An Isoperimetric Inequality for the Heisenberg Groups

Daniel Allcock*
July 21, 1995; revised 5 February 1997
allcock@math.utah.edu
web page: *http://www.math.utah.edu/~allcock*
Deptartment of Mathematics
University of Utah
Salt Lake City, UT 84112

1991 mathematics subject classification: 20F32 (53C20, 22E25)

Abstract.
We show that the Heisenberg groups $\mathcal{H}^{2n+1}$ of dimension five and higher, considered as Riemannian manifolds, satisfy a quadratic isoperimetric inequality. (This means that each loop of length L bounds a disk of area $\sim L^2$). This implies several important results about isoperimetric inequalities for discrete groups that act either on $\mathcal{H}^{2n+1}$ or on complex hyperbolic space, and provides interesting examples in geometric group theory. The proof consists of explicit construction of a disk spanning each loop in $\mathcal{H}^{2n+1}$.

1 Introduction

The Heisenberg groups $\mathcal{H}^3, \mathcal{H}^5, \mathcal{H}^7, \ldots$ are a sequence of nilpotent Lie groups that arise in geometry in several ways. For example, $\mathcal{H}^3$ is known to three-dimensional geometers as *Nilgeometry*, and arises in the study of Seifert-fibered three-manifolds [14]. The $\mathcal{H}^{2n+1}$ also appear in hyperbolic geometry: a horosphere in complex hyperbolic space $\mathbb{C}H^n$ is a copy of $\mathcal{H}^{2n-1}$ (see [7], [11]). Thurston [3], without proof, and Gromov [9][10], outlining a proof, have stated that $\mathcal{H}^5, \mathcal{H}^7, \ldots$ satisfy quadratic isoperimetric inequalities (defined below). This paper provides a new and complete proof of this theorem, by explicitly exhibiting spanning disks and estimating their areas. Our main purpose in proving this theorem is to obtain isoperimetric inequalities for certain finitely presented groups, in particular, for the discrete Heisenberg groups and for nonuniform lattices in $\mathrm{Isom}\,(\mathbb{C}H^n)$. In a sense, we use elementary differential geometry to obtain results in the geometric theory of discrete groups. Our inequalities for $\mathcal{H}^{2n+1}$ $(n > 1)$ contrast with the case of $\mathcal{H}^3$, which satisfies a cubic (but no quadratic) isoperimetric inequality; see [3]. For more information about isoperimetric inequalities for finitely presented groups, see [5], [6].

A Riemannian manifold M is said to satisfy the isoperimetric inequality f, where f is a function from the positive real numbers to themselves, if any smooth closed curve in M of length at most L bounds a disk in M of area at most $f(L)$. One says that M satisfies a quadratic isoperimetric inequality if f may be taken to be a quadratic polynomial, and one makes similar statements for cubic bounds of f, etc.

It turns out that our problem may be reduced to a problem in symplectic geometry which is interesting in its own right. Namely, we consider loops in $\mathbb{R}^{2n}$, which we equip with both its usual Euclidean metric and its usual symplectic form. The problem of finding spanning disks of appropriate area in $\mathcal{H}^{2n+1}$ reduces to the following problem in $\mathbb{R}^{2n}$: given a loop γ in $\mathbb{R}^{2n}$, enclosing zero symplectic area and having length L, does γ bound an isotropic disk with Euclidean area of order L^2? The answer is yes if $n > 1$, as we prove in section 2.

Section 3 introduces the Heisenberg groups and explains the connections between them and the symplectic geometry, and section 4 obtains the advertised quadratic isoperimetric inequality for

* Supported by a National Science Foundation Graduate Fellowship.

$\mathcal{H}^{2n+1}$. Finally, section 5 offers applications of our results to complex and quaternionic hyperbolic geometry and to geometric group theory; it also compares our techniques with those of Gromov [10] and Lee [12].

The proofs are arranged so that a reader interested only in the existence of isotropic spanning disks (with or without the area estimate) need read only a minimal amount.

