theorem 4.3 requires some properties of the quantitative boundary for δ-monotone sets which are derived in the next subsection using Proposition 7.3. The assumption $L(0)\in\mathrm{int}_{1-\eta,\varrho} E$ is guaranteed by assuming that $L(0)\in\partial_{\eta,\varrho}E$; see (7.5).

Remark 7.6. Note that the constants $0<\varkappa,\eta,\xi,\varrho\leqslant 1$ can be chosen arbitrarily small, provided r is chosen to satisfy (7.8). The particular form of (7.8) reflects the multiplicative structure of $\mathbb{H}$; compare Remark 12.4. In view of (7.4), by taking ξ a definite amount smaller if necessary, we can replace csr in (7.13) by any smaller positive radius.

7.3. Lines in quantitative boundaries of δ-monotone sets

Given that any proper monotone subset E is a half-space, and hence that ∂E is a 2-plane, any line $L\in\mathrm{lines}(\mathbb{H})$ intersecting ∂E in more than one point is entirely contained in ∂E. Moreover, ∂E is a union of such lines $L\in\mathrm{lines}(\mathbb{H})$. Conversely, these statements form two key substeps in the proof that monotone subsets $\mathbb{H}$ are half-spaces; see Steps B.1 and B.2 in §8.

Essentially, Corollaries 7.8 and 7.10 of Proposition 7.3 constitute the corresponding substeps in the proof that δ-monotone subsets are close to half-spaces. Specifically, if we quantify the hypotheses of the above two statements, then the conclusions hold in a weaker quantitative sense. Prior to tackling these corollaries, whose statements are somewhat complicated, the reader may wish to look at the proofs of the above mentioned substeps given in §8. Here and in §8, the skeleton of the argument is precisely the same, but in §8 the technical complications are absent.

Below, we have sacrificed some sharpness to avoid further complication in the relevant expressions.

COROLLARY 7.7. *Let C and c be the constants from Proposition* 7.3. *For every $\varkappa,\alpha_1,\alpha_2\in(0,1)$, $u_1\in\left(0,\frac{1}{2}\varkappa^3\right)$ and $u_2\in(0,c\varkappa^2)$ define*

$$\gamma=C\max\left\{\sqrt{\frac{2u_1}{\varkappa}},\frac{u_2}{c\varkappa}\right\}, \tag{7.14}$$

$$\beta=\frac{C^5\alpha_1^2\alpha_2u_1^3u_2^4\varkappa}{c^3\gamma^5}, \tag{7.15}$$

$$\delta = \frac{C^2 \alpha_1^3 \alpha_2^2 u_1^3 u_2^8 \varkappa}{c^7 \gamma^2}. \tag{7.16}$$

Fix $L \in \text{lines}(\mathbb{H})$ *which is parameterized by arc length and* $E \subset B_{2C}(L(0))$ *which is* δ*-monotone on* $B_{2C}(L(0))$. *Assume that*

$$x_1 = L(0) \in \partial_{\alpha_1,u_1}(E) \quad \textit{and} \quad x_2 = L(\varkappa) \in \partial_{\alpha_2,u_2}(E),$$

so that in particular $d^{\mathbb{H}}(x,y) = \varkappa$. *Then, for all* $t \in [\varkappa, 1]$, *we have*

$$L(t) \in \partial_{\beta,\gamma}(E).$$

Proof. Assume, for the sake of contradiction, that $L(t) \notin \partial_{\beta,\gamma}(E)$. Thus, without loss of generality, we may assume that, say, $x_1 \in \text{int}_{1-\alpha_1,u_1}(E)$ and $L(t) \in \text{int}_{\beta,\gamma}(E)$. In order to apply Proposition 7.3, rescale the metric $d^{\mathbb{H}} \mapsto d^{\mathbb{H}}/t$, so that the rescaled distance between x_1 and $L(t)$ is equal to 1. Let $L'(s) = L(ts)$, so that L' is parameterized by arc length in the rescaled metric. We will apply Proposition 7.3 with the parameters

$$\eta = \alpha_1, \quad \varrho = \frac{u_1}{t}, \quad r = \frac{\gamma}{Ct}, \quad s = \frac{\varkappa}{t} \quad \text{and} \quad \xi = \alpha_2 \left(\frac{Ctu_2}{c\gamma\varkappa} \right)^4. \tag{7.17}$$

Note that with this notation the parameter β, as defined in (7.15), is at most δ_1 as given in (7.9) (where we used the assumption $t \geqslant \varkappa$). Hence in the rescaled metric, and using the notation in (7.17), we have $x_1 = L'(0) \in \text{int}_{1-\eta,\varrho}(E)$ and $L(t) = L'(1) \in \text{int}_{\delta_1,Cr}(E)$. Moreover, after rescaling, E is δ-monotone on $B_{2C/t}(x_1)$, and hence E is (δ/t^4)-monotone on $B_{2C}(x_1)$. Observe that δ/t^4, with δ given in (7.16), is at most δ_2 as defined in (7.10) (using the assumption $t \geqslant \varkappa$). We are therefore in position to apply Proposition 7.3, provided that we check that with our definitions $\varkappa, \eta, \xi, r, \varrho \in (0,1)$, $s \in [\varkappa, 1]$ and (7.8) is satisfied. These facts are ensured by the assumptions $u_1 < \frac{1}{2}\varkappa^3$, $u_2 < C\varkappa^2$, $t \in [\varkappa, 1]$ and (7.14). So, the conclusion of Proposition 7.3 says that in the rescaled metric we have $x_2 = L'(s) \in \text{int}_{\xi,csr}(E)$.

Rescaling back to the original metric $d^{\mathbb{H}}$, we see that $x_2 \in \text{int}_{\xi,csrt}(E) = \text{int}_{\xi,c\gamma\varkappa/Ct}(E)$. In other words,

$$\begin{aligned} \mathcal{L}_3(E^c \cap B_{c\gamma\varkappa/Ct}(x_2)) &\leqslant \xi \mathcal{L}_3(B_{c\gamma\varkappa/Ct}(x_2)) \\ &= \xi \left(\frac{c\gamma\varkappa}{Ctu_2} \right)^4 \mathcal{L}_3(B_{u_2}(x_2)) \\ &= \alpha_2 \mathcal{L}_3(B_{u_2}(x_2)), \end{aligned} \tag{7.18}$$

where the last equality follows by (7.17). Note that $c\gamma\varkappa/Ct \geqslant u_2$, as ensured by (7.14) (since $t \leqslant 1$), and hence $B_{c\gamma\varkappa/Ct}(x_2) \supset B_{u_2}(x_2)$. Moreover, the assumption $x_2 \in \partial_{\alpha_2,u_2}(E)$ implies that

$$\mathcal{L}_3(E^c \cap B_{c\gamma\varkappa/Ct}(x_2)) \geqslant \mathcal{L}_3(E^c \cap B_{u_2}(x_2)) > \alpha_2 \mathcal{L}_3(B_{u_2}(x_2)). \tag{7.19}$$

Inequalities (7.18) and (7.19) yield the desired contradiction. □

Corollary 7.8. *Let C and c be the constants from Proposition* 7.3. *For every* $\varkappa\in\left(0,\frac{1}{2}\right]$, $\alpha_1,\alpha_2\in(0,1)$, $u_1\in(0,\varkappa^7/8C^2)$ *and* $u_2\in(0,c\varkappa^4/2C)$, *define*

$$\gamma_*=C^{3/2}\left(\frac{2}{\varkappa}\right)^{1/2}\max\left\{\left(\frac{2u_1}{\varkappa}\right)^{1/4},\left(\frac{u_2}{c\varkappa}\right)^{1/2}\right\}, \tag{7.20}$$

$$\beta_*=\frac{128C^{29}\alpha_1^4\alpha_2^3u_1^6u_2^{12}}{\varkappa^5\gamma_*^{19}}, \tag{7.21}$$

$$\delta_*=\frac{2^{16}C^{41}\alpha_1^6\alpha_2^5u_1^9u_2^{20}}{81c^{16}\varkappa^8\gamma_*^{26}}. \tag{7.22}$$

Fix $L\in\text{lines}(\mathbb{H})$, *parameterized by arc length, and a subset* $E\subset B_{3C}(L(0))$, *which is* δ_**-monotone on* $B_{3C}(L(0))$. *Assume that*

$$x_1=L(0)\in\partial_{\alpha_1,u_1}(E)\quad\textit{and}\quad x_2=L(\varkappa)\in\partial_{\alpha_2,u_2}(E),$$

so that in particular $d^{\mathbb{H}}(x,y)=\varkappa$. *Then, for all* $t\in[0,1]$, *we have*

$$L(t)\in\partial_{\beta_*,\gamma_*}(E). \tag{7.23}$$

Proof. The case $t\in[\varkappa,1]$ is a simple consequence of Corollary 7.7. Indeed, since E is δ_*-monotone on $B_{3C}(L(0))$, it is also $\left(\frac{3}{2}\right)^4\delta_*$-monotone on $B_{2C}(L(0))$. Note that $\left(\frac{3}{2}\right)^4\delta_*\leqslant\delta$, where δ is given in (7.16) (this follows immediately from the definitions and our assumed upper bounds on u_1 and u_2). We can therefore apply Corollary 7.7 to deduce that $L(t)\in\partial_{\beta,\gamma}(E)$, where β and γ are given in (7.15) and (7.14), respectively. Note that $\gamma_*\geqslant\gamma$, and hence, by (7.4), we have

$$L(t)\in\partial_{(\gamma/\gamma_*)^4\beta,\gamma_*}(E). \tag{7.24}$$

Since $\beta_*\leqslant(\gamma/\gamma_*)^4\beta$, we can use (7.3) to deduce that $L(t)\in\partial_{\beta_*,\gamma_*}(E)$, as required.

To deal with the case $t\in[0,\varkappa]$, note that, since as explained above E is δ-monotone on $B_{2C}(L(0))$, by Corollary 7.7 we have $L(2\varkappa)\in\partial_{\beta,\gamma}(E)$. We wish to apply Corollary 7.7 to the points $L(2\varkappa)$, $x_2=L(\varkappa)$ and $L(t)$, and to the geodesic $L'(s)=L(-s+2\varkappa)$ (so that $L'(0)=L(2\varkappa)$, $L'(\varkappa)=L(\varkappa)$ and $L'(2\varkappa-t)=L(t)$, where $2\varkappa-t\in[\varkappa,1]$), with (α_1,u_1) replaced by (β,γ) and (α_2,u_2) unchanged. It follows from (7.14) and (7.20) that

$$\gamma_*=C\sqrt{\frac{2\gamma}{\varkappa}}=C\max\left\{\sqrt{\frac{2\gamma}{\varkappa}},\frac{u_2}{c\varkappa}\right\}, \tag{7.25}$$

where in the last step above we used the fact that $\gamma\geqslant Cu_2/c\varkappa$, which implies that

$$\sqrt{\frac{2\gamma}{\varkappa}}\geqslant\frac{u_2}{c\varkappa},$$

by our assumption on u_2. We can restate (7.25) saying that γ_* corresponds to the quantity γ of Corollary 7.7, where we substitute our new choice of parameters. Moreover, by substituting these values into (7.16) and (7.15), we see that for these new parameters $\left(\frac{3}{2}\right)^4\delta_*$ is equal to the value of δ from (7.16), and β_* is equal to the value of β from (7.21) (we chose the values of β_* and δ_* precisely for this purpose). Note that since E is δ_*-monotone on $B_{3C}(L(0))\supset B_{2C}(L(2\varkappa))$, it is also $\left(\frac{3}{2}\right)^4\delta_*$-monotone on $B_{2C}(L(2\varkappa))$. In order to apply Corollary 7.7, we also need to ensure that

$$\gamma = C\max\left\{\sqrt{\frac{2u_1}{\varkappa}}, \frac{u_2}{c\varkappa}\right\} \leqslant \frac{\varkappa^3}{2} \quad \text{and} \quad \beta = \frac{C^5\alpha_1^2\alpha_2 u_1^3 u_2^4\varkappa}{c^3\gamma^5} \leqslant 1,$$

which is indeed the case, as follows from the assumed upper bounds on u_1 and u_2. The conclusion of Corollary 7.7 is precisely (7.23). □

COROLLARY 7.9. *Let C and c be the constants from Proposition 7.3. Fix $\alpha, u\in(0,1)$, $t\in\left(0,\frac{1}{3}\right]$ and $d>0$ such that*

$$\frac{2C}{c}u \leqslant d \leqslant \min\left\{2C, \frac{C^2}{2}\right\}t, \tag{7.26}$$

$$\alpha u^4 \leqslant \min\left\{\frac{c^3}{2C^7}d^3t, \frac{8c^4}{C^5}dt^3\right\}. \tag{7.27}$$

Set

$$\phi = \frac{C^{20}}{2^9c^{12}}\frac{\alpha^4u^{16}}{d^8t^8}, \tag{7.28}$$

$$\zeta = \frac{C^{17}}{2^{12}c^{16}}\frac{\alpha^5u^{20}}{t^{11}d^5}. \tag{7.29}$$

Assume that $x\in\mathbb{H}$ and $E\subset B_{2C}(x)$ are such that $x\in\partial_{\alpha,u}(E)$. Assume also that E is ζ-monotone on $B_{2C}(x)$. Then there exists $L\in\mathrm{lines}(\mathbb{H})$, with $L(0)=x$, such that either $L(t)$ or $L(-t)$ is in $\partial_{\phi,d}(E)$.

Proof. Take any $L\in\mathrm{lines}(\mathbb{H})$ with $L(0)=x$. Note that (7.26)–(7.28) imply that $\phi\leqslant\frac{1}{2}$. Hence, we are done if either $L(-t)$ or $L(t)$ is in $\mathrm{int}_{\phi,d}(E)$ and the other is in $\mathrm{int}_{\phi,d}(E^c)$. Indeed in this case, since the collection of all oriented lines through x is connected (it can be identified with the unit circle), the required result follows from the intermediate-value theorem.

Assume for the sake of contradiction that the assertion of Corollary 7.9 is false. Then, by the above discussion, $\{L(-t),L(t)\}\subset\mathrm{int}_{\phi,d}(E)$ or $\{L(-t),L(t)\}\subset\mathrm{int}_{\phi,d}(E^c)$. So, assume without loss of generality that $\{L(-t),L(t)\}\subset\mathrm{int}_{\phi,d}(E)$. In order to apply Proposition 7.3 we rescale the metric $d^{\mathbb{H}}\mapsto d^{\mathbb{H}}/2t$. Set $y_1=L(-t)$ and $y_2=L(t)$, so that in the rescaled metric the distance between y_1 and y_2 is 1, and their distance from x is $\frac{1}{2}$. The geodesic $L'(s)=L(2ts-t)$ is parameterized by arc length in the rescaled metric with $L'(0)=y_1$, $L'(\frac{1}{2})=x$ and $L'(1)=y_2$. We will apply Proposition 7.3 to the geodesic L' with the parameters

$$\varkappa=\eta=s=\frac{1}{2}, \quad \xi=\alpha\left(\frac{2Cu}{cd}\right)^4, \quad r=\frac{d}{2Ct} \quad \text{and} \quad \varrho=\frac{C^5}{8c^3}\frac{\alpha u^4}{dt^3}. \tag{7.30}$$

Note that (7.26) and (7.27) ensure that $\varrho\leqslant\min\left\{\frac{1}{2}\varkappa r^2, c\right\}$ and $r,\xi\leqslant 1$. In the rescaled metric $y_1,y_2\in\mathrm{int}_{\phi,d/2t}(E)=\mathrm{int}_{\phi,Cr}(E)$. Since $\varrho\leqslant d/2t$, we may conclude from (7.4) that

$$y_2\in\mathrm{int}_{(d/2t\varrho)^4\phi,\varrho}(E)=\mathrm{int}_{\eta,\varrho}(E),$$

where the equality follows by (7.28) and (7.30).

In the rescaled metric E is ζ-monotone on $B_{C/t}(x)$. Since $t\leqslant\frac{1}{3}$, we have $B_{2C}(y_1)\subset B_{C/t}(x)$, and therefore E is $\zeta/(2t)^4$-monotone on $B_{2C}(y_1)$. Observe that with the parameters set as in (7.30), we have $\delta_1=\phi$ and $\delta_2=\zeta/(2t)^4$, where δ_1 and δ_2 are as in (7.9) and (7.10). Therefore, we can apply Proposition 7.3, which implies that (in the rescaled metric)

$$x\in\mathrm{int}_{\xi,csr}(E)=\mathrm{int}_{\alpha(2Cu/cd)^4,cd/4Ct}(E), \tag{7.31}$$

where the equality follows by (7.30).

But, in the rescaled metric we know that $x\in\partial_{\alpha,u/2t}(E)$. It follows from (7.26) that $u/2t\leqslant cd/4Ct$. Hence, by (7.4) and (7.6), we deduce from (7.31) that

$$x\in\mathrm{int}_{\alpha,u/2t}(E)\subset\partial_{\alpha,u/2t}(E)^c,$$

which yields the desired contradiction. □

Corollary 7.10. *There exist universal constants $\bar{c},\overline{C}>0$ with the following properties. Fix $\alpha\in(0,1)$ and $u\in(0,\bar{c})$. Let $x\in\mathbb{H}$ and $E\subset B_{\overline{C}}(x)$ be δ-monotone on $B_{\overline{C}}(x)$, where*

$$\delta=\bar{c}\alpha^{29}u^{71}. \tag{7.32}$$

Assume also that $x\in\partial_{\alpha,u}(E)$. Then there exists $L\in\mathrm{lines}(\mathbb{H})$, parameterized by arc length, such that $x=L(0)$ and for all $s\in\left[-\frac{1}{3},\frac{1}{3}\right]$ we have

$$L(s)\in\partial_{\bar{c}\alpha^{19}u^{46},\overline{C}u^{1/4}}(E). \tag{7.33}$$

Proof. Below we will have $\overline{C}\geqslant 4C$, where C is the constant from Proposition 7.3. We will first apply Corollary 7.9 with $t=\frac{1}{3}$ and $d\asymp u$, while noting that if d is a large enough multiple of u and $\bar{c}$ is small enough (i.e., u is small enough), then the conditions (7.26) and (7.27) are satisfied. For ϕ as in (7.28) we have $\phi \asymp \alpha^4 u^8$, and for ζ as in (7.29) we have $\zeta \asymp \alpha^5 u^{15}$. Note that, since E is δ-monotone on $B_{\overline{C}}(x)$, it is also ζ-monotone on $B_{2C}(x)$, provided $\bar{c}$ is small enough. Hence, it follows from Corollary 7.9 that there exists $L \in \mathrm{lines}(\mathbb{H})$ with $L(0)=x$ and $L\left(-\frac{1}{3}\right) \in \partial_{\phi,d}(E)$.