The author is pleased to S. Gersten and J. Burillo for working through the paper in their seminar and for suggesting improvements. This paper is derived from part of the author's Ph.D. thesis at U.C. Berkeley and the author would like to thank his advisor, A. Casson, for his suggestions on the exposition.

2 Isotropic Disks in $\mathbb{R}^{2n}$

We consider $\mathbb{R}^{2n}$ as both a symplectic and Euclidean space, with the orthonormal basis x_1, y_1, ..., x_n,y_n, and the standard symplectic form ω, defined by

$$\omega(x_i, y_i) = -\omega(y_i, x_i) = 1,$$

with all other pairings of basis vectors vanishing. (Recall that a symplectic form is a nondegenerate antisymmetric bilinear form on a vector space.) As usual, we also consider $\mathbb{R}^{2n}$ as a manifold and ω as a differential 2-form thereon. If α is a smooth closed path $\alpha : [0, 1] \to \mathbb{R}^{2n}$, we may choose a smooth disk D spanning α and evaulate $\int_D \omega$. We call this quantity the symplectic area enclosed by α; since ω is exact, this area does not depend on the disk chosen.

We say that a smooth map from a manifold, $f : M \to \mathbb{R}^{2n}$, is isotropic if $f^*\omega = 0$; in the following constructions M will be the unit disk or a rectangle. Stokes' theorem implies that two closed loops joined by an isotropic homotopy enclose the same symplectic area.

The goal of this section is to prove theorem 2.3 below, and the strategy is to first prove two lemmas that provide us with 'legal moves' on loops in $\mathbb{R}^{2n}$. That is, they provide a variety of isotropic homotopies with small Euclidean areas. We will prove theorem 2.3 by weaving these moves together.

Lemma 2.1. *Suppose $\mathbb{R}^{2n} = V_1 \oplus V_2$, where V_1 and V_2 are symplectically orthogonal subspaces, and that $\alpha : [0, 1] \to \mathbb{R}^{2n}$ is a smooth loop based at the origin, with $\alpha(s) = \alpha_1(s) + \alpha_2(s)$ for all s, where the image of α_i lies in V_i. Then there is a smooth isotropic homotopy between α and the loop obtained by first traversing α_1 and then α_2.*

If V_1 and V_2 are orthogonal (with respect to the Euclidean metric) and α has length L, then the isotropic homotopy may be taken to have (Euclidean) area at most L^2.

Proof: By reparameterizing and extending its domain, we may (and do) suppose that α is a smooth map from $\mathbb{R}$ to $\mathbb{R}^{2n}$ which vanishes except on $[0, 1]$. Consider the homotopy

$$\Gamma(s, t) = \alpha_1(s + t) + \alpha_2(s),$$

where $t \in [0, 1]$ and $s \in [-1, 1]$. $\Gamma(s, 0)$ is the restriction of α to $[-1, 1]$, and $\Gamma(s, 1)$ is the loop obtained by traversing first α_1 and then α_2. We compute

$$\partial\Gamma/\partial t = \alpha_1'(s + t)$$
$$\partial\Gamma/\partial s = \alpha_1'(s + t) + \alpha_2'(s);$$

Γ is isotropic: $\omega(\partial\Gamma/\partial s, \partial\Gamma/\partial t) = 0$ because ω is antisymmetric and because the α_i' lie in symplectically orthogonal subspaces. We estimate the area of the homotopy as follows. Writing X for

$[0,1] \times [-1,1]$, we have

$$\begin{aligned}
\text{Area}(\Gamma(X)) &= \int_X ds\, dt \sqrt{\|\partial\Gamma/\partial t\|^2 \, \|\partial\Gamma/\partial s\|^2 - ((\partial\Gamma/\partial s)\cdot(\partial\Gamma/\partial t))^2} \\
&= \int_X ds\, dt \sqrt{\|\alpha_1'(s+t)\|^2 \left[\|\alpha_1'(s+t)\|^2 + \|\alpha_2'(s)\|^2\right] - \|\alpha_1'(s+t)\|^4} \\
&= \int_X \|\alpha_1'(s+t)\| \, \|\alpha_2'(s)\| ds\, dt \\
&= \int_0^1 \|\alpha_2'(s)\| dt \int_{-1}^1 \|\alpha_1'(s+t)\| ds \le L^2.
\end{aligned}$$

In the last step the integrals are bounded by the lengths of α_1 and α_2, respectively, which are in turn bounded by L. □

Lemma 2.2. *Let α and β be two smooth paths in $\mathbb{R}^{2n}$. If their images lie in symplectically orthogonal subspaces and they are parameterized such that*

$$\omega(\alpha(s), \alpha'(s)) = \omega(\beta(s), \beta'(s)), \tag{2.1}$$

then there is a smooth isotropic homotopy between them. If α and β are closed paths, then this may be taken to be a homotopy through closed paths.

If α has length L and there are constants A and B such that for all s we have $\|\beta'(s)\| \le \|\alpha'(s)\|$, $\|\alpha(s)\| \le A$ and $\|\beta(s)\| \le B$, then the homotopy may be taken to have (Euclidean) area at most $(A+B)\pi L$.

Remark: The quantity on the left side of (2.1) may be interepreted geometrically as (twice) the rate at which the family of segments joining 0 and $\alpha(s)$ sweeps out symplectic area. If α lies in a symplectic plane then it may also be interpreted as the angular momentum (about the origin) of a particle moving along α. Similar interpretations in terms of β describe the right hand side.

Proof: Define the homotopy

$$\Gamma(s,t) = \alpha(s)\cos t + \beta(s)\sin t,$$

where $t \in [0, \pi/2]$. It is obvious that if α and β are loops then Γ is a homotopy through loops. We have

$$\begin{aligned}
\partial\Gamma/\partial t &= -\alpha(s)\sin t + \beta(s)\cos t \\
\partial\Gamma/\partial s &= \alpha'(s)\cos t + \beta'(s)\sin t.
\end{aligned}$$

To evaluate $\Gamma^*\omega$, one computes $\omega(\partial\Gamma/\partial s, \partial\Gamma/\partial t)$, which vanishes, establishing that Γ is isotropic. To estimate the area of the homotopy under the conditions of the last claim of the theorem, we observe

$$\begin{aligned}
\|\partial\Gamma/\partial t\| &\le \|\alpha(s)\| + \|\beta(s)\| \le A + B \\
\|\partial\Gamma/\partial s\| &\le \|\alpha'(s)\| + \|\beta'(s)\| \le 2\|\alpha'(s)\|.
\end{aligned}$$

Writing X for $[0,1] \times [0, \pi/2]$, we have

$$\begin{aligned}
\text{Area}(\Gamma(X)) &\le \int_X \|\partial\Gamma/\partial t\| \, \|\partial\Gamma/\partial s\| ds\, dt \\
&\le \int_X 2(A+B)\|\alpha'(s)\| ds\, dt \\
&= (A+B)\pi L.
\end{aligned}$$

□

Theorem 2.3. *Let $\gamma : S^1 \to \mathbb{R}^{2n}$ be a smooth loop enclosing zero symplectic area, and suppose $n > 1$. Then there is a smooth isotropic map of the unit disk $f : D \to \mathbb{R}^{2n}$ such that $f|\partial D = \gamma$.*

If γ has length L, then the spanning disk $f(D)$ may be chosen to have Euclidean area at most kL^2, where k is a constant.

Remark: The proof below shows that k may be taken to be $1 + \pi(2 + \sqrt{5})$.

Proof: The proof proceeds by applying a sequence of isotropic homotopies to γ; their composition shrinks γ to a point.

STEP 1: Consider γ to be a closed path $\gamma = \gamma_0 : I \to \mathbb{R}^{2n}$, and suppose that $\gamma_0(0) = 0$. By lemma 2.1, γ_0 is homotopic by a smooth isotropic homotopy to a smooth curve γ_1 which is a composition of two loops, the first being the projection of γ_0 to the x_1, y_1 plane and the second the projection of γ_0 to the span of $x_2, y_2, \ldots, x_n, y_n$.