Next, we will apply Corollary 7.8 with $x_1 = L\left(-\frac{1}{3}\right)$, $x_2 = x$, $\varkappa = \frac{1}{3}$, $\alpha_1 = \phi$, $u_1 = d$, $\alpha_2 = \alpha$ and $u_2 = u$. Provided $\bar{c}$ is small enough, u_1 and u_2 satisfy the conditions of Corollary 7.8. For γ_* as in (7.20) we have $\gamma_* \asymp u^{1/4}$, for β_* as in (7.21) we have $\beta_* \asymp \alpha^{19} u^{181/4}$, and for δ_* as in (7.22) we have $\delta_* \asymp \alpha^{29} u^{141/2}$. Moreover, since E is δ-monotone on $B_{\overline{C}}(x)$, it is also δ_*-monotone on $B_{3C}(x_1)$, provided $\bar{c}$ is small enough. Corollary 7.8 now implies that $L(s) \in \partial_{\beta_*,\gamma_*}(E)$ for all $s \in \left[-\frac{1}{3}, \frac{1}{3}\right]$, as required. □

8. Classification of monotone sets

It was shown in [21] that a proper non-empty monotone set $E \subset \mathbb{H}$ is, up to a set of measure zero, a half-space $\mathcal{P}$. In this section we recall the proof of this result, which is an essential preliminary for the proof of Theorem 4.3, the stability version for δ-monotone sets, given in §9. To give the key ideas, it suffices to consider the case of precisely monotone sets. We call a set E *precisely monotone* if $E \cap L$ and $E^c \cap L$ are connected for all L (rather than connected modulo subsets of measure zero).

There are two main steps:

Step A. For every $r>0$ and $p \in \mathbb{H}$, either $E \cap B_r(p)$ or $E' \cap B_r(p)$ has non-empty interior.

Step B. Either ∂E is contained in a 2-plane, in which case (by a trivial connectedness argument) the theorem holds, or ∂E has non-empty interior.

Step A follows from the special case of Lemma 7.2 in which $\varepsilon=0$; the general case of Lemma 7.2 is Step A′, the quantitative version of Step A, which is used in the proof of Theorem 4.3. Below, we prove Step B.

8.1. Proof of Step B

It will suffice to assume that ∂E is non-empty.

The proof of Step B has the following substeps:

Step B1. If $L \in \mathrm{lines}(\mathbb{H})$ and $L \cap \partial E$ contains at least two points, then $L \subset \partial E$.

Step B2. ∂E is a union of lines.

Step B3. Steps B1 and B2 imply Step B.

Steps B1 and B2 are special cases of Corollaries 7.8 and 7.10, respectively. But, for clarity, we will repeat the proofs in the present simpler (non-quantitative) situation. The proof of Step B3 requires some additional properties of pairs of lines, which are established in the next subsection.

8.2. Properties of pairs of lines

Two lines $L_1, L_2 \in \mathrm{lines}(\mathbb{H})$ are called *parallel* if $L_2 = g \cdot L_1$, or equivalently $L_1 = g^{-1} \cdot L_2$, for some $g \in \mathbb{H}$. This holds if and only if the projections $\pi(L_1)$ and $\pi(L_2)$ are either parallel or coincide. Unless $\pi(L_1) = \pi(L_2)$, the lines L_1 and L_2 are *not* parallel as lines in $\mathbb{R}^3$.

If L_1 and L_2 are not parallel, then their projections intersect in a unique point. If, in addition, $L_1 \cap L_2 = \varnothing$, then the pair (L_1, L_2) is called *skew*. We put

$$F(L_1, L_2) = \pi^{-1}(\pi(L_1) \cap \pi(L_2)). \tag{8.1}$$

Define the vertical distance between L_1 and L_2 by

$$d^V(L_1, L_2) = d^{\mathbb{H}}(L_1 \cap F(L_1, L_2), L_2 \cap F(L_1, L_2)). \tag{8.2}$$

The following lemmas describe the family of lines L which intersect both L_1 and L_2, where L_1 and L_2 are either parallel lines with distinct projections, or skew. At this point, the roles of L_1 and L_2 are symmetric and assertions about L_1 should be understood as applying to L_2 as well.

LEMMA 8.1. *If $L_1, L_2 \in \mathrm{lines}(\mathbb{H})$ are parallel and $\pi(L_1) \neq \pi(L_2)$, then the following properties hold*:

(1) *For $x \in L_1$ there is a unique $x^* \in L_2$ such that x and x^* lie on a line $L(x) \in \mathrm{lines}(\mathbb{H})$. Moreover, there is a unique shortest line segment $b(L_1, L_2)$ from L_1 to L_2 which is orthogonal to both L_1 and L_2.*

(2) *There exists a unique point $m = m(L_1, L_2) \in \mathbb{R}^2$ lying halfway between $\pi(L_1)$ and $\pi(L_2)$ such that, for all $x \in L_1$, the line $L(x)$ intersects the fiber $\pi^{-1}(m)$. Moreover, every point on $\pi^{-1}(m)$ intersects some line $L(x)$.*

(3) *The union of the lines $\{L(x)\}_{x \in L_1}$ is a smooth ruled surface $X = X(L_1, L_2)$ defined over $\mathbb{R}^2 \setminus T$, where $T = T(L_1, L_2)$ denotes the line parallel to $\pi(L_1)$ and $\pi(L_2)$ which passes through $m(L_1, L_2)$. At no point $x \in X$ is X tangent to the horizontal subspace H_x.*

Proof. Let O_θ be as in §2.3. By applying O_θ for suitable θ, and then applying a suitable left translation, we may assume that $L_1 = (0,0,0) \cdot (t,0,0)$. Next, after a suitable translation $(a'',0,0)$ (which leaves L_1 invariant), we may assume that L_2 intersects the plane $(u,v,0)$ along the v-axis, and hence can be written as $(0,b'',0) \cdot (s,0,0)$ (we are using here the fact that $\pi(L_1)\neq\pi(L_2)$). Finally, after a translation of the form $\left(0,-\frac{1}{2}b'',0\right)$, we may assume that

$$L_1=(t,e,-et) \quad \text{and} \quad L_2=(s,-e,es) \tag{8.3}$$

for some $e\in\mathbb{R}$.

For this pair, if $x=(t,e,-et)$, then

$$x^*=(-t,-e,-et). \tag{8.4}$$

To see this, observe that the line $L(x)$ in $\mathbb{R}^3$ which joins these points, is parallel to the plane $(u,v,0)$ and passes through the point $(0,0,-et)$. So $L(x)\in\text{lines}(\mathbb{H})$; see (2.13). The fiber $\pi^{-1}((0,0))$ is the unique fiber through which all of these lines pass as t varies in $(-\infty,\infty)$, and every point on $\pi^{-1}((0,0))$ arises in this way. Thus, for this pair,

$$m(L_1,L_2)=(0,0). \tag{8.5}$$

The unique shortest line segment joining L_1 and L_2 has length $2e$ and is contained in the line

$$b(L_1,L_2):=(0,s,0). \tag{8.6}$$

The surface X is given by

$$\left(u,v,-\frac{e^2u}{v}\right). \tag{8.7}$$

From these facts, all the remaining statements are straightforward to verify. □

Lemma 8.2. *If $L_1,L_2\in\text{lines}(\mathbb{H})$ are skew, then there is a hyperbola*

$$Y=Y(L_1,L_2)\subset\mathbb{R}^2$$

with asymptotes $\pi(L_1)$ and $\pi(L_2)$, such that every tangent line of Y has a unique lift $L\in\text{lines}(\mathbb{H})$ which intersects both L_1 and L_2. Conversely, if $L\in\text{lines}(\mathbb{H})$ and $L\cap L_i\neq\varnothing$, $i=1,2$, then $\pi(L)$ is tangent to Y. Except for the unique point $L_1\cap F(L_1,L_2)$ (resp. $L_2\cap F(L_1,L_2)$), every point x of L_1 lies on a unique line $\underline{L}(x)$ intersecting L_2 (resp. every point x of L_2 lies on a unique line $\underline{L}(x)$ intersecting L_1).

Proof. By applying a suitable left translation and then an isometry O_θ, and after possibly interchanging the subscripts of L_1 and L_2, we may assume the canonical normalization

$$L_1=(t,wt,c), \quad L_2=(s,-ws,-c), \quad w,c>0 \quad \text{and} \quad \pi(L_1)\cap\pi(L_2)=(0,0). \tag{8.8}$$

From the geometric interpretation of the correction term given after (2.11), it follows that the points $p=(t,wt,c)\in L_1$ and $q=(s,-ws,-c)\in L_2$ lie on some $L\in\text{lines}(\mathbb{H})$, precisely when there is parallelogram in $\mathbb{R}^2$ spanned by $(s,-ws)$ and (t,wt), with oriented area $2c$. Define $\theta(L_1,L_2)\in(0,\pi)$ by

$$\tan\tfrac{1}{2}\theta(L_1,L_2)=w, \tag{8.9}$$

or, equivalently,

$$ts\tan\tfrac{1}{2}\theta(L_1,L_2)=wts=c. \tag{8.10}$$

Then

$$(1+w^2)ts\sin\theta(L_1,L_2)=2c.$$

The envelope of the resulting family of lines is the hyperbola Y with equation

$$w^2u^2-v^2=cw. \qquad \square$$

COROLLARY 8.3. *Let L_1 and L_2 be skew lines as in the proof of Lemma* 8.2. *Put $x=(t,wt,c)$ and $\hat{x}=(-t,-wt,c)$ (note that $x,\hat{x}\in L_1$). Then, as elements of* $\mathrm{lines}(\mathbb{H})$, *the lines $\underline{L}(x)$ and $\underline{L}(\hat{x})$ are parallel, and*

$$m(\underline{L}(x),\underline{L}(\hat{x}))=(0,0).$$

There are precisely two *shortest* horizontal segments joining L_1 and L_2. In the notation of the proof of Lemma 8.2, these are the segments which project to the segments that touch $\pi(L_1)$ and $\pi(L_2)$, and are tangent to the hyperbola Y at its focal points $\pm\sqrt{c/w}$. Thus, these segments are contained in the lines

$$b_+(L_1,L_2):=\left(\sqrt{\frac{c}{w}},s,s\sqrt{\frac{c}{w}}\right)\quad\text{and}\quad b_-(L_1,L_2):=\left(-\sqrt{\frac{c}{w}},s,-s\sqrt{\frac{c}{w}}\right). \tag{8.11}$$

Remark 8.4. A pair of parallel lines L_1,L_2 with distinct projections, can be viewed as a limiting case of a pair of skew lines $L_{1,t}$, $L_{2,t}$ in which

$$\lim_{t\to\infty}\theta(L_{1,t},L_{2,t})=0$$

(where $\theta(L_{1,t},L_{2,t})$ is the angle between $\pi(L_{1,t})$ and $\pi(L_{2,t})$) and, as $t\to\infty$, one of $b_\pm(L_{1,t},L_{2,t})$ moves off to infinity, while the other converges to $b(L_1,L_2)$; cf. (8.6).

Proof of Step B1. Let $p\in\partial E$. Then the exists $p_i\in E$, with $p_i\to p$. Let $q\in\mathrm{int}(E)$, and assume that there is a line $L\in\mathrm{lines}(\mathbb{H})$ with $L(0)=p$ and $L(\ell)=q$. If L_i denotes the line parallel to L passing through p_i, then $L_i(\ell)\to L(\ell)$, and hence $L_i(\ell)\in\mathrm{int}(E)$ for i sufficiently large. Since E is convex, by (the non-quantitative version of) Corollary 7.8, we get $L_i(s)\in\mathrm{int}(E)$ for $0<s\leqslant\ell$ and i sufficiently large, and, by passing to the limit, $L(s)\in\mathrm{int}(E)$ for $0<s\leqslant\ell$. Since E^c is also convex, the corresponding statement holds for E^c as well. This completes the proof of Step B1. $\square$

Proof of Step B2. Let $p\in\partial E$ and $L\in\text{lines}(\mathbb{H})$ with $p=L(0)$ and $q=L(\ell)\neq p$. We need to show that $q\in\partial E$. By Step B1, we may assume that $L(\ell)\in\text{int}(E)$, say. Similarly, $L(-\ell)\in\text{int}(E)$ or $L(-\ell)\in\text{int}(E^c)$. By the same argument as in the proof of Step B1, if $L(\ell)\in\text{int}(E)$ and $L(-\ell)\in\text{int}(E)$, then $p=L(0)\in\text{int}(E)$. Therefore, we may assume that $L(-\ell)\in\text{int}(E^c)$. Since the collection of unit tangent vectors to the horizontal geodesics at p is a circle, and in particular connected, it follows by continuity that there exists a horizontal geodesic $\underline{L}$ with $\underline{L}(0)=p$ and $\underline{L}(\ell)\in\partial E$. The conclusion of Step B2 now follows from Step B1. □

Proof of Step B3. By Step B2, ∂E is a union of lines. If ∂E contains only parallel lines with the same projection, then it is contained in a vertical plane. Likewise, if ∂E contains only lines passing through some fixed point, then it is contained in the horizontal 2-plane through that point. Thus, we may assume that ∂E contains either a pair of parallel lines with distinct projections, or a pair of skew lines. By Lemma 8.1, if ∂E contains a pair of parallel lines L_1, L_2 with distinct projections, then distinct members of the family of lines $\{L(x)\}_{x\in L_1}$ are skew lines which intersect $\pi^{-1}(m(L_1, L_2))$. So we may assume that ∂E contains a pair of skew lines L_1, L_2.

Claim. If L_1 and L_2 are skew lines, with $L_1\cup L_2\subset\partial E$, then

$$\pi^{-1}(\pi(L_1)\cap\pi(L_2))\subset\partial E.$$

To see this, note that, by Corollary 8.3, for every line L_3 intersecting L_1 and L_2, there is a parallel line L_4 intersecting L_1 and L_2, such that $m(L_3, L_4)=\pi(L_1)\cap\pi(L_2)$. Since, by Step B1, $L_3\cup L_4\subset\partial E$, it follows from Lemma 8.1 that $\pi^{-1}(m(L_3, L_4))\in\partial E$.

Let $Y=Y(L_1, L_2)$ be the hyperbola as in Lemma 8.2 and let U be the component of $\mathbb{R}^2\setminus Y$ whose closure intersects both branches of Y. Every point of $V=U\setminus(\pi(L_1)\cup\pi(L_2))$ is the intersection of a pair of distinct tangent lines of Y. Each such pair of lines lifts to a pair of skew lines L_5, L_6 in ∂E. By the claim, $\pi^{-1}(\pi(L_5)\cap\pi(L_6))\in\partial E$, so $\pi^{-1}(V)\subset\partial E$. Thus ∂E has non-empty interior, as required. □

9. Proof of the stability of monotone sets

The proof of Theorem 4.3, that is, the stability of monotone sets, has two main steps, Steps A′ and B′, which are quantitative versions of Steps A and B in the classification of monotone subsets of $\mathbb{H}$ given in §8. In fact, Step A′ is just Lemma 7.2, which is proved in §11. In this section, assuming some technical geometric preliminaries which are proved in §10, we carry out Step B′, thereby completing the proof of Theorem 4.3.

The proof of Step B′ has three substeps, Steps B1′–B3′, which correspond to Steps B1–B3. Steps B1′ and B2′ are just Corollaries 7.8 and 7.10, specialized to our specific context. Step B3′ requires a substantial preliminary discussion; see §10 and, in particular, Lemma 10.1 and Corollaries 10.2 and 10.4. The point of the discussion in §10 is to quantify the following statement which occurs at the end of the first paragraph of the proof of Step B3: "Thus, we may assume that ∂E contains either a pair of parallel lines with distinct projections, or a pair of skew lines". Once this statement has been quantified, the proof of Step B3′ is completed by repeating, mutadis mutandis, the proof of Step B3.

The preliminaries to Step B3′, i.e., §10, constitute the part of the argument which necessitates assuming that E is ε^a-monotone on $B_{\varepsilon^{-3}}(x)$, rather than on $B_1(x)$. The following example shows that such an assumption cannot be entirely avoided, even if ε^a-monotonicity is strengthened to monotonicity.

Example 9.1. Consider the set

$$E=\{(x,y,z): z\leqslant xy \text{ and } y>0\}. \tag{9.1}$$

We claim that if $B_r(p)\subset\{(x,y,z):y>0\}$, then $E\cap B_r(p)$ is monotone. This is equivalent to the statement that if a line $L\in\text{lines}(\mathbb{H})$ passes through two points in

$$\partial E=\{(x,y,z): z=xy \text{ and } y>0\},$$

then L is contained in ∂E. Note that the horizontal line $L=(a,t,at)$ passes through (a,b,ab). If $b>0$, then L is contained in ∂E. Conversely, if $(a,b,ab),(a',b',a'b')\in\partial E$, then $b,b'>0$. If these points lie on a horizontal line, since $H_{(a,b,ab)}=(x,y,ab-bx+ay)$, we get

$$a'b'=ab-ba'+ab'; \tag{9.2}$$

see (2.13). The solutions of (9.2) are either $a'=a$ and b' arbitrary, which gives the above line L, or a' arbitrary and $b'=-b$, which contradicts the assumption $b'>0$.

Proof of Theorem 4.3. Recall that in Theorem 4.3, we are given $E\subset B_1(x)$ which is ε^a-monotone on $B_1(x)$. Our goal is to find a half-space $\mathcal{P}\subseteq\mathbb{H}$ such that

$$\frac{\mathcal{L}_3((E\cap B_{\varepsilon^3}(x))\triangle\mathcal{P})}{\mathcal{L}_3(B_{\varepsilon^3}(x))}\lesssim\varepsilon.$$

For convenience of notation in the ensuing argument, we will rescale the ball $B_1(x)$ above by ε^{-3}. Thus our assumption is that $E\subset B_{\varepsilon^{-3}}(x)$ is ε^a-monotone on $B_{\varepsilon^{-3}}(x)$ and, for the sake of contradiction, that for no half-space $\mathcal{P}$ we have

$$\mathcal{L}_3((E\cap B_1(x))\triangle\mathcal{P})\lesssim\varepsilon.$$

This contrapositive assumption will be used via the following simple lemma. In what follows, for $A\subset\mathbb{H}$ and $\varepsilon>0$, the closed ε-tubular neighborhood of A is denoted by

$$\overline{T_\varepsilon(A)}=\{x\in\mathbb{H}:d^{\mathbb{H}}(x,A)\leqslant\varepsilon\}. \tag{9.3}$$

LEMMA 9.2. *Let $Q\subset\mathbb{H}$ be a 2-plane. Assume that, for some $\varepsilon,r\in\left(0,\frac{1}{2}\right)$, a subset $E\subset\mathbb{H}$ satisfies*

$$\partial_{\varepsilon,r}(E)\cap B_1(x)\subset\overline{T_\varepsilon(Q)}.$$

Then there exists a half-space $\mathcal{P}\subset\mathbb{H}$ such that

$$\mathcal{L}_3((E\cap B_1(x))\triangle\mathcal{P})\lesssim\varepsilon.$$

Proof. The set $B_1(x)\setminus\overline{T_\varepsilon(Q)}$ has at most two connected components, say C_1 and C_2 (one of which might be empty). Since $\varepsilon<\frac{1}{2}$, by continuity, for $j\in\{1,2\}$ we have either $C_j\subset\operatorname{int}_{\varepsilon,r}(E)$ or $C_j\subset\operatorname{int}_{\varepsilon,r}(E^c)$. For definiteness assume that $C_1\subset\operatorname{int}_{\varepsilon,r}(E)$. Let $\mathfrak{N}\subset C_1$ be an r-net in C_1. Thus the balls $\{B_{r/2}(y)\}_{y\in\mathfrak{N}}$ are disjoint, and the balls $\{B_r(y)\}_{y\in\mathfrak{N}}$ cover C_1. Moreover, since for each $y\in\mathfrak{N}$ we have $y\in\operatorname{int}_{\varepsilon,r}(E)$, we know that

$$\mathcal{L}_3(B_r(y)\cap E^c)\leqslant\varepsilon\mathcal{L}_3(B_r(y)).$$

Hence

$$\begin{aligned}\mathcal{L}_3(C_1\cap E^c)\leqslant\sum_{y\in\mathcal{N}}\mathcal{L}_3(B_r(y)\cap E^c)\leqslant\sum_{y\in\mathcal{N}}\varepsilon\mathcal{L}_3(B_r(y))=16\varepsilon\sum_{y\in\mathfrak{N}}\mathcal{L}_3(B_{r/2}(y))\\ =16\varepsilon\mathcal{L}_3\bigg(\bigcup_{y\in\mathfrak{N}}B_{r/2}(y)\bigg)\leqslant16\varepsilon\mathcal{L}_3(B_{1+r/2}(x))\lesssim\varepsilon.\end{aligned}$$

This argument shows that for $j\in\{1,2\}$ either $\mathcal{L}_3(C_j\cap E^c)\lesssim\varepsilon$ or $\mathcal{L}_3(C_j\cap E)\lesssim\varepsilon$. Thus we can take $\mathcal{P}$ to be either one of the half-spaces bounded by Q, or a half space that contains $B_1(x)$. □

By virtue of Lemma 9.2, we may assume from now on that for every 2-plane $P\subset\mathbb{H}$ we have

$$\partial_{c_1\varepsilon,\varepsilon^{k_1}}(E)\cap B_1(x)\not\subset\overline{T_{c_1\varepsilon}(P)}, \tag{9.4}$$

where $k_1>1$ and $c_1>0$ are constants which will be determined presently.

By rescaling the metric $d^{\mathbb{H}}$ by a suitable multiple of ε^{-3}, we may apply Corollary 7.10 with $\alpha\asymp\varepsilon$ and $u\asymp\varepsilon^{k_1+3}$, and (after rescaling back to our present setting) deduce that, provided

$$\begin{aligned}a&>29+71(k_1+3)=242+71k_1,\\ k_2&<\tfrac{1}{4}(k_1+3),\\ h_2&>19+46(k_1+3)=157+46k_1,\end{aligned} \tag{9.5}$$

and ε is smaller than a small enough universal constant, we may associate with every point $y\in\partial_{\varepsilon,\varepsilon^{k_1}}(E)\cap B_1(x)$ an open set $\mathcal{O}_y$ of lines through the origin of H_y, such that for all $L\in\mathcal{O}_y$ we have

$$L([-c_2\varepsilon^{-3}, c_2\varepsilon^{-3}]) \subset \partial_{\varepsilon^{h_2},\varepsilon^{k_2}}(E) \tag{9.6}$$

for some small enough constant $c_2>0$. We have used here the fact that the definition of the quantitative boundary in (7.2) is an open condition.

As we shall see in §10 (see Lemma 10.1 and Corollaries 10.2, 10.3 and 10.4), the above discussion implies the following: if c_1 in assumption (9.4) is a small enough universal constant, then there exists a pair of skew lines $L, L'\in\operatorname{lines}(\mathbb{H})$, with the following properties:

- the distances of L and L' from x are at most $\frac{1}{10}c_2\varepsilon^{-3}$;
- both L and L' intersect $\partial_{\varepsilon^{h_2},\varepsilon^{k_2}}(E)\cap B_{c_2\varepsilon^{-3}/10}(x)$;
- the angle $\theta(L, L')$ between L and L' (for the definition, see (8.9)) is bounded away from 0 and π, say, $\tan\theta(L, L')\in\left[\frac{2}{5}, \frac{5}{2}\right]$;
- L and L' have separation $\gtrsim\varepsilon^3$, i.e. $d^V(L, L')\gtrsim\varepsilon^3$, where the vertical distance $d^V(\cdot\,,\cdot)$ is as in (8.2).

We now use the lines L and L' to perform Step B3′, that is, to produce a quantitative version of the argument in Step B3.

First of all, we will apply again Corollary 7.10 (rescaled, as before, by $\asymp\varepsilon^{-3}$) to deduce, from the fact that L and L' intersect $\partial_{\varepsilon^{h_2},\varepsilon^{k_2}}(E)\cap B_{c_2\varepsilon^{-3}/10}(x)$, that

$$(L\cup L')\cap B_{c_3\varepsilon^{-3}}(x) \subset \partial_{\varepsilon^{h_3},\varepsilon^{k_3}}(E), \tag{9.7}$$

where $c_3>0$ is a universal constant, provided that

$$\begin{aligned} a &> 29h_2+71(k_2+3) = 29h_2+71k_2+213, \\ k_3 &< \tfrac{1}{4}(k_2+3), \\ h_3 &> 19h_2+46(k_2+3) = 19h_2+46k_2+138. \end{aligned} \tag{9.8}$$

Now, as in the non-quantitative proof, let $Y=Y(L, L')\subset\mathbb{R}^2$ be the hyperbola from Lemma 8.2, and let U be the component of $\mathbb{R}^2\setminus Y$ whose closure intersects both branches of Y. Let c_4 be a small enough constant, and take a point $q\in U$ such that $B_{c_4\varepsilon^3}(q)\cap\mathbb{R}^2\subset U$ and $d^{\mathbb{H}}(q, \pi(L)\cap\pi(L'))\leqslant 10\varepsilon^3$. As in the proof of Step B3, using Lemma 8.2, any point $z\in B_{c_4\varepsilon^3}(q)\cap\mathbb{R}^2$ is of the form $z=\pi(L^*)\cap\pi(L^{**})$, where, provided c_4 is small enough, L^* and L^{**} are skew lines which intersect $L\cap B_{c_3\varepsilon^{-3}}(x)$ and $L'\cap B_{c_3\varepsilon^{-3}}(x)$, both of which are contained in $\partial_{\varepsilon^{h_3},\varepsilon^{k_3}}(E)$. Another application of Corollary 7.10 implies that

$$(L^*\cup L^{**})\cap B_{c_5\varepsilon^{-3}}(x) \subset \partial_{\varepsilon^{h_4},\varepsilon^{k_4}}(E), \tag{9.9}$$

provided that

$$\begin{aligned} a &> 29h_3+71(k_3+3)=29h_3+71k_3+213,\\ k_4 &< \tfrac{1}{4}(k_3+3),\\ h_4 &> 19h_3+46(k_3+3)=19h_3+46k_3+138. \end{aligned} \tag{9.10}$$

We continue to argue as in Step B3. We use Corollary 8.3 to deduce that $z=m(\underline{L}^*,\underline{L}^{**})$, where (once more, provided c_4 is small enough), $\underline{L}^*$ and $\underline{L}^{**}$ are parallel lines which intersect the line segments $L^*\cap B_{c_5\varepsilon^{-3}}(x)$ and $L^{**}\cap B_{c_5\varepsilon^{-3}}(x)$, both of which are contained in $\partial_{\varepsilon^{h_4},\varepsilon^{k_4}}(E)$. So, another application of Corollary 7.10 gives

$$(\underline{L}^*\cup\underline{L}^{**})\cap B_{c_6\varepsilon^{-3}}(x)\subset\partial_{\varepsilon^{h_5},\varepsilon^{k_5}}(E), \tag{9.11}$$

provided that

$$\begin{aligned} a &> 29h_4+71(k_4+3)=29h_4+71k_4+213,\\ k_5 &< \tfrac{1}{4}(k_4+3),\\ h_5 &> 19h_4+46(k_4+3)=19h_4+46k_4+138. \end{aligned} \tag{9.12}$$

By Lemma 8.1, any point on the fiber $\pi^{-1}(z)$ lies on a line which touches both of $\underline{L}^*$ and $\underline{L}^{**}$. Moreover (using the explicit formula for this line, which is contained in the proof of Lemma 8.1), since the vertical separation between L and L' is $\gtrsim\varepsilon^3$, and z is an arbitrary point in $B_{c_4\varepsilon^3}(q)\cap\mathbb{R}^2$, a (final) application of Corollary 7.10, together with the box-ball principle, shows that there exists $p\in\mathbb{H}$ such that

$$B_r(p)\subset\partial_{\varepsilon^{h_6},\varepsilon^{k_6}}(E)\cap B_{\varepsilon^{-3}}(x), \tag{9.13}$$

where $r\asymp\varepsilon^3$, provided that

$$\begin{aligned} a &> 29h_5+71(k_5+3)=29h_5+71k_5+213,\\ k_6 &< \tfrac{1}{4}(k_5+3),\\ h_6 &> 19h_5+46(k_5+3)=19h_5+46k_5+138. \end{aligned} \tag{9.14}$$

Now, by choosing k_1 sufficiently large, we will use (9.13) to contradict Lemma 7.2, provided k_1 is large enough. Modulo the proofs of technical lemmas which were postponed to the following sections, this will conclude the proof of Theorem 4.3, Before doing so, note that for the conditions (9.5), (9.8), (9.10), (9.12) and (9.14) to be satisfied, we can ensure that (9.13) is satisfied for, say, $a=2^{40}k_1$, $k_6=2^{-10}k_1$ and $h_6=2^{30}k_1$. (In actuality, these are big overestimates.)

It remains to show how to choose k_1 so that (9.13) will contradict Lemma 7.2. Since E is ε^a-monotone on $B_{\varepsilon^{-3}}(x)$, it is also $\asymp\varepsilon^{a-24}$-monotone on $B_r(p)$ (recall that $r\asymp\varepsilon^3$). By Lemma 7.2, we can find $p'\in B_{r/2}(p)$ such that $p'\notin\partial_{\alpha,r'}(E)$, where $\alpha\asymp\varepsilon^{(a-24)/2}$ and $r'\asymp r\asymp\varepsilon^3$. Without loss of generality, we may assume that $p'\in\mathrm{int}_{\alpha,r'}(E)$, so that

$$\mathcal{L}_3(E^c\cap B_{r'}(p'))\leqslant\alpha\mathcal{L}_3(B_{r'}(p'))\asymp\varepsilon^{a/2-12}\varepsilon^{12}=\varepsilon^{a/2}. \tag{9.15}$$

But, at the same time, $p'\in\partial_{\varepsilon^{h_6},\varepsilon^{k_6}}(E)$, which means that

$$\mathcal{L}_3(E^c\cap B_{\varepsilon^{k_6}}(p'))>\varepsilon^{h_6}\mathcal{L}_3(B_{\varepsilon^{k_6}}(p'))\asymp\varepsilon^{h_6+4k_6}. \tag{9.16}$$

In order for (9.16) to contradict (9.15), assume that $\varepsilon^{k_6}>r'$, which would hold for small enough ε if $k_6=k_1/2^{10}>3$. In this case, $B_{\varepsilon^{k_6}}(p')\subset B_{r'}(p')$, and the desired contradiction would follow (for small enough ε) if $\frac{1}{2}a>h_6+4k_6$. Choosing $k_1=2^{12}$, and thus $a=2^{52}$ and $h_6=2^{42}$, yields the required contradiction, and completes the proof of Theorem 4.3. □

10. Non-degeneracy of the initial configuration.

In this section, we prove the assertion on non-degeneracy of the initial configuration, which was used in the proof of Step B3′ given in §9; see the four items marked with bullets in the paragraph following (9.6). Specifically under the assumptions of Lemma 10.1 below, we show that at distance $\lesssim\varepsilon^{-3}$ from x, we can find a pair of skew lines in the controlled quantitative boundary which make standard angle as in (10.48), and have separation $\gtrsim\varepsilon^3$; see (10.4) and (10.49).

Given a pair of (unordered) skew lines $L,\tilde{L}$, define $0<\theta(L,\tilde{L})<\pi$ to be the angle between $\pi(L)$ and $\pi(\tilde{L})$ which faces the sector which contains the hyperbola $Y(L,\tilde{L})$ from Lemma 8.2. The vertical distance $d^V(L,\tilde{L})$ is defined as in (8.2). As in (8.11), denote by $b(L,\tilde{L})$ the union of the (two) shortest horizontal segments joining L and $\tilde{L}$, and denote the length of each of these segments by $\mathfrak{d}(L,\tilde{L})$. Then, with the geometric interpretation of the multiplication in $\mathbb{H}$, we get

$$\tfrac{1}{2}\big(\cot\tfrac{1}{2}\theta(L,\tilde{L})\big)\mathfrak{d}(L,\tilde{L})^2=d^V(L,\tilde{L})^2. \tag{10.1}$$

Since $\tan\frac{1}{2}\theta\geqslant\frac{1}{2}\sin\theta$ for all $\theta\in[0,\pi]$, it follows from (10.1) that

$$\mathfrak{d}(L,\tilde{L})^2\gtrsim(\sin\theta(L,\tilde{L}))d^V(L,\tilde{L})^2. \tag{10.2}$$

In what follows, for $A\subset\mathbb{H}$ and $\varepsilon>0$, the closed ε-tubular neighborhood of A is

$$\overline{T_\varepsilon(A)}=\{x\in\mathbb{H}:d^{\mathbb{H}}(x,A)\leqslant\varepsilon\}.$$

The required non-degeneracy of the initial configuration, namely the existence of the "quantitatively skew" lines L and L' which are described in the paragraph preceding (9.7), is a consequence of the following lemma. The hypothesis (10.3) below corresponds to the assumption concerning 2-planes which we arrived at in (9.4), and the discussion preceding (9.6).

LEMMA 10.1. *Fix $x\in\mathbb{H}$, $U\subset B_1(x)$ and $\varepsilon\in(0,1)$. For all $y\in U$ fix an open set $\mathcal{O}_y$ of lines through the origin of H_y. Assume that for all 2-planes P we have*

$$U\not\subset\overline{T_\varepsilon(P)}. \tag{10.3}$$

Then there exist $y,z\in U$, $L_y\in\mathcal{O}_y$ and $L_z\in\mathcal{O}_z$ such that L_y and L_z are skew and

$$\mathfrak{d}(L_y,L_z)^2\gtrsim(\sin\theta(L_y,L_z))d^V(L_y,L_z)^2\gtrsim\varepsilon^6. \tag{10.4}$$

Proof. Denote the Euclidean distance in $\mathbb{R}^2$ by $d(\cdot,\cdot)$. Fix $y_1\in U$ and let V denote the unique vertical 2-plane containing L_{y_1}. It follows from (10.3), with $P=V$, that there exists $y_2\in U$ such that

$$d(\pi(L_{y_1}),\pi(y_2))\geqslant\varepsilon. \tag{10.5}$$

Let $H=H_q$ denote the unique horizontal 2-plane containing the line L_{y_1} and the point $y_2\notin L_{y_1}$. Note that $q\in L_{y_1}$. By (10.3), there exists y_3 such that

$$d^{\mathbb{H}}(y_3,H_q)\geqslant\varepsilon. \tag{10.6}$$

For arbitrary $q,y_3\in\mathbb{H}$, we have

$$d^{\mathbb{H}}(q,H_{y_3})=d^{\mathbb{H}}(y_3,H_q). \tag{10.7}$$

In fact, by the left-invariance of $d^{\mathbb{H}}$, the left-hand side of (10.7) equals $d^{\mathbb{H}}(y_3^{-1}\cdot q,H_{(0,0,0)})$, whereas its right-hand side equals $d^{\mathbb{H}}(q^{-1}\cdot y_3,H_{(0,0,0)})$. If $z\in H_{(0,0,0)}$ is closest to $y_3^{-1}\cdot q$, then $z^{-1}\in H_{(0,0,0)}$ and, by (2.15), $d^{\mathbb{H}}(y_3^{-1}\cdot q,z)=d^{\mathbb{H}}(q^{-1}\cdot y_3,z^{-1})$. By symmetry, this gives (10.6).

Since $d^{\mathbb{H}}(q,L_{y_3})\geqslant d^{\mathbb{H}}(q,H_{y_3})$, from (10.6) and (10.7) we get

$$d^{\mathbb{H}}(q,L_{y_3})\geqslant\varepsilon. \tag{10.8}$$

Since $\mathcal{O}_{y_1}$, $\mathcal{O}_{y_2}$ and $\mathcal{O}_{y_3}$ are open, we may assume that no two of L_{y_1}, L_{y_2} and L_{y_3} are parallel. By applying a translation, we may also assume that $q=(0,0,0)$. Thus, after applying a suitable rotation, we may assume that L_{y_1} is the x-axis and that $y_1=(a,0,0)$

and $y_2=(b,c,0)$ for some $a,b,c\in\mathbb{R}$. Set $L_{y_1}\cap\pi(L_{y_2})=n_3=(s,0,0)$ and let $\alpha\in(0,\pi)$ denote the angle between L_{y_1} and $\pi(L_{y_2})$.

It follows from (10.5) that $|c|\geqslant\varepsilon$. As $|a-b|+|c|+\sqrt{|ac|}\asymp d^{\mathbb{H}}(y_1,y_2)\leqslant 2$, we have that

$$|a|\lesssim\frac{1}{\varepsilon}\quad\text{and}\quad|a-b|\lesssim 1. \tag{10.9}$$

Note that

$$\sin\alpha=\frac{c}{d(\pi(y_2),n_3)}\asymp\frac{c}{c+|b-s|}\geqslant\frac{\varepsilon}{\varepsilon+|b|+|s|} \tag{10.10}$$

and, by the geometric interpretation of the multiplication,

$$d^V(L_{y_1},L_{y_2})^2=|c|\,|s|\geqslant\varepsilon|s|. \tag{10.11}$$

Thus, by multiplying together (10.10) and (10.11), we may assume that $\varepsilon+|b|\geqslant|s|$, since otherwise (10.4) holds with ε^6 replaced by ε^2. Using (10.9), we deduce that $|b|,|s|\lesssim 1/\varepsilon$, and therefore it follows from (10.10) that

$$\sin\alpha\gtrsim\varepsilon^2. \tag{10.12}$$

By (10.11) and (10.12), we may assume that

$$d(n_3,q)=|s|\lesssim\varepsilon^3, \tag{10.13}$$

since otherwise (10.4) holds.