STEP 2: If $\alpha(s) = (x(s), y(s), 0, 0)$ is a smooth curve in $\mathbb{R}^4$ then define $\beta(s) = (0, 0, x(s), y(s))$, and apply lemma 2.2 to deduce that there is an isotropic homotopy carrying α to β. Applying this to the first of the two loops comprising γ_1 produces an isotropic homotopy between γ_1 and a loop γ_2 lying entirely in the span of $x_2, y_2, \ldots, x_n, y_n$.

STEP 3: Define $\gamma_3(s) = (L/2, h(s), 0, \ldots, 0)$, where $h(s)$ is defined by the conditions $h(0) = 0$ and

$$\omega(\gamma_3(s), \gamma_3'(s)) = \omega(\gamma_2(s), \gamma_2'(s)).$$

The latter condition is equivalent to $h'(s) = 2\omega(\gamma_2(s), \gamma_2'(s))/L$. Since the symplectic area enclosed by γ_2 is zero, we have $h(1) = 0$, so γ_3 is a closed path. Applying lemma 2.2 to γ_2 and γ_3 yields an isotropic homotopy between them.

STEP 4: Steps 1 through 3 have provided an isotropic homotopy between our given loop $\gamma_0 = \gamma$ and the loop γ_3, which lies in a line of $\mathbb{R}^{2n}$. As our last step, we contract the loop γ_3 to a point in this one-dimensional subspace by a linear (and obviously isotropic) homotopy. By changing the parametrization of the homotopy parameter s, we may paste the four homotopies together to obtain a smooth isotropic disk $f : D \to \mathbb{R}^{2n}$ bounding γ. This completes the proof of the first part of the theorem.

To prove the second part, we must estimate the areas of the homotopies used above. By lemma 2.1, the homotopy in step one has area at most L^2, and we observe that γ_1 consists of two loops, each of length at most L. This implies that $\|\gamma_1(s)\| \leq L$ for all s. Lemma 2.2 then proves that the second homotopy has area at most πL^2. We will also apply lemma 2.2 to the homotopy of step three. Since $\|\gamma_2(s)\| \leq L/2$ for all s, we have $|h'(s)| \leq \|\gamma_2'(s)\|$, which implies that the length of γ_3 is at most $2L$, and hence that $|h(s)| \leq L$ and $\|\gamma_3(t)\| \leq L\sqrt{5}/2$. Since the length of γ_2 is at most $2L$, lemma 2.2 shows that the area of this homotopy is at most $\pi L^2(1 + \sqrt{5})$. Observing that the fourth homotopy has zero area, we add all these areas together to obtain the advertised bound. □

3 The Heisenberg Groups

For background on the Heisenberg groups, see [7]. The Heisenberg group $\mathcal{H}^{2n+1}$ is the connected and simply connected Lie group with Lie algebra $\mathfrak{h}^{2n+1}$, which has a basis consisting of the $2n+1$ vectors $x_1, y_1, \ldots, x_n, y_n, z$, and Lie bracket defined by the relations that $[x_i, y_i] = z$ and that all other pairings of basis elements vanish. The center $\mathfrak{z}$ of $\mathfrak{h}^{2n+1}$ is the span of z, and we denote by $\mathcal{Z}$ the center of the group $\mathcal{H}^{2n+1}$. Let $\pi : \mathcal{H}^{2n+1} \to \mathcal{H}^{2n+1}/\mathcal{Z} \cong \mathbb{R}^{2n}$ be the canonical projection map.

We may equip $\mathcal{H}^{2n+1}$ with a left-invariant Riemannian metric g by declaring the $2n+1$ basis vectors given above to be an orthonormal basis for $\mathfrak{h}^{2n+1}$, and translating this inner product to other points of the group by (left) multiplication. This metric is not canonical, but in this investigation, the choice of metric does not matter. When $n > 1$, theorem 4.1 below provides a quadratic isoperimetric inequality for $\mathcal{H}^{2n+1}$ equipped with the metric g. Since any other (left invariant) metric on $\mathcal{H}^{2n+1}$ disagrees with g about lengths and areas by some uniformly bounded amount, we will be able to conclude that $\mathcal{H}^{2n+1}$ satisfies a quadratic isoperimetric inequality when equipped with *any* (left-invariant) Riemannian metric.