Let S be the triangle with vertices $\pi(L_{y_2})\cap\pi(L_{y_3})=n_1$, $L_{y_1}\cap\pi(L_{y_3})=n_2$ and n_3. Let α_j denote the angle at the vertex n_j (thus α_3 is either α or $\pi-\alpha$). Put $\ell_j=d(n_{j+1},n_{j+2})$. (Here and in the rest of the proof of Lemma 10.1, indices are taken mod 3.) It follows from (10.12) that there exists $j\in\{1,2\}$ such that

$$\sin\alpha_j\geqslant\tfrac{1}{2}c\varepsilon^2. \tag{10.14}$$

Write $\{k\}=\{1,2\}\setminus\{j\}$. From (10.12) and (10.14), we may assume that

$$d^V(L_{y_1},L_{y_2})^2\lesssim\varepsilon^4, \tag{10.15}$$

$$d^V(L_{y_k},L_{y_3})^2\lesssim\varepsilon^4, \tag{10.16}$$

since otherwise (10.4) holds.

From the geometric interpretation of the multiplication, it follows that there exists $r\in\{1,2,3\}$ such that

$$d^V(L_{r+1},L_{r+2})^2\gtrsim\operatorname{area}(S)=\tfrac{1}{2}\ell_1\ell_2\sin\alpha_3. \tag{10.17}$$

By multiplying both sides of (10.17) by $\sin\alpha_r$ and using the law of sines to express $\sin\alpha_r$ in terms of $\sin\alpha_3$, we obtain, with (10.12) and the triangle inequality,

$$d^V(L_{r+1},L_{r+2})^2\sin\alpha_r \gtrsim (\sin^2\alpha_3)\min\{\ell_1^2,\ell_2^2\} \gtrsim \varepsilon^4\min\{\ell_1^2,\ell_2^2\}. \tag{10.18}$$

Arguing similarly with (10.14), we have

$$d^V(L_{r+1},L_{r+2})^2\sin\alpha_r \gtrsim \varepsilon^4\min\{\ell_k^2,\ell_3^2\}. \tag{10.19}$$

We may therefore assume that $\min\{\ell_1^2,\ell_2^2\}\leqslant\frac{1}{16}\varepsilon^2$ and $\min\{\ell_k^2,\ell_3^2\}\leqslant\frac{1}{16}\varepsilon^2$, since otherwise (10.4) holds. By the triangle inequality, $\ell_k\leqslant\ell_j+\ell_3$, and therefore

$$\ell_k\leqslant\tfrac{1}{2}\varepsilon. \tag{10.20}$$

By adding (10.13), (10.15), (10.16) and (10.20), we get $d^{\mathbb{H}}(L_{y_3},q)\leqslant\frac{1}{2}\varepsilon+O(\varepsilon^2)$, which contradicts (10.8) for ε small enough. This contradiction completes the proof. □

In deriving the consequences of Lemma 10.1, we will work in coordinates. Thus, we write L_1 and L_2 for L_y and L_z, respectively. As in (8.8), we may canonically assume that

$$L_1=(t,wt,c) \quad\text{and}\quad L_2=(s,-ws,-c).$$

As in (8.9), the angle $0<\theta(L_1,L_2)<\pi$ is determined by $\tan\frac{1}{2}\theta(L_1,L_2)=w$. Let $b_+(L,\tilde{L})$ and $b_-(L,\tilde{L})$ contain the shortest line joining L_1 and L_2; see (8.11). Thus,

$$b_\pm(L_1,L_2)=\left(\pm\sqrt{\frac{c}{w}},u,\pm\sqrt{\frac{c}{w}}u\right),$$

and so

$$d^{\mathbb{H}}(b_+(L_1,L_2),b_-(L_1,L_2))=2\sqrt{\frac{c}{w}}. \tag{10.21}$$

Put

$$e_{i,\pm}(L_1,L_2)=L_i\cap b_\pm(L_1,L_2), \tag{10.22}$$

or, equivalently,

$$\begin{aligned} e_{1,+}(L_1,L_2)&=\left(\sqrt{\frac{c}{w}},\sqrt{cw},c\right),\\ e_{2,+}(L_1,L_2)&=\left(\sqrt{\frac{c}{w}},-\sqrt{cw},-c\right). \end{aligned} \tag{10.23}$$

Then, with the geometric interpretation of the multiplication in $\mathbb{H}$, $d^V(L_1,L_2)^2$ is twice the area of the triangle with vertices $\pi(L_1)\cap\pi(L_2)$, $\pi(e_{1,+}(L_1,L_2))$ and $\pi(e_{2,+}(L_1,L_2))$. It follows that $d^V(L_1,L_2)^2=2c$. Note that

$$\mathfrak{d}(L_1,L_2)=d^{\mathbb{H}}(e_{1,\pm}(L_1,L_2),e_{2,\pm}(L_1,L_2)).$$

Therefore,

$$\mathfrak{d}(L_1, L_2) = 2\sqrt{cw}. \tag{10.24}$$

Equation (10.1) now becomes

$$\mathfrak{d}(L_1, L_2)^2 = 2wd^V(L_1, L_2)^2. \tag{10.25}$$

Corollary 10.2 below will be derived directly from Lemma 10.1 without further reference to the assumption on 2-planes. The proof will rely on the following additional information. Write $o_1 = y$ and $o_2 = z$. Since $o_1, o_2 \in B_1(x)$, we have $d^{\mathbb{H}}(o_1, o_2) < 2$, which implies that $d^V(o_1, o_2) < 2$ and $d(\pi(o_1), \pi(o_2)) < 2$.

We put

$$o_1 = (t_1, wt_1, c) \quad \text{and} \quad o_2 = (s_2, -ws_2, -c),$$

where, without loss of generality, we may assume that either (i) $t_1 \geqslant 0$ and $s_2 \leqslant 0$, or (ii) $t_1 \geqslant 0$ and $s_2 \geqslant 0$.

Below, to avoid confusion, we will write Ω for the constant c, in (10.4). We also assume that $\varepsilon \leqslant 1$.

Corollary 10.2. *We have the bound*

$$c\frac{w}{1+w^2} \gtrsim \varepsilon^6. \tag{10.26}$$

Moreover,

(i) *If* $t_1 \geqslant 0$ *and* $s_2 \leqslant 0$, *then*

$$\varepsilon^6 \lesssim w \lesssim \varepsilon^{-6}, \tag{10.27}$$

$$\varepsilon^3 \sqrt{1+w^2} \lesssim \mathfrak{d}(L_1, L_2) \lesssim \sqrt{w}, \tag{10.28}$$

$$d^{\mathbb{H}}(o_1, e_{1,+}(L_1, L_2)) \lesssim \sqrt{w} + \frac{1}{\sqrt{w}}, \tag{10.29}$$

$$\varepsilon^6 \lesssim c \lesssim 1, \tag{10.30}$$

$$|t_1| + |s_2| \lesssim 1, \tag{10.31}$$

$$|t_1 + s_2| \lesssim \frac{1}{w}, \tag{10.32}$$

$$w|t_1|\,|s_2| \lesssim 1. \tag{10.33}$$

(ii) *If* $t_1 \geqslant 0$ *and* $s_2 \geqslant 0$, *then*

$$w \lesssim \varepsilon^{-6}, \tag{10.34}$$

$$\varepsilon^3 \sqrt{1+w^2} \lesssim \mathfrak{d}(L_1, L_2) \lesssim \sqrt{1+w}, \tag{10.35}$$

$$d^{\mathbb{H}}(o_1, e_{1,+}(L_1, L_2)) \lesssim \varepsilon^{-3}, \tag{10.36}$$

$$\varepsilon^6 \lesssim c \lesssim 1+\frac{1}{w}, \tag{10.37}$$

$$|t_1-s_2| \lesssim 1, \tag{10.38}$$

$$t_1+s_2 \lesssim \frac{1}{w}, \tag{10.39}$$

$$|c-wt_1s_2| \lesssim 1. \tag{10.40}$$

Proof. As $\sin\theta(L_1, L_2)=2w/(1+w^2)$ and $d^V(L_1, L_2)^2=2c$, (10.26) is just a rewriting of (10.4). The lower bounds in (10.28) and (10.35) follow from (10.26) and (10.24).

Next note that, since $d^{\mathbb{H}}(o_1, o_2)\leqslant 2$, from the box-ball principle we have

$$|t_1-s_2|+w|t_1+s_2|+\sqrt{|2c-2wt_1s_2|} \lesssim 1. \tag{10.41}$$

Case (i). Since the inequalities $t_1\geqslant 0$ and $s_2\leqslant 0$ imply that in the third term of the left-hand side of (10.41) we have $-wt_1s_2>0$, we get $c\lesssim 1$, which gives the upper bound in (10.30) and, since $\mathfrak{d}(L_1, L_2)=2\sqrt{cw}$, the upper bound in (10.27) as well. Using $c\lesssim 1$ and multiplying both sides of (10.26) by $(1+w^2)/w$ gives (10.27). Relations (10.31)–(10.33) follow immediately from (10.41), $t_1\geqslant 0$ and $s_2\leqslant 0$.

To prove (10.29), by considering the cases $w\leqslant 1$ and $w\geqslant 1$, and using the box-ball principle (10.41), it suffices to show that

$$t_1+\sqrt{\frac{c}{w}} \lesssim \frac{1}{\sqrt{w}}.$$

Since $c\lesssim 1$, we get $\sqrt{c/w}\lesssim 1/\sqrt{w}$. From (10.31), we get $t_1\lesssim 1$, which gives the case $w\leqslant 1$. For $w\geqslant 1$, relations (10.32) and (10.33) imply that $t_1\lesssim 1/\sqrt{w}$, which completes the proof.

Case (ii). Relation (10.37) was already shown in the proof of case (i). By using $t_1\geqslant 0$ and $s_2\geqslant 0$, relations (10.38)–(10.40) follow directly from (10.41). From (10.39),

$$c \lesssim 1+wt_1s_2 \lesssim 1+w(t_1+s_2)^2,$$

which, by (10.39), gives the upper bound for c in (10.37). From this, the upper bound in (10.35) follows immediately, as does the implication $c\leqslant 5$ if $w\geqslant 1$. This, together with (10.26), gives the upper bound for w in (10.34).

To prove (10.36), note that, from (10.40), by dividing through by w and factoring, we get

$$\left|\sqrt{\frac{c}{w}}-\sqrt{t_1s_2}\right|\left|\sqrt{\frac{c}{w}}+\sqrt{t_1s_2}\right| \lesssim \frac{1}{w},$$

which, together with (10.39), gives

$$(1+w)\left|\sqrt{\frac{c}{w}}-t_1\right|\lesssim(1+w)\frac{1}{\sqrt{wc}}+(1+w)|\sqrt{t_1s_2}-t_1|.$$

From (10.38), in case $w\leqslant 1$, and (10.39), in case $w\geqslant 1$, it follows that the second term of the right-hand side is $\lesssim 1$. By using the lower bound in (10.28) to bound the first term of the right-hand side, the proof is complete. □

In the next corollary, by considering separately the cases $0<w\leqslant 1$ and $1\leqslant w<\infty$, we obtain estimates which depend only on ε.

Corollary 10.3. (1) *If* $0<w\leqslant 1$, *then*

$$\varepsilon^3\lesssim\mathfrak{d}(L_1,L_2)\lesssim 1, \tag{10.42}$$

$$d^{\mathbb{H}}(o_1,e_{1,+}(L_1,L_2))\lesssim\varepsilon^{-3}. \tag{10.43}$$

(2) *If* $1\leqslant w<\infty$, *then*

$$\varepsilon^3\lesssim d^{\mathbb{H}}(b_+(L_1,L_2),b_-(L_1,L_2))\lesssim 1, \tag{10.44}$$

$$\varepsilon^3\lesssim\mathfrak{d}(L_1,L_2)\lesssim\varepsilon^{-3}, \tag{10.45}$$

$$d^{\mathbb{H}}(o_1,e_{1,+}(L_1,L_2))\lesssim\varepsilon^{-3}. \tag{10.46}$$

Proof. Relation (10.44), i.e., the bound on $2\sqrt{c/w}$ for $w\geqslant 1$, is a direct consequence of (10.27), (10.30), (10.34) and (10.37). The remaining relations can be read off from Corollary 10.4. □

If $w\geqslant 1$, as a consequence of (10.44) we obtain a pair of parallel lines $b_\pm(L_1,L_2)$ whose separation $2\sqrt{c/w}$ is bounded below by $\gtrsim\varepsilon^3$ and above by a universal constant, and with a transversal L_1 such that (10.45) and (10.46) hold. This will suffice for our application.

If $w\leqslant 1$, although $2\sqrt{c/w}$ is not bounded above, $\mathfrak{d}(L_1,L_2)$ is bounded above by a universal constant and below by $\gtrsim\varepsilon^3$, and $d^{\mathbb{H}}(o_1,e_{1,+}(L_1,L_2))$ is bounded by a definite multiple of ε^{-3}. In this case, we obtain a pair of skew lines making a standard angle and a controlled separation as follows.

Let

$$r=\sqrt{1+w^2}\sqrt{\frac{c}{w}}$$

denote the distance from the origin in $\mathbb{R}^2$ to $(\sqrt{c/w},\pm\sqrt{cw})=\pi(e_{1,\pm}(L_1,L_2))$. The points

$$\frac{\mathfrak{d}(L_1,L_2)+r}{r}\left(\sqrt{\frac{c}{w}},\sqrt{cw}\right)\quad\text{and}\quad\frac{r}{\mathfrak{d}(L_1,L_2)+r}\left(\sqrt{\frac{c}{w}},\sqrt{cw}\right)$$

lie on the projection of a line $L \in \text{lines}(H)$, with $z_i = L \cap L_i$, where

$$d^{\mathbb{H}}(z_1, e_{1,+}(L_1, L_2)) = \mathfrak{d}(L_1, L_2),$$
$$d^{\mathbb{H}}(z_2, e_{2,+}(L_1, L_2)) = \frac{r}{r + \mathfrak{d}(L_1, L_2)} \mathfrak{d}(L_1, L_2).$$

By direct computation, the slope $m(L)$ of $\pi(L)$ satisfies

$$\tfrac{2}{3} \leqslant m(L) \leqslant \tfrac{5}{2}. \tag{10.47}$$

From this, we get the following result.

Corollary 10.4. *The lines L and $b_+(L_1, L_2)$ are skew. The point*

$$\pi(L) \cap \pi(b_+(L_1, L_2))$$

lies on the line segment from $\pi(e_{1,+}(L_1, L_2))$ to $\pi(e_{2,+}(L_1, L_2))$. Moreover,

$$\tfrac{2}{5} \leqslant \tan\theta(L, b_+(L_1, L_2)) \leqslant \tfrac{3}{2} \tag{10.48}$$

and

$$\mathfrak{d}(L_1, L_2)\sqrt{\tfrac{2}{3}} \leqslant d^V(L, b_+(L_1, L_2)) \leqslant \mathfrak{d}(L_1, L_2). \tag{10.49}$$

Proof. Relation (10.48) follows directly from (10.47). By the discussion above, the triangle with vertices $(0,0)$, $\pi(e_{1,+}(L_1, L_2))$ and $\pi(e_{2,+}(L_1, L_2))$, and the triangle with vertices $(0,0)$, $\pi(z_1)$ and $\pi(z_2)$ have equal areas. It follows that the triangle with vertices $\pi(e_{1,+}(L_1, L_2))$, $\pi(z_1)$ and $\pi(L) \cap \pi(b_+(L_1, L_2))$, and that with vertices $\pi(e_{2,+}(L_1, L_2))$, $\pi(z_2)$ and $\pi(L) \cap \pi(b_+(L_1, L_2))$ (whose union is the symmetric difference of the previous ones), have equal areas as well. From this and the geometric interpretation of multiplication, we get (10.49). □

11. Proof of Lemma 7.2

In this section we prove Lemma 7.2. We begin with some measure-theoretic preliminaries which play a role here and in §12. Then we give the proof for the case of precisely monotone sets (for which the preliminaries are not required); compare also the proof in [21]. Finally, we show how to modify the argument to obtain the general case of Lemma 7.2.

11.1. Measure-theoretic preliminaries

Recall that $\mathcal{L}_3$ denotes the Lebesgue measure on $\mathbb{H}=\mathbb{R}^3$, which is a Haar measure of $\mathbb{H}$. Given $L\in\text{lines}(\mathbb{H})$, we denote by $\mathcal{H}^1_L$ the Hausdorff measure induced by the metric $d^{\mathbb{H}}$ on L. Recall that $\mathcal{N}$ denotes the unique left-invariant measure on lines($\mathbb{H}$) which is normalized so that $\mathcal{N}(\text{lines}(B_1(p)))=1$. Given $L\in\text{lines}(\mathbb{H})$, we denote its Heisenberg parallelism class by $[L]$, i.e., $[L]=\{gL:g\in\mathbb{H}\}$. $\mathbb{H}$ acts transitively on $[L]$. Therefore there exists a left-invariant measure $\mu_{[L]}$ on $[L]$, normalized so that for every measurable $A\subset\mathbb{H}$ we have

$$\int_{[L]}\mathcal{H}^1_{L'}(L'\cap A)\,d\mu_{[L]}(L')=\mathcal{L}_3(A). \tag{11.1}$$

The space of all parallelism classes will be denoted by $\mathfrak{P}=\{[L]:L\in\text{lines}(\mathbb{H})\}$. Each such parallelism class $[L]$ is uniquely determined by an angle $\theta\in[0,2\pi)$, corresponding to the angle of $\pi(L)\subset\mathbb{R}^2$ (recall that the lines $L\in\text{lines}(\mathbb{H})$ are oriented). Thus, there is an induced measure $d\theta$ on $\mathfrak{P}$ which corresponds to the standard measure on the circle S^1. By uniqueness, there exists a constant $c>0$ such that for all integrable $f:\text{lines}(\mathbb{H})\to\mathbb{R}$ we have

$$\int_{\text{lines}(\mathbb{H})} f\,d\mathcal{N}=c\int_{\mathfrak{P}}\int_{L'\in[L]} f(L')\,d\mu_{[L]}(L')\,d\theta([L]). \tag{11.2}$$

Let $\mathbb{P}\subset\text{lines}(\mathbb{H})\times\mathbb{H}$ denote the space of oriented pointed lines, i.e., $\mathbb{P}=\{(L,x):x\in L\}$. We shall use below the measure ν on $\mathbb{P}$, which is defined by setting, for every compactly supported continuous $f:\mathbb{P}\to\mathbb{R}$,

$$\int_{\mathbb{P}} f\,d\nu=\int_{\text{lines}(\mathbb{H})}\int_L f(L,x)\,d\mathcal{H}^1_L(x)\,d\mathcal{N}(L). \tag{11.3}$$

In the proof of Lemma 7.2 we shall use the space $\mathcal{C}$ of configurations, where a *configuration* is a quadruple $(L,x_1,x_2,[L'])$, where $L\in\text{lines}(\mathbb{H})$, $x_1,x_2\in L$ and $[L']\in\mathfrak{P}$ (i.e., a doubly pointed line and a parallelism class). The space $\mathcal{C}$ carries two measures σ_1 and σ_2, which are defined as follows. Given $f:\mathcal{C}\to\mathbb{R}$, which is continuous and compactly supported, let

$$\int_{\mathcal{C}} f\,d\sigma_1=\int_{\text{lines}(\mathbb{H})}\int_{\mathfrak{P}}\int_{L\times L} f(L,x_1,x_2,[L'])\,d(\mathcal{H}^1_L\times\mathcal{H}^1_L)(x_1,x_2)\,d\theta([L'])\,d\mathcal{N}(L), \tag{11.4}$$

and

$$\int_{\mathcal{C}} f\,d\sigma_2 \tag{11.5}$$
$$=\int_{\mathfrak{P}}\int_{[L']\times[L']}\int_{L_1} f(L(x),x,L(x)\cap L_2,[L'])\,d\mathcal{H}^1_{L_1}(x)\,d(\mu_{[L']}\times\mu_{[L']})(L_1,L_2)\,d\theta([L']),$$

where in (11.5), as in §8, given two parallel lines $L_1, L_2 \in [L']$ with distinct projections and $x \in L_1$, $L(x)$ denotes the unique element of lines($\mathbb{H}$) which passes through x and intersects L_2. (In §8, $L(x) \cap L_2$ was denoted x^*.)

The measures σ_1 and σ_2 are mutually absolutely continuous. Moreover, given a compact subset $K \subset \mathbb{H}$ of the Heisenberg group and $a>0$, let $\mathcal{C}(K,a)$ denote the set of configurations $(L, x_1, x_2, [L']) \in \mathcal{C}$ with $x_1, x_2 \in K$ and

$$d(\pi(x_1 L'(0)^{-1} L'), \pi(x_2 L'(0)^{-1} L')) \geqslant a,$$

i.e., the unique lines in $[L']$ which pass through x_1 and x_2 have projections of distance at least a, and x_1 and x_2 are in the compact set K. One checks from the definitions that, on the compact set $\mathcal{C}(K,a)$, the Radon–Nikodym derivatives $d\sigma_1/d\sigma_2$ and $d\sigma_2/d\sigma_1$ are continuous. Hence, for every measurable $A \subset \mathcal{C}(K,a)$, we have

$$\sigma_1(A) \asymp \sigma_2(A), \tag{11.6}$$

where the implied constants depend only on K and a. This observation will be used in the proof of Lemma 7.2 below.

11.2. A consequence of small non-convexity

Recall that the non-convexity $\mathrm{NC}_{B_r(x)}(E,L)$ was defined in (4.6). The following consequence of small non-convexity will be used in the present section and repeatedly in §12 (where the main result concerns δ-convex sets, which are more general than δ-monotone sets).

LEMMA 11.1. *Fix $p \in \mathbb{H}$ and $E \subset \mathbb{H}$. Let $L \in$ lines($\mathbb{H}$) be such that $\mathrm{NC}_{B_1(p)}(E,L) < \delta$. Assume that $[c,d] \subset [a,b] \subset L \cap B_1(p)$ and that*

$$\mathcal{H}^1_L([a,c] \cap E) > \delta \quad \text{and} \quad \mathcal{H}^1_L([d,b] \cap E) > \delta. \tag{11.7}$$

Then

$$\mathcal{H}^1_L([c,d] \cap E^c) \leqslant \delta.$$

Proof. For all sufficiently small $\eta>0$, we have $\mathcal{H}^1_L([a,c] \cap E), \mathcal{H}^1_L([d,b] \cap E) \geqslant \delta + \eta$. If $I \subset L \cap B_1(p)$ is an interval which exceeds the infimum on the right-hand side of (4.6) by at most $\frac{1}{2}\eta$, then the intersection of I with both $[a,c]$ and $[d,b]$ must have positive $\mathcal{H}^1_L$ measure. Thus, $I = [e,f] \supset [c,d]$. By the choice of I, we also have $\mathcal{H}^1_L([e,f] \cap E^c) < \delta + \frac{1}{2}\eta$, which implies that $\mathcal{H}^1_L([c,d] \cap E^c) < \delta + \frac{1}{2}\eta$. Letting η tend to 0 completes the proof. □

11.3. The case of precisely monotone sets

We call a set $E\subset B_1(p)$ *precisely monotone* if $E\cap L$ and $E^c\cap L$ are connected for all $L\in\text{lines}(\mathbb{H})$. Therefore, given such a pair (E,L), either $L\cap B_1(p)\subset E$, $L\cap B_1(p)\subset E^c$, or there exists a unique $q_{L,E}\in L$ such that $(L\setminus\{q_{L,E}\})\cap B_1(p)$ consists of two open intervals, one of which is contained in E and the other in E^c.

Choose a pair of parallel lines L_1 and L_2 with distinct projections, and recall that $X=X(L_1,L_2)$ denotes the ruled surface which is the union of lines $L(x)$ passing through L_1 and L_2. Here, $\{x\}=L(x)\cap L_1$. Note that $\mathbb{H}$ acts on L_1, L_2 and $X(L_1,L_2)$ by left translation.

LEMMA 11.2. *Fix $p\in\mathbb{H}$. There is a left-invariant function defined on pairs L_1, L_2 of parallel lines with distinct projections and taking values in 6-tuples of relatively open subsets of the surface $X(L_1,L_2)\cap B_1(p)$, $\{W_j(L_1,L_2)\}_{j=1}^6$, such that if E is a precisely monotone subset of $B_1(p)$ then for some $j(L_1,L_2,E)\in\{1,...,6\}$,*

$$W(L_1,L_2,E):=W_{j(L_1,L_2,E)}(L_1,L_2)$$

consists either entirely of points of E, or entirely of points of E^c.

Proof. We may assume that, say, $L_1=(t,b,-bt)$ and $L_2=(s,-b,-bs)$, with $\frac{1}{2}\leqslant b\leqslant 1$. Then, in the notation of Lemma 8.1, for $x=(t,b,-bt)\in L_1$ we have $x^*=(-t,-b,bt)$ and $L(x)=(r,r/b,-bt)$.

Consider the intervals

$$I_1:=\left(0,\tfrac{1}{6}\right),\quad I_2:=\left(\tfrac{1}{6},\tfrac{1}{3}\right)\quad\text{and}\quad I_3:=\left(\tfrac{1}{3},\tfrac{1}{2}\right),$$

and let $-I_j=\{t:-t\in I_j\}$. Define $W_1(L_1,L_2)$ to be $B_1(p)$ intersected with the union of those segments of the lines $L(x)$ which join $x\in L_1(I_1)$ to $x^*\in L_2(-I_1)$. Also, define $W_2(L_1,L_2)$ to be $B_1(p)$ intersected with the union of those subrays of the lines $L(x)$ for which $x\in L_1(I_1)$ and $x^*\in L_2(-I_1)$, whose endpoint is x and which are disjoint to the segment $[x,x^*]$. Define $W_3(L_1,L_2)$, $W_4(L_1,L_2)$ and $W_5(L_1,L_2)$, $W_6(L_1,L_2)$ analogously, corresponding to $j=2,3$, respectively.

There exists some $i(L_1,L_2,E)\in\{1,2,3\}$ such that $q_{E,L_1}\notin L_1(I_i)$ and $q_{E,L_2}\notin L_2(-I_i)$ (where possibly, one or both of q_{E,L_1} and q_{E,L_1} do not exist at all).

Fix $i=i(L_1,L_2,E)$ as above. There are four possibilities:

(1a) $L_1(I_i)\subset E$ and $L_1(-I_i)\subset E$,

(1b) $L_1(I_i)\subset E^c$ and $L_1(-I_i)\subset E^c$,

(2a) $L_1(I_i)\subset E$ and $L_2(-I_i)\subset E^c$,

(2b) $L_1(I_i)\subset E^c$ and $L_2(-I_i)\subset E$.

In cases (1a) and (1b) we can take $j(L_1,L_2,E)=2i-1$, while in cases (2a) and (2b) we can take $j(L_1,L_2,E)=2i$. □

Take a small interval $[0,d]$ in the center of $\mathbb{H}$ and consider all left translates

$$g(X(L_1,L_2)) = X(g(L_1),g(L_2)),$$

where $g\in[0,d]$. Clearly, for some subset $S=S(L_1,L_2,E)\subset[0,d]$ of measure $\geqslant\frac{1}{12}d$, if $g\in S$ then the integers $j(g(L_1),g(L_2),E)$ will all coincide and either $W(g(L_1),g(L_2),E)\subset E$ for all $g\in S$, or $W(g(L_1),g(L_2),E)\subset E'$ for all $g\in S$. Without essential loss of generality, we may assume that $0\in S$ and, say, $W(L_1,L_2,E)\subset E$. Hence, for all $g\in S$ we have

$$W(g(L_1),g(L_2),E) = g(W(L_1,L_2,E)) \subset E.$$

Choose an interior point $y\in W(L_1,L_2,E)\cap B_1(p)$, lying at a definite distance from the boundary, choose a line $L(x)$ such that $y\in L(x)$ and choose $L\in\text{lines}(\mathbb{H})$ making a definite angle with $L(x)$. Let $[L]$ denote the collection of lines parallel to L. By Lemma 8.1 (and continuity) we may assume that d has been chosen so small that if $g\in[0,d]$, we have $L'\in[L]$ and $L'\cap g(X(L_1,L_2))\cap B_{10d}(y)\neq\varnothing$. Then L' intersects $g(X(L_1,L_2))$ transversely. Parameterize each such L' so that $L'(0)=L'\cap W(L_1,L_2,E)$. Let J denote the smallest interval containing S. Then the length of J is $\geqslant\frac{1}{12}d$ and, by the monotonicity of E, we have $L'(J)\subset E$. This completes the proof in the precisely monotone case.

11.4. Proof of Lemma 7.2

Suppose now that $E\subset B_1(p)$ is ε^2-monotone on $B_1(p)$. Choose a measurable mapping $L\mapsto I_{L,E}\subset L$ (e.g., via an application of the measurable selection theorem in [44]) such that $I_{L,E}\cap B_1(p)$ and $I^c_{L,E}\cap B_1(p)$ are intervals and, for almost all L,

$$\mathcal{H}^1_L(I_{L,E}\triangle(E\cap L\cap B_1(p))) \leqslant 2\,\mathrm{NM}_{B_1(p)}(L,E). \tag{11.8}$$

Let $\bar{q}_{L,E}$ denote the common boundary point of $I_{L,E}$ and $I^c_{L,E}$.

Define I_1, I_2 and I_3 as above. Given a pair of parallel lines L_1,L_2 with distinct projections, define $i(L_1,L_2,E)$ as above, replacing $q_{L,E}$ by $\bar{q}_{L,E}$. Then, mutatis mutandis, define cases (1a)–(2b), $\{W_j(L_1,L_2)\}_{j=1}^6$, $j(L_1,L_2,E)$ and $W(L_1,L_2,E)$ as above.

In the claim below, Center($\mathbb{H}$) is equipped with its natural measure $\mathcal{L}$ and

$$W(L_1,L_2,E)\subset X(L_1,L_2)$$

is equipped with the natural surface measure $\mathcal{M}$ on $X(L_1,L_2)$.

Claim. There exists a universal constant $c>0$ with the following properties. Assume that $E\subset B_1(p)$ is ε^2-monotone on $B_1(p)$. Then there exists a pair of parallel lines L_1, L_2 whose projections lie at distance $\geqslant c$, such that for a fraction $\geqslant c$ of $g\in\text{Center}(\mathbb{H})\cap B_1(p)$ the surface $W(g(L_1), g(L_2), E)$ has measure $\geqslant c$ and, apart from a subset of measure $\leqslant\varepsilon/c$, it consists either entirely of points of E or entirely of points in E'.

Assume for the moment that the claim holds. Since $E\subset B_1(p)$ is ε^2-monotone, by (11.2) with $f(L)=\text{NM}_{B_1(p)}(E, L)$ and Markov's inequality (3.14), we can choose a parallelism class $[L]$ of lines which are all a definite amount transverse to $X(L_1, L_2)$ (i.e., the angle between these lines and the surface $X(L_1, L_2)$ is larger than a universal constant), such that apart from a subset of measure $\lesssim\varepsilon$ of lines in $[L]$, the remaining lines are all $\lesssim\varepsilon$-monotone. From this, together with the claim and Lemma 11.1, we directly obtain Lemma 7.2.

Proof of the claim. Rather than studying the space of surfaces $X(L_1, L_2)$ directly, it is convenient to decompose each such $X(L_1, L_2)$ into its collection of ruling lines L, and then to lift considerations to the space of all triples (L_1, L_2, L), where a *triple* (L_1, L_2, L) is a set of lines $L_1, L_2, L\in\text{lines}(B_1(p))$ such that L_1 and L_2 are parallel in the Heisenberg sense, and $L_1\cap L=\{x_1\}$ and $L_2\cap L=\{x_2\}$ are non-empty and lie in $B_1(p)$. Recall the set of configurations $\mathcal{C}$ that was defined in the paragraph following (11.3). Each configuration $(L, x_1, x_2, [L'])\in\mathcal{C}$ with $x_1, x_2\in B_1(p)$ determines, and is determined by, the triple (L_1, L_2, L), where $L_1=x_1L'(0)^{-1}L'$ and $L_2=x_2L'(0)^{-1}L'$ (thus, using previous notation, $L=L(x_1)$ and $x_2=x_1^*$). Let $\mathcal{C}_1$ denote the set of configurations $(L, x_1, x_2, [L'])\in\mathcal{C}$ with $x_1, x_2\in B_1(p)$.

Definition 11.3. A pointed line (L, x) is called *consistent* if either $x\in I_{L,E}$ and $x\in E$, or $x\in I^c_{L,E}\cap B_1(p)$ and $x\in E^c$. A configuration $(L, x_1, x_2, [L'])$ is called *consistent* if the pointed lines (L_1, x_1), (L_2, x_2), (L, x_1) and (L, x_2) are all consistent, where

$$L_1=x_1L'(0)^{-1}L' \quad \text{and} \quad L_2=x_2L'(0)^{-1}L'.$$

Such a configuration is called ε-*monotone* on $B_1(p)$ with respect to E if L_1, L_2 and L are all ε-monotone on $B_1(p)$ with respect to E.

Let G_ε denote the set of configurations which are consistent and ε-monotone on $B_1(p)$ and let $G'_\varepsilon=\mathcal{C}_1\setminus G_\varepsilon$. By using Lemma 11.1, it is immediate to verify from the definitions that if for some parallel lines $L_1, L_2\in\text{lines}(B_1(p))$ we have

$$\mathcal{H}^1_{L_1}(x\in L_1(I_{j(L_1,L_2,E)}) : (L(x), x, x^*, [L_1])\in G'_\varepsilon)\lesssim\varepsilon, \tag{11.9}$$

then, apart from a set of measure $\lesssim\varepsilon$, the set $W(L_1, L_2, E)$ consists either entirely of points of E or entirely of points in E^c.

LEMMA 11.4. *If E is ε^2-monotone on $B_1(p)$ then*

$$\sigma_1(G'_\varepsilon) \lesssim \varepsilon, \tag{11.10}$$

where σ_1 is defined as in (11.4).

Proof. We use the notation introduced in our discussion of measure-theoretical preliminaries at the beginning of this section. Fix $L\in\operatorname{lines}(B_1(p))$ and let $G'(L,E)$ denote the set of $x\in L\cap B_1(p)$ such that (L,x) is *not* consistent. By definition, we have

$$\mathcal{H}^1_L(G'(L,E)) \leqslant 2\,\mathrm{NM}_{B_1(p)}(L,E), \tag{11.11}$$

so

$$\int_{\operatorname{lines}(B_1(p))} \mathcal{H}^1_L(G'(L,E))\,d\mathcal{N}(L) \leqslant 2\,\mathrm{NM}_{B_1(p)}(E) \leqslant \varepsilon^2. \tag{11.12}$$

For $x\in B_1(p)$, denote by $m(x)$ the measure (with respect to $d\theta$) of the set of those $[L]\in\mathcal{P}$ such that, for the unique $L'\in[L]$ such that $x\in L'$, the pointed line (L',x) is *not* consistent. Then, by (11.1) and (11.2), we deduce from (11.12) that

$$\int_{B_1(p)} m(x)\,d\mathcal{L}_3(x) \lesssim \varepsilon^2. \tag{11.13}$$

For $x_1,x_2\in L$, define $F(x_1,x_2)$ by

$$F(x_1,x_2) = m(x_1)+m(x_2)+\chi_{I_{L,E}\triangle(E\cap L\cap B_1(p))}(x_1)+\chi_{I_{L,E}\triangle(E\cap L\cap B_1(p))}(x_2). \tag{11.14}$$

Then,

$$\begin{aligned}\int_{\operatorname{lines}(B_1(p))}\int_{x_1,x_2\in L\cap B_1(p)} & F(x_1,x_2)\,d(\mathcal{H}^1_L\times\mathcal{H}^1_L)(x_1,x_2)\,d\mathcal{N}(L)\\ &\overset{(11.13)}{\lesssim} \varepsilon^2+\int_{\operatorname{lines}(B_1(p))}\mathcal{H}^1_L(I_{L,E}\triangle(E\cap L\cap B_1(p))\,d\mathcal{N}(L)\\ &\overset{(11.8)}{\lesssim} \varepsilon^2+\mathrm{NM}_{B_1(p)}(E)\\ &\lesssim \varepsilon^2.\end{aligned} \tag{11.15}$$

It follows from (11.15) and (11.4) that, if we denote by $A\subseteq\mathcal{C}$ the set of configurations $(L,x_1,x_2,[L'])$ which are not consistent, then $\sigma_1(A)\lesssim\varepsilon^2$.