Two important notions are the *vertical* and *horizontal* subspaces at a point $x \in \mathcal{H}^{2n+1}$. The vertical subspace V_x is the set of vectors in $T_x\mathcal{H}^{2n+1}$ that are tangent to the coset of $\mathcal{Z}$ passing through x. (Note that the left and right cosets of $\mathcal{Z}$ coincide.) V_x is one-dimensional. We define the horizontal subspace H_x to be the orthogonal complement of V_x in $T_x\mathcal{H}^{2n+1}$; the distribution of $2n$-dimensional horizontal subspaces is invariant under the action of (left) group multiplication. We may use the horizontal subspaces to equip the central quotient $\mathbb{R}^{2n}$ of $\mathcal{H}^{2n+1}$ with a Riemannian metric: if $x \in \mathbb{R}^{2n}$, then choose any representative $\tilde{x}$ for x in $\mathcal{H}^{2n+1}$, and observe that the projection π establishes an isomorphism between $H_{\tilde{x}}$ and $T_x\mathbb{R}^{2n}$. Define the inner product on $T_x\mathbb{R}^{2n}$ via this identification; this is independent of the choice of $\tilde{x}$, and yields a 'quotient' metric on $\mathbb{R}^{2n}$. It is not hard to see that this is the standard Euclidean metric. We say that a vector in $T_x\mathcal{H}^{2n+1}$ is horizontal (resp. vertical) if it lies in H_x (resp. V_x). If M is a manifold and $f : M \to \mathcal{H}^{2n+1}$ is smooth, then we say that f and $f(M)$ are horizontal if every element of f_*TM is horizontal. (In our applications, M will be an interval or a surface.)

We may define a symplectic form ω on the vector space $H_1 \subseteq \mathfrak{h}^{2n+1}$ by identifying $\mathfrak{z}$ with $\mathbb{R}$ (by identifying z with $1 \in \mathbb{R}$), and setting $\omega(v, w) = [v, w] \in \mathfrak{z} \cong \mathbb{R}$, for $v, w \in H_1$. We may translate this structure to other points of $\mathcal{H}^{2n+1}$ by (left) group multiplication, and then follow the procedure above to equip the quotient $\mathbb{R}^{2n}$ with a symplectic structure, which we also call ω. It is easy to see that ω is the standard symplectic form relative to the basis $x_1, y_1, \ldots, x_n, y_n$.

When equipped with both its Euclidean and symplectic structures, $\mathbb{R}^{2n}$ encodes much information about $\mathcal{H}^{2n+1}$, and the family of horizontal subspaces allows us to 'lift' objects in $\mathbb{R}^{2n}$ to objects in $\mathcal{H}^{2n+1}$. Namely, if $\alpha : [0, 1] \to \mathbb{R}^{2n}$ is a smooth curve and we select a point $x \in \pi^{-1}(\alpha(0))$, then there is a unique horizontal lift $\tilde{\alpha} : [0, 1] \to \mathcal{H}^{2n+1}$ having the properties that $\tilde{\alpha}(0) = x$ and that $\pi \circ \tilde{\alpha}(t) = \alpha(t)$ for all $t \in [0, 1]$. If α is a closed loop in $\mathbb{R}^{2n}$, then $\tilde{\alpha}$ typically fails to be a closed loop in $\mathcal{H}^{2n+1}$; the symplectic structure on $\mathbb{R}^{2n}$ tells us the amount by which it fails to close. The vertical distance between $\tilde{\alpha}(0)$ and $\tilde{\alpha}(1)$ is equal to the absolute value of the symplectic area enclosed by the loop α in $\mathbb{R}^{2n}$, and the sign of the symplectic area tells us which of $\tilde{\alpha}(0)$ and $\tilde{\alpha}(1)$ is 'above' the other.

More generally, it is natural to ask whether can we find a horizontal lift of a map $f : M \to \mathbb{R}^{2n}$, where M is some manifold. If M is simply connected, $f : M \to \mathbb{R}^{2n}$ is isotropic, $x \in M$, and $\tilde{x} \in \pi^{-1}(f(x))$, then there is a unique horizontal map $\tilde{f} : M \to \mathcal{H}^{2n+1}$ with the properties that $\tilde{f}(x) = \tilde{x}$ and that for all $y \in M$, $\pi \circ \tilde{f}(y) = f(y)$. We will use this to build horizontal lifts of the isotropic disks $f : D \to \mathbb{R}^{2n}$ obtained in theorem 2.3. The lift $\tilde{f}$ may be defined as follows: if $y \in M$, choose a smooth path $\alpha : [0, 1] \to M$ from x to y. Define $\tilde{f}(y)$ to be the endpoint $\widetilde{f \circ \alpha}(1)$ of the horizontal lift $\widetilde{f \circ \alpha}$ satisfying $\widetilde{f \circ \alpha}(0) = \tilde{x}$. This definition does not depend on the choice of path α because if β is another path in M from x to y, then α and β together bound a disk in M, and so $f \circ \alpha$ and $f \circ \beta$ together bound a disk in $\mathbb{R}^{2n}$ with zero symplectic area. Hence, any horizontal lift of the boundary of this disk is a closed loop, which is to say that $\widetilde{f \circ \alpha}(1) = \widetilde{f \circ \beta}(1)$.

We note the simple yet important fact that if α is a horizontal path in $\mathcal{H}^{2n+1}$, then its length measured with respect to g is the same as the Euclidean length of its projection in $\mathbb{R}^{2n}$. Similarly, the area of a horizontal surface in $\mathcal{H}^{2n+1}$ is the same as the Euclidean area of its image in $\mathbb{R}^{2n}$. These properties hold because π carries each horizontal space H_x isometrically to $T_{\pi(x)}\mathbb{R}^{2n}$.

4 An Isoperimetric Inequality for $\mathcal{H}^{2n+1}$

Theorem 4.1. *The Heisenberg groups $\mathcal{H}^{2n+1}$ with $n > 1$ satisfy a quadratic isoperimetric inequality.*

Remarks: The proof below show that if a loop has length L, then it spans a disk of area less than $(1+k)L^2 + 4\pi^{1/2}(k+1/3)L^{3/2} + (4\pi k + 2)L$, where k is the constant of theorem 2.3. The proof uses the following lemmas, whose proofs appear after the proof of the theorem.

Lemma 4.2. *A path in $\mathcal{H}^{2n+1}$ $(n \geq 1)$ of length L is homotopic (rel endpoints) to a path which is the composition of two paths, each of length at most L, the first being horizontal and the second being vertical; such a homotopy may be chosen with area at most L^2.*

Lemma 4.3. *A vertical path in $\mathcal{H}^{2n+1}$ $(n \geq 1)$ of length L is homotopic (rel endpoints) to a horizontal path of length $2(\pi L)^{1/2}$, by a homotopy of area $\leq 2L + 4\pi^{1/2}L^{3/2}/3$.*

Proof of theorem 4.1: First suppose that γ is a smooth horizontal loop of length L. Then $\pi \circ \gamma$ in $\mathbb{R}^{2n}$ also has length L, and since one of its horizontal lifts is the closed loop γ, it encloses zero symplectic area. By theorem 2.3, $\pi \circ \gamma$ bounds a smooth isotropic disk D with area at most kL^2. A horizontal lift of this isotropic disk has the same area as D, and one of the horizontal lifts of D bounds γ. This completes the proof in the case that γ is horizontal.