Let $B\subset\mathcal{C}$ denote the set of configurations $(L,x_1,x_2,[L'])$ which are not ε-monotone on $B_1(p)$ with respect to E. This implies that, for $(L,x_1,x_2,[L'])\in B$, we have

$$\mathrm{NM}_{B_1(p)}(L,E)+\mathrm{NM}_{B_1(p)}(x_1L'(0)^{-1}L',E)+\mathrm{NM}_{B_1(p)}(x_2L'(0)^{-1}L',E)\geqslant\varepsilon.$$

Hence, using Markov's inequality (3.14), combined with the identities (11.4) and (11.2), we deduce that

$$\sigma_1(B)\lesssim\frac{\mathrm{NM}_{B_1(p)}(E)}{\varepsilon}\leqslant\varepsilon.$$

Since, by definition, $\sigma_1(G'_\varepsilon)\leqslant\sigma_1(A)+\sigma_1(B)$, the proof of (11.10) is complete. □

Set $a=\frac{1}{100}$ and recall that, as defined before (11.6), $\widetilde{\mathcal{C}}=\mathcal{C}(B_1(p),a)$ is the set of all configurations $(L,x_1,x_2,[L'])\in\mathcal{C}$ such that $x_1,x_2\in B_1(p)$ and the projections of the parallel lines $x_1L'(0)^{-1}L'$ and $x_2L'(0)^{-1}L'$ have distance at least a. It follows from (11.6) and (11.10) that $\sigma_2(G'_\varepsilon\cap\widetilde{\mathcal{C}})\lesssim\varepsilon$, where the measure σ_2 is as in (11.5). It follows from (11.5) that there exists a parallelism class $[L']\in\mathfrak{P}$ such that, if we define

$$F=\{L_1,L_2\in[L']:d(\pi(L_1),\pi(L_2))\geqslant a\},$$

then

$$\int_F \mathcal{H}^1_{L_1}(\{x\in L_1\cap B_1(p):(L(x),x,x^*,[L'])\in G'_\varepsilon\})\,d(\mu_{[L']}\times\mu_{[L']})(L_1,L_2)\lesssim\varepsilon. \tag{11.16}$$

Partition the set of all $(L_1,L_2)\in F\times F$ into equivalence classes, where (L_1,L_2) and (L'_1,L'_2) are equivalent if there exists $g\in\text{Center}(\mathbb{H})$ such that $L'_1=g(L_1)$ and $L'_2=g(L_2)$. We deduce from (11.16) that there exist $L_1,L_2\in F$, an index $j\in\{1,...,6\}$ and a subset $D\subset\text{Center}(\mathbb{H})\cap B_1(p)$ of measure $\gtrsim 1$ such that for all $g\in D$ we have $j(g(L_1),g(L_2),E)=j$, the surface $W(g(L_1),g(L_2),E)$ has measure $\gtrsim 1$ and, by another application of Markov's inequality (3.14),

$$\mathcal{H}^1_{L_1}(\{x\in L_1(I_j):(L(x),x,x^*,[L_1])\in G'_\varepsilon\})\lesssim\varepsilon.$$

The claim now follows from the discussion following (11.9), and hence the proof of Lemma 7.2 is complete. □

12. Proof of Proposition 7.3

In this section we will consider both lines in $\mathbb{R}^2$ and their horizontal lifts to $\mathbb{H}$. It will be convenient to use a notation which differs somewhat from that employed elsewhere in the paper. This change will also prevent an undesirable proliferation of subscripts, which would result if we continued to write L (exclusively) for horizontal lines in $\mathbb{H}$.

Let $\pi:\mathbb{H}\to\mathbb{R}^2$ be the canonical map. Given a line segment $\tau_1\subset\mathbb{R}^2$ with $\tau_1(0)=(0,0)$, denote by $\tilde{\tau}_1$ its unique lift to a segment of a horizontal line in $\mathbb{H}$ emanating from the origin. Similarly, consider a once-broken line segment $\tau_1\cup\tau_2$, i.e., a pair of line segments such that the initial point of the line segment τ_2 is the final point of τ_1. Again, we assume that $\tau_1(0)=(0,0)$. Denote by $\widetilde{\tau_1\cup\tau_2}$ the lift of $\tau_1\cup\tau_2$ to a continuous broken horizontal line in $\mathbb{H}=\mathbb{R}^3$, emanating from the origin.

In place of the line L in Proposition 7.3, we will write $\widetilde{\gamma}$. Since both the statement and proof are somewhat technical, for purposes of exposition, we first consider the model case in which $\widetilde{\gamma}(0)\in E$, $B_{Cr}(\widetilde{\gamma}(1))\subset E$ and E is precisely convex. In this case, we will show that for a large enough universal constant C, there exists $c>0$ such that $B_{csr}(\widetilde{\gamma}(s))\subset E$ for all $0\leqslant s\leqslant 1$.

Recall that in Proposition 7.3 we are given parameters $\varkappa, \eta, \xi, r, \varrho\in(0,1)$ such that $\varrho\leqslant\frac{1}{2}\varkappa r^2$ and $\varrho\leqslant c$, where $c\in(0,1)$ will be a small enough universal constant to be determined below. For the reader's convenience, we recall here the values of δ_1 and δ_2 in Proposition 7.3, namely (7.9) and (7.10):

$$\delta_1 = \frac{c\varkappa^3\eta^2\xi\varrho^3}{r}, \tag{12.1}$$

$$\delta_2 = c\varkappa^6\eta^3\xi^2\varrho^3 r^6. \tag{12.2}$$

In the proof we will also use an auxiliary parameter $\omega>0$, which will be given by

$$\omega = \bar{c}\varkappa^3\eta\xi r^2, \tag{12.3}$$

where $\bar{c}\geqslant\sqrt{c}$ will be an appropriately chosen universal constant.

12.1. Values of constants

For the reader's convenience, we include here a summary of the steps of the proof of Proposition 7.3, in which the particular values of the constants enter explicitly. For the constraint $\varrho\leqslant\frac{1}{2}\varkappa r^2$, (12.15), (12.45) and (12.61); for ω, (12.25), (12.33) and (12.53); for δ_1, (12.25) and (12.59); and for δ_2, (12.33) and (12.55).

12.2. Proof of model case

For reasons of exposition, we begin with the proof of the model case described above.

Remark 12.1. Even though the statement of Proposition 7.3 pertains to balls, the multiplicative structure of $\mathbb{H}$ and its relation to the geometry will necessitate the introduction of certain product cylinders, where the product structure corresponds to $\mathbb{H}=\mathbb{R}^2\times\mathbb{R}$. In fact, our argument will show that if a cylinder centered at $\widetilde{\gamma}(1)$ with base radius r^2 and height r is contained in E, then a cylinder with base radius $\gtrsim sr^2$ and height sr centered at $\widetilde{\gamma}(s)$ will be contained in E. Here, both heights are measured with respect to the metric $d^{\mathbb{H}}$.

Remark 12.2. Let O_θ be as in §2.4. By applying a suitable left translation and action by O_θ, we may assume that $\widetilde{\gamma}$ is the line $(s,0,0)$. In the present subsection and in the proof of Proposition 7.3, we consider only points $(\bar{a},\bar{b},\bar{c})\in B_{csr}(\widetilde{\gamma}(s))$ with $\bar{c}\geqslant 0$. Points with $\bar{c}\leqslant 0$ are handled by the symmetric argument, i.e., by interchanging the roles of ϕ and ψ below.

Let $\Theta\subset\mathbb{R}^2$ denote the angular sector which is the union of rays τ such that $\tau(0)=\gamma(0)=(0,0)$ and $\tau(1)\in B_{r/2}(\gamma(1))$, where here $B_{r/2}(\gamma(1))$ denotes the Euclidean disk in $\mathbb{R}^2$. Let $\widetilde{\Theta}$ denote the union of the lifts of the rays $\tau\subset\Theta$. By the convexity of E, it follows that $\tau\subset\Theta$ implies that $\tilde{\tau}(s)\in E$ for $0\leqslant s\leqslant 1$, where s denotes arc length. To prove the non-quantititative version of (7.13), it suffices to show that for any $\tau\subset\Theta$,

$$\pi^{-1}(\tau(s))\cap B_{csr}(\tilde{\tau}(s))\subset E. \tag{12.4}$$

Set

$$\lambda=\arctan r^2, \tag{12.5}$$

and let ϕ and ψ denote the rays in Θ which make an angle λ with τ. For the sake of proving (12.4), by applying a suitable rotation, assume that τ is the x-axis, so that we have $\tilde{\tau}(s)=(s,0,0)$. Choose the parameterization u such that $\tilde{\phi}(u)=(u,-ur^2,0)$ and $\tilde{\psi}(u)=(u,ur^2,0)$.

Fix $s_0\in[0,1]$ and let u_0 be the unique value of u such that the line segment χ, from $\phi(u_0)$ to $\psi(1)$, intersects τ at $\tau(s_0)$, where $\chi(0)=\phi(u_0)$. Thus

$$u_0=\frac{s_0}{2-s_0}. \tag{12.6}$$

Let ϕ_1 denote the segment of ϕ from $\tau(0)=\phi(0)$ to $\phi(u_0)$. Let $\widetilde{\chi}$ denote the segment of $\widetilde{\phi_1\cup\chi}$ which lifts χ. Choose the parameterization t such that

$$\widetilde{\chi}(t)=u_0(1,-r^2,0)\cdot t((1,r^2,0)-u_0(1,-r^2,0)), \tag{12.7}$$

where above, as usual, "$\cdot$" denotes multiplication in $\mathbb{H}$. If we define d_0 by $\chi(d_0)=\tau(s_0)$, then from (12.7) it follows that

$$d_0=\frac{s_0}{2}, \tag{12.8}$$

$$\widetilde{\chi}(d_0)=\left(s_0,0,\frac{2s_0^2r^2}{2-s_0}\right). \tag{12.9}$$

By the assumed convexity of E, we have $\tilde{\phi}(u)\subset E$ for all $0\leqslant u\leqslant 1$. From (12.7) for $t=1$, and the box-ball principle (2.17), we may assume that C has been chosen such that $\widetilde{\chi}(1)\in B_{Cr}(\widetilde{\gamma}(1))\subset E$. Since E is convex, this gives $\widetilde{\chi}(d_0)\in E$. Similarly, by (2.17), for

$$0\leqslant v\leqslant\frac{2s_0^2r^2}{2-s_0}$$

the line $(0,0,-v)\cdot\widetilde{\chi}$ intersects $\widetilde{\Theta}\cap E$ and satisfies $(0,0,-v)\cdot\widetilde{\chi}(1)\in B_{Cr}(\widetilde{\gamma}(1))\subset E$. Again, by convexity of E, we get $(0,0,-v)\cdot\widetilde{\chi}(d_0)\in E$. It is straightforward to check that this implies that $B_{crs_0}(\tilde{\tau}(s_0))\subset E$ for some universal constant $c>0$, as required.

Remark 12.3. In our proof of Proposition 7.3, it is necessary to replace $\widetilde{\chi}$ by a 1-parameter family of lines $\widetilde{\chi}_\theta$, to be defined below. To this end, we observe that there is some flexibility in the definition of $\widetilde{\chi}$. First note that, from (12.5), (12.6) and (12.8) and the law of sines, it follows that for some universal constant $c_1>0$ we have

$$c_1 r^2 \leqslant \angle(\tilde{\tau}'(s_0), \widetilde{\chi}'(d_0)). \tag{12.10}$$

For all

$$0 \leqslant \theta \leqslant c_2 r^2, \tag{12.11}$$

with a small constant $c_2>0$, let $\widetilde{\chi}_\theta$ denote the horizontal line through $\widetilde{\chi}(d_0)=\tilde{\tau}(s_0)$ obtained by rotating $\widetilde{\chi}$ by an angle θ about the point $\widetilde{\chi}(d_0)$, in the clockwise direction. The constant C can be chosen such that $\widetilde{\chi}_\theta$ intersects $B_{Cr}(\widetilde{\gamma}(1))$. It will also intersect $\widetilde{\Theta}\cap E$. In fact, if $q=g(\tilde{\tau}(s_0))\in B_{cs_0r^2}(\tilde{\tau}(s_0))$, then $g(\widetilde{\chi}_\theta)$ intersects both $\Theta\cap E$ and $B_{Cr}(\widetilde{\gamma}(1))$. It follows that for *any* $q\in B_{cs_0r^2}(\tilde{\tau}(s_0))$ and any $\widetilde{\chi}_\theta$ with θ as in (12.11), in verifying that $\pi^{-1}(\pi(q))\cap B_{crs_0}(\tilde{\tau}(s_0))\subset E$, we can use only the lines which are parallel to some $\widetilde{\chi}_\theta$.

12.3. Intermediate case and proof

Before going to the proof of Proposition 7.3, we consider a slight generalization of the model case. As before, assume that $B_{Cr}(\widetilde{\gamma}(0))\in E$ and that E is convex. Assume in addition that $g(\widetilde{\gamma}(0))\in E$ for some g, with

$$d^{\mathbb{H}}(g(\widetilde{\gamma}(0)), \widetilde{\gamma}(0)) \leqslant \varrho, \tag{12.12}$$

where we recall that we are assuming that $\varrho\leqslant\varkappa r^2$. The box-ball principle (2.17), together with (12.12), implies that for all $s>0$ we have

$$d^{\mathbb{H}}(\widetilde{\gamma}(s), g(\widetilde{\gamma}(s)) \lesssim \max\{\varrho, \sqrt{s\varrho}\}. \tag{12.13}$$

Since $\varrho\leqslant r^2$, it follows from (12.13) that, provided C is large enough, the assumption $B_{Cr}(\widetilde{\gamma}(1))\subseteq E$ implies that $B_{Cr/2}(g\widetilde{\gamma}(1))\subseteq E$. Now, a computation like those above shows that for all $0\leqslant s\leqslant 1$ we have

$$B_{csr}(g\widetilde{\gamma}(s)) \subset E \quad \text{for all } s\in[0,1], \tag{12.14}$$

where $c>0$ is a universal constant. It follows from (12.12), the bound $\varrho\leqslant\varkappa r^2$ and (12.14), that there exists a constant $c'>0$ such that

$$B_{csr/2}(\widetilde{\gamma}(s)) \subset E \quad \text{for all } s\in[c'\varkappa, 1]. \tag{12.15}$$

Remark 12.4. Here it is worthwhile to emphasize that, in order to obtain (12.15), we must take ϱ proportional to r^2 (rather than to r); compare also the use of cylinders $B_{csr^2}(\tau(0))\times[-(sr)^2,(sr)^2]$ in the model case above and below.

Proof of Proposition 7.3. The arguments in the model and intermediate cases use $\widetilde{\Theta}\subset E$ and $g(\widetilde{\Theta})\cap B_1(\widetilde{\gamma}(0))\subset E$, respectively. Since in the general case our hypotheses are measure theoretic, we will employ integral geometry and, in particular, the measure-theoretical preliminaries given in §11.

The first (and most involved) step is to show the existence of a thickened version $\mathbb{S}$ of the angular sector $g(\widetilde{\Theta})\subset E$ of the intermediate case:

$$\mathbb{S}=\bigcup_{\zeta\in A} g\zeta(\widetilde{\Theta}) \tag{12.16}$$

for some $g\in\mathbb{H}$, where A is a suitably chosen subset of the center of $\mathbb{H}$ such that

$$g\zeta(\widetilde{\gamma}(0))\in B_\varrho(\widetilde{\gamma}(0))\cap E$$

for all $\zeta\in A$. We think of $\mathbb{S}$ as a "thin slab".

The slices $g\zeta(\widetilde{\Theta})$ of $\mathbb{S}$ will be shown to consist almost entirely of points of E. Hence, the vertical structure exhibited in (12.16) implies that the mass of $E\cap\mathbb{S}$ is very evenly distributed. In the presence of sufficiently small non-convexity of E, this compensates for the fact that the mass of $E\cap\mathbb{S}$, while subject to a definite lower bound depending on η (see (12.19)) might still be very small, which might otherwise prevent $\mathbb{S}$ from serving as an adequate substitute for $\widetilde{\Theta}$.

The following lemma summarizes the essential properties of $\mathbb{S}$. Eventually, we will be interested in the behavior of the intersection of $\mathbb{S}$ with the cylinder (with coordinate description)

$$B_{csr^2}(\tau(s))\times[-c(sr)^2,c(sr)^2],$$

whose intrinsic height is $2sr$ and whose measure is $\asymp(sr^2)^2(sr)^2$. This may help to explain the right-hand sides of (12.17)–(12.20).

Lemma 12.5. *There exist* $g\in\mathbb{H}$ *satisfying* (12.12) *and* $A\subset\operatorname{Center}(\mathbb{H})$ *such that, for* $\mathbb{S}$ *as in* (12.16), *the following properties hold*:

(a)

$$\mathbb{S}\subset g(\mathbb{R}^2\times I) \tag{12.17}$$

for some $I\subset[-\varrho^2,\varrho^2]\subset\operatorname{Center}(\mathbb{H})$, *with*

$$\mathcal{L}(I)\asymp\xi\varrho^2, \tag{12.18}$$

where $\mathcal{L}$ *denotes the Hausdorff measure on the center of* $\mathbb{H}$, *with its induced* $\frac{1}{2}$-*snowflake metric*;

(b) $\mathbb{S}$ *has a definite thickness*:

$$\mathcal{L}(A) \gtrsim \eta\xi\varrho^2; \tag{12.19}$$

(c) $\mathbb{S}$ *consists almost entirely of points of* E, *i.e., for all* $\zeta\in A$,

$$\mathcal{L}_2((g\zeta(\widetilde{\Theta})\setminus(B_\varrho(\widetilde{\gamma}(0))\cup B_r(\widetilde{\gamma}(1))))\cap E^c) \lesssim \omega r^2. \tag{12.20}$$

Granted Lemma 12.5, the proof of Proposition 7.3 is not difficult to complete. We will now give a brief overview of the argument to come. Since E is δ_2-convex, we can chose $\widetilde{\chi}_\theta$, as in Remark 12.3, such that for a fraction very close to 1 of lines $L\in[\widetilde{\chi}_\theta]$, the non-convexity $\mathrm{NC}_{B_{Cr}(\widetilde{\gamma}(0))}(E,L)$ is very small. By a quantitative version of the proofs of the previously considered model and intermediate cases, it follows that such lines consist almost entirely of points of E. The argument uses (12.19), (12.20) and Lemma 11.1 below. In fact, Lemma 11.1 enters, in a similar fashion, in the proof of Lemma 12.5.

Proof of Lemma 12.5. In the next few subsections we will show that for most points $g(\widetilde{\gamma}(0))\in E\cap B_\varrho(\widetilde{\gamma}(0))$, the sector $g(\widetilde{\Theta})$ consists almost entirely of points of E. The remainder of the construction will be a simple consequence of this fact.

12.4. Implication of almost full measure of E in $B_{Cr}(\widetilde{\gamma}(1))$

By the assumption (7.11), we have

$$\mathcal{L}_3(E\cap B_\varrho(\widetilde{\gamma}(0))) \geqslant \eta\mathcal{L}_3(B_\varrho(\widetilde{\gamma}(0))) \gtrsim \eta\varrho^4. \tag{12.21}$$

Let $C>0$ be the (large enough, but fixed) constant that was chosen in the intermediate case of the proof. Assume, as in (7.12), that $L(1)\in\mathrm{int}_{\delta_1,Cr}(E)$. Recall that the parameter ω was defined in (12.3). At the end of the proof, it will become clear why it is necessary to choose this value of ω; see (12.25), (12.33) and (12.53).

For $\tau\subset\Theta$, set

$$S_{[\tilde\tau]} = \{\hat\tau\in[\tilde\tau] : 0<\mathcal{H}^1_{\hat\tau}(\hat\tau\cap E\cap B_\varrho(\widetilde{\gamma}(0)))\leqslant\omega\eta\varrho\}, \tag{12.22}$$

$$T_{[\tilde\tau]} = \big\{\hat\tau\in[\tilde\tau] : 0<\mathcal{H}^1_{\hat\tau}(\hat\tau\cap E\cap B_\varrho(\widetilde{\gamma}(0))) \text{ and } \mathcal{H}^1_{\hat\tau}(\hat\tau\cap E\cap B_{r/2}(\widetilde{\gamma}(1)))\leqslant\tfrac{1}{2}r\big\}. \tag{12.23}$$

Without loss of generality, we may assume that $\varkappa$ is small enough that if $\hat\tau\in S_{[\tilde\tau]}$ then $\hat\tau(1)\in B_{r/2}(\widetilde{\gamma}(1))$; see (12.13). From (7.12), (7.9) and (11.1), we get

$$\mu_{[\tilde\tau]}(T_{[\tilde\tau]}) \lesssim \delta_1 r^3 \leqslant \omega\eta\varrho^3, \tag{12.24}$$

provided that the lower bound

$$\omega \geqslant \frac{\delta_1}{\eta}\left(\frac{r}{\varrho}\right)^3 \tag{12.25}$$

holds. An inspection of our choice of parameters in (12.1) and (12.3) shows that (12.25) does indeed hold in our setting. (Note that here we used the requirement $\bar{c}\geqslant\sqrt{c}\geqslant c$.)

12.5. Estimates in the space of pointed lines

Recall that ν denotes the measure on $\mathbb{P}$, the space of pointed lines, as in (11.3). Let

$$\begin{aligned} U^{\#} &= \{(\hat\tau,\hat\tau(0)) : \hat\tau(0)\in E\cap B_\varrho(\widetilde\gamma(0)) \text{ and } \hat\tau\subset\hat\tau(0)\widetilde\gamma(0)^{-1}(\widetilde\Theta)\},\\ S^{\#} &= \{(\hat\tau,\hat\tau(0)) : \hat\tau\in S\cap g(\widetilde\Theta) \text{ and } \hat\tau(0)=g(\widetilde\gamma(0))\in E\cap B_\varrho(\widetilde\gamma(0))\},\\ T^{\#} &= \{(\hat\tau,\hat\tau(0)) : \hat\tau(0)\in E\cap B_\varrho(\widetilde\gamma(0)), \hat\tau\in T_{[\hat\tau]} \text{ and } \hat\tau\subset\hat\tau(0)\widetilde\gamma(0)^{-1}(\widetilde\Theta)\}. \end{aligned} \tag{12.26}$$

Then

$$\nu(U^{\#}) \asymp r^2\mathcal{L}_3(E\cap B_\varrho(\widetilde\gamma(0))) \gtrsim \eta r^2\varrho^4, \tag{12.27}$$

$$\nu(S^{\#}) \lesssim \omega\eta r^2\varrho^4, \tag{12.28}$$

$$\nu(T^{\#}) \lesssim \omega\eta r^2\varrho^4. \tag{12.29}$$

Relation (12.27) follows from (11.1)–(11.3), (12.5) and (12.21). Relations (12.28) and (12.29) follow similarly, using (12.22) and (12.24), respectively.

Recall that E is δ_2-convex on $B_{2C}(\widetilde\gamma(0))$, for δ_2 as in (7.10). Define

$$V = \big\{L\in \operatorname{lines}(B_{2C}(\widetilde\gamma(0))) : \mathrm{NC}_{B_{2C}(\widetilde\gamma(0))}(E,L)\geqslant \tfrac12\omega\big\}, \tag{12.30}$$

$$V^{\#} = \{(L,x)\in\mathbb{P} : L\in V \text{ and } x\in E\cap B_\varrho(\widetilde\gamma(0))\}. \tag{12.31}$$

By (4.6), Markov's inequality (3.14) and (11.3), we have

$$\nu(V^{\#}) \lesssim \frac{\delta_2}{\omega}\varrho \leqslant \omega\eta r^2\varrho^4, \tag{12.32}$$

provided that, in addition to (12.25), we also assume that

$$\omega \geqslant \sqrt{\frac{\delta_2}{\eta r^2\varrho^3}}. \tag{12.33}$$

An inspection of our choice of parameters in (12.2) and (12.3) shows that (12.33) holds in our setting (note that we used here the requirement $\bar{c}\geqslant\sqrt{c}$).

It follows from (12.27)–(12.29) and (12.32) that

$$\nu(S^{\#}\cup T^{\#}\cup V^{\#})\lesssim\omega\nu(U^{\#}). \tag{12.34}$$

Define

$$W^{\#}=U^{\#}\setminus(S^{\#}\cup T^{\#}\cup V^{\#}). \tag{12.35}$$

Then, for $(\hat{\tau},\hat{\tau}(0))\in W^{\#}$, the definitions (12.26) and (12.31), together with Lemma 11.1, imply that if we set

$$J=\left[\varrho+\tfrac{1}{2}\omega, 1-\tfrac{1}{2}r-\tfrac{1}{2}\omega\right],$$

then

$$\mathcal{H}^1_{\hat{\tau}}(E^c\cap\hat{\tau}(J))\leqslant\tfrac{1}{2}\omega. \tag{12.36}$$

For $x\in E\cap B_\varrho(\tilde{\gamma}(0))$, $x=g(\tilde{\gamma}(0))$, put

$$\mathfrak{G}(x)=\{\hat{\tau}\subset g(\widetilde{\Theta}):(\hat{\tau},x)\in W^{\#}\}. \tag{12.37}$$

Note that, using the above notation, the set of all $\hat{\tau}\subset g(\widetilde{\Theta})$ with $\hat{\tau}(0)=x$ is parameterized by the corresponding angle ϕ in the sector given by (12.5). It follows from (12.34), (12.27) and (12.21) that there exists a large enough universal constant $c>0$ such that, if we put

$$E_1=\{x\in E\cap B_\varrho(\tilde{\gamma}(0)):d\phi(\mathfrak{G}(x))\leqslant c\omega r^2\}, \tag{12.38}$$

then

$$\mathcal{L}_3(E_1)\geqslant\tfrac{1}{2}\mathcal{L}_3(E\cap B_\varrho(\tilde{\gamma}(0)))\gtrsim\eta\varrho^4. \tag{12.39}$$

In the following corollary and below, we consider the splitting of the measure

$$\mathcal{L}_3=\mathcal{L}\times\mathcal{L}_2=\mathcal{H}^2\times\mathcal{L}_2, \tag{12.40}$$

relative to the product structure on $\mathbb{H}=\mathbb{R}^3$, where $\mathcal{L}=\mathcal{H}^2$ corresponds to the center of $\mathbb{H}$ (with its induced $\frac{1}{2}$-snowflake metric) and $\mathcal{L}_2$ to the plane $(x,y,0)\subset\mathbb{R}^3$. In connection with (12.41) below, recall that $\mathcal{L}_3(B_\varrho(\tilde{\gamma}(0)))\asymp\varrho^4=\varrho^2\times\varrho^2$, where the product corresponds to the decomposition in (12.40).

COROLLARY 12.6. *There exists* $q\in\pi(E_1)$ *such that*

$$\eta\varrho^2\lesssim\mathcal{L}(\pi^{-1}(q)\cap E_1)\lesssim\varrho^2. \tag{12.41}$$

If $g(\tilde{\gamma}(0))\in\pi^{-1}(q)\cap E_1$ *then*

$$\mathcal{L}_2(E^c\cap(g(\widetilde{\Theta})\setminus(B_\varrho(\tilde{\gamma}(0))\cup B_r(\tilde{\gamma}(1)))))\lesssim\omega r^2. \tag{12.42}$$

Proof. Relation (12.41) follows from the splitting of the measure in (12.40) and (12.39), whereas relation (12.42) follows from the definitions of $U^{\#}$, $S^{\#}$, $T^{\#}$, $V^{\#}$ and $W^{\#}$, together with (12.36)–(12.38). □

12.6. Construction of $\mathbb{S}$ and completion of the proof of Lemma 12.5

By virtue of (12.41), we can choose an interval $I\subset[-\varrho^2,\varrho^2]$ such that

$$\mathcal{L}(I)\asymp\xi\varrho^2, \tag{12.43}$$

$$\mathcal{L}(\pi^{-1}(q)\cap E_1\cap q(I))\gtrsim\eta\xi\varrho^2. \tag{12.44}$$

Define a subset A of the center of $\mathbb{H}$ by

$$A=\pi^{-1}(0,0)\cap q^{-1}(\pi(E_1))\cap I.$$

Choose $g=q\tilde{\gamma}(0)^{-1}$. Then g satisfies (12.12) and, by (12.43) and (12.44), statements (a) and (b) of Lemma 12.5 hold. Since, by the definition of q, for all $\zeta\in A$ we have $g\zeta\tilde{\gamma}(0)\in\pi^{-1}(q)\cap E_1$, Corollary 12.6 implies that part (c) holds as well. □

12.7. Completion of the proof of Proposition 7.3

By choosing c in Proposition 7.3 sufficiently small, we may assume without essential loss of generality that g of Lemma 12.5 is the point $(0,0,0)$ and that I is symmetric about the origin. Fix $s\in[\varkappa,1]$ and let τ and $\tilde{\chi}_\theta$ be as in Remark 12.3, with $0\leqslant\theta\leqslant\frac{1}{100}r^2$ to be determined below.

Let $\mathcal{S}$ denote the square with center $\tau(s)$ and side length csr^2, with one side parallel to χ_θ. Of the two sides of $\mathcal{S}$ which are orthogonal to χ_θ, let α denote the one which is furthest from the origin $(0,0)\in\mathbb{R}^2$. Let $\mathcal{R}$ denote the rectangle with one side parallel to χ_θ and of length $2sr^2+\frac{1}{2}s$, and such that of the two sides which are orthogonal to χ_θ the one furthest from $(0,0)\in\mathbb{R}^2$ is α. In particular, $\mathcal{S}\subset\mathcal{R}$. By a standard covering argument, it will suffice to show that at most a fraction ξ of points of $\mathcal{S}\times[0,(csr)^2]$ lie in E^c; see (7.13). Since, by our assumptions $\varrho\leqslant\frac{1}{2}\varkappa r^2$ and $s\geqslant\varkappa$, we have

$$\xi\varrho^2\leqslant\tfrac{1}{2}\xi(sr)^2, \tag{12.45}$$

it suffices to show that at most a fraction ξ of the points of $\mathcal{S}\times([0,(csr)^2]\setminus I)$ lie in E^c.

Let

$$\mathfrak{F}_\theta=\{L\in[\tilde{\chi}_\theta]:L\cap(\alpha\times([0,(csr)^2]\setminus I))\neq\varnothing\}.$$

Note that $\mathfrak{F}$ can (essentially) be described alternatively as consisting of those lines in $[\tilde{\chi}_\theta]$ which intersect $\mathcal{R}$. If follows from (11.1) that

$$\mu_{[\tilde{\chi}_\theta]}(\mathfrak{F}_\theta)\gtrsim(sr^2)(sr)^2=s^3r^4. \tag{12.46}$$

Moreover, since E is δ_2-convex on $B_{2C}(\widetilde{\gamma}(0))$, there exists $\theta\in\left(0,\frac{1}{100}r^2\right)$ such that

$$\int_{L\in\mathfrak{F}_\theta}\mathrm{NC}_{B_{2C}(\widetilde{\gamma}(0))}(E,L)\,d\mu_{[\widetilde{\chi}_\theta]}(L)\leqslant\frac{100\delta_2}{r^2}. \tag{12.47}$$

We shall fix θ as in (12.47) from now on. It follows from (12.46) and (12.47) that there exists a universal constant $c''>0$ such that, if we define

$$\mathfrak{F}_\theta^*=\left\{L\in\mathfrak{F}_\theta:\mathrm{NC}_{B_{2C}(\widetilde{\gamma}(0))}(E,L)\leqslant\frac{c''\delta_2}{\xi s^3r^6}\right\}, \tag{12.48}$$

then

$$\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta^*)\geqslant\left(1-\tfrac{1}{8}\xi\right)\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta). \tag{12.49}$$

Lemma 12.5 implies that

$$\mathcal{L}_3((\mathbb{S}\setminus(B_\varrho(\widetilde{\gamma}(0))\cup B_r(\widetilde{\gamma}(1))))\cap E^c)\lesssim\omega r^2\xi\varrho^2,$$

and hence, using (11.1) and Markov's inequality (3.14), we see that there exists a universal constant $c'''>0$ such that, if we define

$$\mathfrak{F}_\theta^{**}=\left\{L\in\mathfrak{F}_\theta^*:\mathcal{H}_L^1((\mathbb{S}\setminus(B_\varrho(\widetilde{\gamma}(0))\cup B_r(\widetilde{\gamma}(1))))\cap E^c\cap L)\leqslant\frac{c'''\omega\varrho^2}{s^3r^2}\right\}, \tag{12.50}$$

then

$$\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta^{**})\geqslant\left(1-\tfrac{1}{4}\xi\right)\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta). \tag{12.51}$$

Note that the definitions of $\mathfrak{F}_\theta$ and $\mathbb{S}$, together with (12.18) and (12.19), imply that for all $L\in\mathfrak{F}_\theta$ we have

$$\mathcal{H}_L^1((\mathbb{S}\setminus(B_\varrho(\widetilde{\gamma}(0))\cup B_r(\widetilde{\gamma}(1))))\cap B_{s(1-r/8)}(\widetilde{\gamma}(0))\cap L)\gtrsim\eta\xi\varrho^2. \tag{12.52}$$

Thus, by the definition of $\mathfrak{F}_\theta^{**}$, and recalling that $s\geqslant\varkappa$, we see that provided that

$$\omega\leqslant c^*\eta\xi\varkappa^3r^2, \tag{12.53}$$

where $c^*>0$ is a small enough absolute constant, we have for all $L\in\mathfrak{F}_\theta^{**}$,

$$\mathcal{H}_L^1(E\cap B_{s(1-r/8)}(\widetilde{\gamma}(0))\cap L)\gtrsim\eta\xi\varrho^2. \tag{12.54}$$

We shall choose $\bar{c}=c^*$ in (12.3), and thus completing our choice of ω.

Assuming also that for a small enough constant $c^{**}>0$ we have

$$\delta_2\leqslant c^{**}\eta\xi^2\varkappa^3r^6\varrho^2, \tag{12.55}$$

which follows from our choice of δ_2 in (12.2), provided c is small enough, we conclude from the definition of $\mathfrak{F}_\theta^*$, together with (12.52), that for all $L\in\mathfrak{F}_\theta^{**}$,

$$\mathrm{NC}_{B_{2C}(\widetilde{\gamma}(0))}(E,L)\leqslant \tfrac{1}{4}\eta\xi\varrho^2 \tag{12.56}$$

and

$$\mathcal{H}^1_L(E\cap B_{s(1-r/8)}(\widetilde{\gamma}(0))\cap L)\geqslant \mathrm{NC}_{B_{2C}(\widetilde{\gamma}(0))}(E,L). \tag{12.57}$$

Recall also assumption (7.12), which implies that

$$\mathcal{L}_3(E^c\cap B_r(\widetilde{\gamma}(1)))\lesssim \delta_1 r^4.$$

Moreover, by the definition of $\mathfrak{F}_\theta$, we have $\mathcal{H}^1_L(L\cap B_r(\widetilde{\gamma}(1)))\gtrsim r$ for all $L\in\mathfrak{F}_\theta$.

For a sufficiently small universal constant $c''''>0$ (see below) define

$$\mathfrak{F}_\theta^{***}=\{L\in\mathfrak{F}_\theta^{**}:\mathcal{H}^1_L(L\cap B_r(\widetilde{\gamma}(1))\cap E)\geqslant c''''r\}. \tag{12.58}$$

By Markov's inequality, we have

$$\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta^{**}\setminus\mathfrak{F}_\theta^{***})\lesssim \frac{\delta_1 r^4}{r}.$$

Therefore, if for a sufficently small universal constant $c^{***}>0$ we have

$$\delta_1\leqslant c^{***}\varkappa^3\xi r, \tag{12.59}$$

it follows that

$$\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta^{***})\geqslant \bigl(1-\tfrac{1}{2}\xi\bigr)\mu_{[\widetilde{\chi}_\theta]}(\mathfrak{F}_\theta). \tag{12.60}$$

Note that our assumption (12.1), together with $\varrho\leqslant\frac{1}{2}\varkappa r^2$, implies that (12.59) holds, provided c is small enough.

For every $L\in\mathfrak{F}_\theta^{***}$ we know that

$$\mathcal{H}^1_L(L\cap B_r(\widetilde{\gamma}(1))\cap E)\geqslant c'''r \overset{(12.56)}{\geqslant} \mathrm{NC}_{B_{2C}(\widetilde{\gamma}(0))}(E,L),$$

provided that

$$\eta\xi\varrho^2\leqslant 4c'''r, \tag{12.61}$$

which holds if c is small enough. In combination with (12.54), (12.56), (12.57), (12.60) and Lemma 11.1, we obtain the required result. □

Appendix A. The sparsest cut problem

As mentioned in §1, Theorem 1.1 and, in particular, Corollary 1.2, leads to an exponential improvement of the previously best known [43] lower bound on the integrality gap of the Goemans–Linial semidefinite relaxation of the sparsest cut problem with general demands. We now add some additional details to the discussion of §1; for further details, we refer also to our paper [24], and the references therein.

The sparsest cut problem with general demands (and capacities) asks for an efficient proceedure to partition a weighted graph into two parts, so as to minimize the interface between them. Formally, we are given an n-vertex graph $G=(V,E)$, with a positive weight (called *capacity*) $c(e)$ associated with each edge $e\in E$, and a non-negative weight (called *demand*) $D(u,v)$ associated with 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

$$\Phi^*(c,D)=\min_{\varnothing\neq S\subsetneq V}\frac{\sum_{uv\in E}c(uv)|\chi_S(u)-\chi_S(v)|}{\sum_{u,v\in V}D(u,v)|\chi_S(u)-\chi_S(v)|}. \tag{A.1}$$

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

$$\Phi^*=\min_{\varnothing\neq S\subsetneq V}\frac{\#\{\text{edges joining } S \text{ and } V\setminus S\}}{|S|\,|V\setminus S|}. \tag{A.2}$$

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.

From now on, the sparsest cut problem will be understood to be with general capacities and demands. These allow one 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 one of the most important problems in the field of approximation algorithms. It is used as a subroutine in many approximation algorithms for NP-hard problems; see the survey article [60], as well as the references in [49], [6], [5] and [24] for some of the (vast) literature on this topic.