Now suppose that γ is an arbitrary smooth loop of length L in $\mathcal{H}^{2n+1}$. Using lemma 4.2 and then applying lemma 4.3 in the obvious way, we apply a homotopy carrying γ to a horizontal loop γ' of length at most $L + 2\sqrt{\pi L}$; the area of this homotopy may be taken to be $\leq L^2 + 2L + 4\pi^{1/2}L^{3/2}/3$. Applying the horizontal case to γ', and adding together the areas of the homotopies used, we find that γ spans a disk of area at most the bound given in the remark. □

The geometric image to keep in mind for the proofs of the two lemmas is that if δ is a simple smooth path in $\mathbb{R}^{2n}$ then $\pi^{-1}(\delta)$ is locally isometric to $\mathbb{R}^2$, having two orthogonal foliations. These are given by the horizontal lifts of δ and the preimages under π of the points of δ.

Proof of lemma 4.2: Fix an isometry $\phi : \mathbb{R} \to \mathcal{Z}$ that carries $0 \in \mathbb{R}$ to the identity element of $\mathcal{Z}$; ϕ is also a group homomorphism. Let $\gamma : [0,1] \to \mathcal{H}^{2n+1}$ be a path and define γ_1 to be the horizontal lift of $\pi \circ \gamma$ with $\gamma_1(0) = \gamma(0)$. Consider the map $f : [0,1] \times \mathbb{R} \to \mathcal{H}^{2n+1}$ given by $f(s,t) = \gamma_1(s)\phi(t)$, juxtaposition denoting multiplication in $\mathcal{H}^{2n+1}$. All computations will take place in $[0,1] \times \mathbb{R}$, so we establish some facts regarding it. We define a horizontal line to be a set of the form $[0,1] \times \{x\}$ $(x \in \mathbb{R})$, and a vertical line to be a set of the form $\{x\} \times \mathbb{R}$ $(x \in [0,1])$. With respect to the pullback "metric" f^*g, vertical and horizontal lines are orthogonal, and vertical line segments have their natural lengths. (Note that f^*g may fail to be a metric by being degenerate at some points.) Length along horizontal lines depends on $\|\gamma_1'\|$, but all we will need about the behavior of f^*g on horizontal lines is that the total length of any horizontal line is the length of γ_1, which is bounded by L.

Define $\beta_1, \beta : [0,1] \to [0,1] \times \mathbb{R}$ to be the unique maps such that $f \circ \beta_1 = \gamma_1$ and $f \circ \beta = \gamma$. We have $\beta_1(s) = (s,0)$ and $\beta(s) = (s, u(s))$ for some function $u : [0,1] \to \mathbb{R}$; this function measures the difference between $\gamma(s)$ and the horizontal path $\gamma_1(s)$. Let B be the homotopy which pushes β along fibers to β_1. Explicitly, $B(s,t) = (s, (1-t)u(s))$, for $t \in [0,1]$. The length of the track s of B is just $u(s)$, which is obviously bounded by the arc length of $\beta([0,s])$, which is in turn bounded by L. Since $|u(s)| \leq L$ for all s, we deduce that the area of B is at most $L \cdot \text{length}(\beta_1) \leq L^2$. Observe that B leaves $\beta(0)$ fixed, while moving $\beta(1)$ toward $\beta_1(1)$; by reparameterizing B, we may regard it as a homotopy rel endpoints between β and the path obtained by first traversing β_1 and then the vertical arc (of length $\leq L$) from $\beta_1(1)$ to $\beta(1)$. The homotopy promised in the lemma is $f \circ B$. Lengths and areas of objects in $[0,1] \times \mathbb{R}$, measured with respect to f^*g, coincide with the lengths and areas of their images under f, measured with respect to g. This delivers the promised bounds for length and area. $\square$

Proof of lemma 4.3 The homotopy we will construct lies entirely in $\mathcal{H}^3$. For each $t \in [0, \infty)$ let $S_t(\theta)$ (for $\theta \in [0, 2\pi]$) be the path in $\mathbb{R}^2$ that begins at 0 and travels counterclockwise with constant speed around the circle with center $(-t, 0)$ and radius t. As t varies between 0 and any given $T > 0$, the S_t sweep out a disk in $\mathbb{R}^2$ and the $\tilde{S}_t$ sweep out a disk $\tilde{D}(T)$ in $\mathcal{H}^3$, where $\tilde{S}_t$ denotes the horizontal lift of S_t beginning at the identity of $\mathcal{H}^3$. For all t, $\tilde{S}_t(2\pi)$ lies over $0 \in \mathbb{R}^2$, so the path $\tilde{S}_t(2\pi)$ (with $t \in [0, T]$) is a vertical path. Setting $T = (L/\pi)^{1/2}$ we see that $\tilde{D}(T)$ provides a homotopy (rel endpoints) between a vertical path of length L and the horizontal path $\widetilde{S_T}$ of length $2(\pi L)^{1/2}$.