The problem of computing $\Phi^*(c,D)$ in polynomial time is known to be NP-hard [59]. The most fruitful approach to finding an approximate solution has been to consider relaxations of the problem. Observe that the term

$$d_S:=|\chi_S(u)-\chi_S(v)|$$

occurring in (A.1) is just the distance from u to v in the elementary cut metric associated with S; see §2. One is therefore led to consider the minimization of the functional

$$\Phi(c,D,d)=\frac{\sum_{uv\in E}c(uv)d(u,v)}{\sum_{u,v\in V}D(u,v)d(u,v)}, \tag{A.3}$$

where d varies over an enlarged class of metrics. The following successively larger classes of metrics have played a key role: elementary cut metrics (as in (A.1)), L_1 metrics (equivalently, cut metrics), metrics of negative type, arbitrary metrics. Denote the last two collections of metrics by NEG and MET, respectively, and denote the corresponding minima by

$$\Phi^*(c,D)\geqslant\Phi^*_{L_1}(c,D)\geqslant\Phi^*_{\mathrm{NEG}}(c,D)\geqslant\Phi^*_{\mathrm{MET}}(c,D).$$

From the standpoint of theoretical computer science, a key point is that $\Phi^*_{\mathrm{MET}}(c,D)$ is a linear program (since the triangle inequality is a linear condition) and $\Phi^*_{\mathrm{NEG}}(c,D)$ can be computed by the ellipsoid algorithm. Thus, both are computable in polynomial time with arbitrarily good precision.

Let d denote an L_1 metric and let Σ_d denote the cut measure occuring in the cut metric representation

$$d=\sum_{S\subset V}\Sigma_d(S)d_S$$

of d; see (2.5). By substituting the cut metric representation into (A.3), it follows directly that in fact $\Phi^*(c,D)=\Phi^*_{L_1}(c,D)$ for all c and D; [10], [53], [9]. The key point is that L_1 metrics are the convex cone generated by elementary cut metrics (and a corresponding statement would hold for any such convex cone and its generators).

If d_1 and d_2 are any two metrics on V, define their distortion by

$$\mathrm{dist}(d_1,d_2)=\left(\max_{u,v\in V}\frac{d_1(u,v)}{d_2(u,v)}\right)\left(\max_{u,v\in V}\frac{d_2(u,v)}{d_1(u,v)}\right).$$

It follows immediately that for all c and D,

$$\max\left\{\frac{\Phi(c,D,d_1)}{\Phi(c,D,d_2)},\frac{\Phi(c,D,d_2)}{\Phi(c,D,d_1)}\right\}\leqslant\mathrm{dist}(d_1,d_2). \tag{A.4}$$

Thus,

$$\sup_{c,D}\frac{\Phi^*_{L_1}(c,D)}{\Phi^*_{\mathrm{NEG}}(c,D)}\leqslant\sup_{d\in\mathrm{NEG}}c_1(V,d), \tag{A.5}$$

$$\sup_{c,D}\frac{\Phi^*_{L_1}(c,D)}{\Phi^*_{\mathrm{MET}}(c,D)}\leqslant\sup_{d\in\mathrm{MET}}c_1(V,d). \tag{A.6}$$

Recall that by definition, the left-hand sides in (A.5) and (A.6) are the integrality gaps for the corresponding relaxations. Since $\Phi^*(c,D)=\Phi^*_{L_1}(c,D)$, Bourgain's embedding theorem implies that the first of these integrality gaps is $\lesssim \log n$; [53], [9]. Similarly, for the Goemans–Linial semidefinite relaxation, the embedding theorem for metrics of negative type given in [5] implies that the integrality gap is $\lesssim (\log n)^{1/2+o(1)}$. (Of course, the upper bound in terms of distortion also applies if L_1 metrics are replaced by the original elementary cut metrics, but since these metrics are so highly degenerate, it does not provide useful information. This illustrates the power of the observation of [53] and [9].)

From now on we restrict our attention to the Goemans–Linial semidefinite relaxation. (What we say also applies mutadis mutandis to the relaxation to MET.) As in (A.5), for all c and D, and all $d_2 \in \mathrm{NEG}$,

$$\frac{\Phi^*_{L_1}(c,D)}{\Phi^*_{\mathrm{NEG}}(c,D)} \leqslant c_1(V,d_2). \tag{A.7}$$

A duality argument (sketched below) shows that for all d_2 there exist c and D for which (A.7) becomes an equality. Therefore, the integrality gap is actually *equal* to

$$\sup_{d\in\mathrm{NEG}} c_1(V,d).$$

Thus, by Corollary 1.2 (see also Remark 1.3) the integrality gap is $\gtrsim (\log n)^\delta$ for some explicit $\delta>0$. Recall that the previous best bound was $\gtrsim \log\log n$; see [43] and [41].

Here is a sketch of the duality argument; for further discussion, see [54, Proposition 15.5.2]. Consider a set V of cardinality n, the vertices of our graph (whose edge structure will be determined below). Define an embedding F of the metrics on V into $\mathbb{R}^{n(n-1)/2}$ as follows: the coordinates $x_{u,v}$ correspond to unordered pairs u,v of distinct vertices in V, and $x_{u,v}(F(d))=d(u,v)$. The image of the L_1 metrics on V is a convex cone $\mathcal{L}\subset\mathbb{R}^{n(n-1)/2}$ generated by the images of the elementary cut metrics d_S. Let d_2 satisfy $F(d_2)\notin\mathcal{L}$. By an easy compactness argument, the distortion $\mathrm{dist}(d,d_2)$, for d with $F(d)\in\mathcal{L}$, is minimized by some d_1, with $F(d_1)\in\mathcal{L}$. Minimality implies that in fact $F(d_1)\in\partial\mathcal{L}$. Take a supporting hyperplane P for $\mathcal{L}$ which passes through $F(d_1)$. Let ℓ denote a linear functional satisfying $\ell|_P\equiv 0$, $\ell|_{\mathcal{L}}\leqslant 0$ and $\ell(F(d_2))>0$. Let

$$\sum_{(u,v)} \ell_{u,v} x^*_{(u,v)}$$

denote the coordinate representation of ℓ. Define capacities and demands by

$$c(uv)=\begin{cases}\ell_{u,v}, & \text{if } \ell_{u,v}>0,\\ 0, & \text{otherwise}\end{cases} \quad\text{and}\quad D(u,v)=\begin{cases}-\ell_{u,v}, & \text{if } \ell_{u,v}<0,\\ 0, & \text{otherwise.}\end{cases}$$

Define a graph structure with vertices V, by stipulating that $uv\in E$ if and only if $c(uv)>0$. It is trivial to check that for the above c and D we have $\Phi(c,D,d)\geqslant 1$ for all d with $F(d)\in\mathcal{L}$. Also, $\Phi(d_1)=1$ and

$$\Phi(c,D,d_2)\leqslant\frac{1}{\operatorname{dist}(d_1,d_2)}=\frac{1}{c_1(V,d_2)}.$$

These relations imply that for these c and D, (A.7) is an equality. By choosing $d_2\in\mathrm{NEG}$ such that $c_1(V,d_2)$ is maximal, it follows that (A.5) is also an equality. (If $F(d_2)\in\mathcal{L}$, this is trivial.)

Appendix B. Quantitative bounds, coercivity and monotonicity

In this appendix, we briefly discuss from a more general standpoint, the structure of our argument, as outlined after the statement of Theorem 1.1 and in §4. We point out that, in essence, what we have done follows the general scheme of other arguments in geometric analysis and non-linear partial differential equations; compare Example B.1 below. Typically, the results are not stated explicitly in quantitative form. Here, we wish to emphasize that the possibility of an estimate of the form of (1.1) is actually implicit in the arguments. In fact, more is true. For any $\varepsilon>0$, there is an explicit bound $c(\varepsilon)<\infty$ on the Carleson measure (in the sense of Semmes) of the set of "ε-bad" balls $B_r(x)\subset B_1(p)$. Note that the Carleson measure $r^{-1}\,dr\times\mathcal{L}$ of the set of *all* $B_r(x)\subset B_1(p)$ is infinite, since

$$\int_0^1\frac{dr}{r}=\infty;$$

for further discussion, see the appendix by Semmes in [33].

The crucial ingredient for obtaining estimates as in (1.1), as well as the more precise Carleson measure estimates, is the existence of a quantity which is coercive, monotone and bounded. The sense in which these terms are to be understood is explained below.

B.1. Coercivity and almost rigidity

The term *rigid* connotes special (i.e. highly constrained) structure. A standard feature of (the statment and proof of) rigidity theorems is the existence of a numerical measurement $Q\geqslant 0$ which is *coercive* in the sense that if $Q=0$, then the desired rigidity holds, and more generally (and often much harder to prove) if $Q<\varepsilon^a$, for some $a<\infty$, then in a suitable sense the structure is ε-close to the one which is obtained in the rigid case. Statments of this type are known as *stability theorems*, *ε-regularity theorems* or *almost* *rigidity theorems*. A classical example from Riemannian geometry is the *sphere theorem*, in which the coercive quantity is minus the logarithm of the pinching; see [13] and the references therein.

In our case, we are given $E\subset\mathbb{H}$, and the coercive quantity is

$$Q(E)=\mathrm{NM}_{B_r(x)}(E),$$

the non-monotonicity of E on $B_r(x)$. Coercivity is the statement that monotone subsets are half-spaces, or more generally, that almost monontone subsets are close to half-spaces; see Theorem 4.3.

B.2. Bounded monotone quantities and existence of a good scale

As in §1 and §2, we point out the general character of the estimate in Proposition 4.6 for the scale on which Theorem 4.3 can be applied. Namely, by Markov's inequality (3.14) (which in this case amounts to the pigeonhole principle), such an estimate for the scale will appear *whenever* we are dealing with an a-priori bounded non-negative quantity, which can be written as a sum of non-negative terms, which correspond to the various scales, such that each term is coercive on its own scale (in the suitably scaled sense). Such a quantity is *monotone* in the sense that the sum is non-decreasing as we include more and more scales; compare (4.28) and (6.11).

Quantities which are coercive and monotone are well known to play a key role in geometric analysis and in partial differential equations. In the latter case, for evolution equations, monotonicity is defined with respect to the time parameter, rather than the scale.

We include below the following illustrative example, which requires familiarity with Riemannian geometry. Numerous other choices from diverse areas would serve equally well; compare Remark 2.1.

Example B.1. A theorem from Riemannian geometry states that for non-collapsed Gromov–Hausdorff limit spaces

$$M_i^n \xrightarrow{d_{GH}} Y^n,$$

such that

$$\operatorname*{Ric}_{M_i^n} \geqslant -(n-1)$$

for all i, every tangent cone Y_y is a metric cone; see [16] and [17, Remark 4.99]. In this case, the rigid objects are metric cones and the relevant coercive quantity Q is derived from the volume ratio,

$$\frac{\mathrm{Vol}(B_r(p))}{\mathrm{Vol}(B_r(\underline{p}))}, \tag{B.1}$$

where $\underline{p} \in \underline{M}^n$ and $\underline{M}^n$ denotes the hyperbolic n-space with curvature identically -1. The coercivity of Q is guaranteed by the fact that "volume cone implies metric cone" and its corresponding almost rigidity theorem, which states that almost volume cones are close in the Gromov–Hausdorff sense to being almost metric cones; see [16] and [17]. The monotonicity of Q is a consequence of the Bishop–Gromov inequality, which asserts that the volume ratio in (B.1) is a monotone non-increasing function of r.

References

[1] Ambrosio, L., Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces. *Adv. Math.*, 159 (2001), 51–67.

[2] — Fine properties of sets of finite perimeter in doubling metric measure spaces. *Set-Valued Anal.*, 10 (2002), 111–128.

[3] Ambrosio, L., Fusco, N. & Pallara, D., *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, New York, 2000.

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

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

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

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

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

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

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

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

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

[13] Brendle, S. & Schoen, R., Manifolds with 1/4-pinched curvature are space forms. *J. Amer. Math. Soc.*, 22 (2009), 287–307.

[14] Burago, D., Burago, Y. & Ivanov, S., *A Course in Metric Geometry.* Graduate Studies in Mathematics, 33. Amer. Math. Soc., Providence, RI, 2001.

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

[16] Cheeger, J. & Colding, T. H., Lower bounds on Ricci curvature and the almost rigidity of warped products. *Ann. of Math.*, 144 (1996), 189–237.

[17] — On the structure of spaces with Ricci curvature bounded below. I. *J. Differential Geom.*, 46 (1997), 406–480.

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

[19] — Differentiability of Lipschitz maps from metric measure spaces to Banach spaces with the Radon–Nikodým property. *Geom. Funct. Anal.*, 19 (2009), 1017–1028.

[20] — Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math.*, 171 (2010), 1347–1385.

[21] — Metric differentiation, monotonicity and maps to L^1. *Invent. Math.*, 182 (2010), 335–370.

[22] — Realization of metric spaces as inverse limits, and bilipschitz embedding in L_1. Preprint, 2011. `arXiv:1110.2406 [math.MG]`.

[23] — Metric differentiation for PI spaces. In preparation.

[24] Cheeger, J., Kleiner, B. & Naor, A., 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), pp. 555–564. IEEE Computer Society, Los Alamitos, CA, 2009.

[25] De Giorgi, E., Su una teoria generale della misura $(r-1)$-dimensionale in uno spazio ad r dimensioni. *Ann. Mat. Pura Appl.*, 36 (1954), 191–213.

[26] — Nuovi teoremi relativi alle misure $(r-1)$-dimensionali in uno spazio ad r dimensioni. *Ricerche Mat.*, 4 (1955), 95–113.

[27] Deza, M. M. & Laurent, M., *Geometry of Cuts and Metrics*. Algorithms and Combinatorics, 15. Springer, Berlin–Heidelberg, 1997.

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

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

[30] — On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.*, 13 (2003), 421–466.

[31] Goemans, M. X., Semidefinite programming in combinatorial optimization. *Math. Program.*, 79 (1997), 143–161.

[32] Gromov, M., Asymptotic invariants of infinite groups, in *Geometric Group Theory*, Vol. 2 (Sussex, 1991), London Math. Soc. Lecture Note Ser., 182, pp. 1–295. Cambridge University Press, Cambridge, 1993.

[33] — *Metric Structures for Riemannian and Non-Riemannian Spaces*. Modern Birkhäuser Classics. Birkhäuser, Boston, MA, 2007. Based on the 1981 French original.

[34] Grötschel, M., Lovász, L. & Schrijver, A., *Geometric Algorithms and Combinatorial Optimization*. Algorithms and Combinatorics, 2. Springer, Berlin–Heidelberg, 1993.

[35] Gupta, A., Krauthgamer, R. & Lee, J. R., Bounded geometries, fractals, and low-distortion embeddings, in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science* (FOCS 2003), pp. 534–543. IEEE Computer Society, Los Alamitos, CA, 2003.

[36] Hajłasz, P. & Koskela, P., Sobolev met Poincaré. *Mem. Amer. Math. Soc.*, 145:688 (2000).

[37] Heinonen, J., *Lectures on Analysis on Metric Spaces*. Universitext. Springer, New York, 2001.

[38] Heinonen, J. & Koskela, P., From local to global in quasiconformal structures. *Proc. Nat. Acad. Sci. U.S.A.*, 93 (1996), 554–556.

[39] Johnson, W. B. & Lindenstrauss, J., Extensions of Lipschitz mappings into a Hilbert space, in *Conference in Modern Analysis and Probability* (New Haven, CT, 1982), Contemp. Math., 26, pp. 189–206. Amer. Math. Soc., Providence, RI, 1984.

[40] Jones, P. W., Lipschitz and bi-Lipschitz functions. *Rev. Mat. Iberoam.*, 4 (1988), 115–121.

[41] Khot, S. & Vishnoi, N., 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), pp. 53–62. IEEE Computer Society, Los Alamitos, CA, 2005.

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

[43] Krauthgamer, R. & Rabani, Y., Improved lower bounds for embeddings into L_1, in *Proceedings of the 17th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 1010–1017. ACM, New York, 2006.

[44] Kuratowski, K. & Ryll-Nardzewski, C., A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 13 (1965), 397–403.

[45] Laakso, T. J., Ahlfors Q-regular spaces with arbitrary $Q>1$ admitting weak Poincaré inequality. *Geom. Funct. Anal.*, 10 (2000), 111–123.

[46] Lang, U. & Plaut, C., Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87 (2001), 285–307.

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

[48] — 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), pp. 99–108. IEEE Computer Society, Los Alamitos, CA, 2006.

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

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

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

[52] — Squared ℓ_2 metrics into ℓ_1, in *Open Problems on Embeddings of Finite Metric Spaces*. 2002. Available at `http://kam.mff.cuni.cz/~matousek/metrop.ps`.

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

[54] Matoušek, J., *Lectures on Discrete Geometry*. Graduate Texts in Mathematics, 212. Springer, New York, 2002.

[55] Montefalcone, F., Some relations among volume, intrinsic perimeter and one-dimensional restrictions of BV functions in Carnot groups. *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, 4 (2005), 79–128.

[56] Montgomery, R., *A Tour of Subriemannian Geometries, their Geodesics and Applications*. Mathematical Surveys and Monographs, 91. Amer. Math. Soc., Providence, RI, 2002.

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

[58] Rao, S., Small distortion and volume preserving embeddings for planar and Euclidean metrics, in *Proceedings of the 15th Annual Symposium on Computational Geometry* (Miami Beach, FL, 1999), pp. 300–306. ACM, New York, 1999.

[59] Shahrokhi, F. & Matula, D. W., The maximum concurrent flow problem. *J. ACM*, 37 (1990), 318–334.

[60] Shmoys, D. B., Cut problems and their application to divide-and-conquer, in *Approximation Algorithms for NP-hard Problems*, pp. 192–235. PWS, Boston, MA, 1997.

[61] Stein, E. M., *Harmonic Analysis*: *Real-Variable Methods*, *Orthogonality*, *and Oscillatory Integrals*. Princeton Mathematical Series, 43. Princeton University Press, Princeton, NJ, 1993.

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

[63] Wells, J. H. & Williams, L. R., *Embeddings and Extensions in Analysis*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 84. Springer, New York, 1975.

[64] Wojtaszczyk, P., *Banach Spaces for Analysts*. Cambridge Studies in Advanced Mathematics, 25. Cambridge University Press, Cambridge, 1991.

Jeff Cheeger
Courant Institute
New York University
251 Mercer Street
New York, NY 10012
U.S.A.
cheeger@cims.nyu.edu

Bruce Kleiner
Courant Institute
New York University
251 Mercer Street
New York, NY 10012
U.S.A.
bkleiner@cims.nyu.edu

Assaf Naor
Courant Institute
New York University
251 Mercer Street
New York, NY 10012
U.S.A.
naor@cims.nyu.edu

Received November 24, 2009