It remains to bound the area of $\tilde{D}(T)$. We may parameterize the homotopy by

$$\Gamma : [0, T] \times [0, 2\pi] \to \mathcal{H}^3$$
$$\Gamma(t, \theta) = \tilde{S}_t(\theta).$$

We observe $\|\partial\Gamma/\partial\theta\| = \|\partial\tilde{S}_t/\partial\theta\| = \|\partial S_t/\partial\theta\| = t$. We write $(\partial\Gamma/\partial t)(t, \theta) = h(t, \theta) + v(t, \theta)$ where $h(t, \theta)$ and $v(t, \theta)$ are horizontal and vertical vectors at $\Gamma(t, \theta)$, respectively. We have $\|h(t, \theta)\| = \|(\partial S_t/\partial t)(t, \theta)\|$. Since $S_t(\theta) = (-t, 0) + (t, 0)\cos\theta + (0, t)\sin\theta$ we see that $\|h(t, \theta)\| \leq 2$ for all t and θ. We bound $\|v(t, \theta)\|$ by observing that $v(t, \theta)$ is given by the infinitesimal area of the region bounded by the arcs $S_t([0, \theta])$ and $S_{t+dt}([0, \theta])$ and the infinitesimal segment joining $S_t(\theta)$ and $S_{t+dt}(\theta)$. Formally, letting $A(t, \theta)$ be the area of $\pi \circ \Gamma([0, t] \times [0, \theta])$, we observe that

$$\|v(t, \theta)\| = \|(\partial A/\partial t)(t, \theta)\|.$$

The right hand side is obviously bounded by $\|(\partial A/\partial t)(t, 2\pi)\| = 2\pi t$. Our bounds on the norms of h and v show that $\|\partial\Gamma/\partial t\| \leq 2 + 2\pi t$. Finally, writing X for $[0, T] \times [0, 2\pi]$, we have

$$\begin{aligned}
\text{Area}(\Gamma(X)) &\leq \int_X \|\partial\Gamma/\partial\theta\| \, \|\partial\Gamma/\partial t\| \, dt \, d\theta \\
&\leq \int_X t(2 + 2\pi t) dt \, d\theta \\
&= 2\pi(T^2 + 2\pi T^3/3) \\
&= 2L + 4\pi^{1/2} L^{3/2}/3.
\end{aligned}$$

$\square$

5 Applications and Remarks

Theorem 4.1 has several applications in the field of geometric group theory. Since $\mathcal{H}^5, \mathcal{H}^7, \ldots$ satisfy a quadratic isoperimetric inequality, so does any group which acts cocompactly, discretely, and isometrically on one of them (see [2]). The main example, the discrete Heisenberg group $\mathcal{H}^{2n+1}_{\mathbb{Z}}$, has generators $x_1, \ldots, x_n, y_1, \ldots, y_n, z$ and relations asserting that $[x_i, y_i] = z$ and that all other pairs of generators commute. $\mathcal{H}^{2n+1}_{\mathbb{Z}}$ is a cocompact discrete subgroup of $\mathcal{H}^{2n+1}$, and we may conclude that this group satisfies a quadratic isoperimetric inequality when $n > 1$. For more information about isoperimetric inequalities for finitely presented groups, see [5], [6]. It is easy to see that with respect to the presentation above, $\mathcal{H}^3_{\mathbb{Z}}$ is isometrically embedded in $\mathcal{H}^{2n+1}_{\mathbb{Z}}$; it is also true that $\mathcal{H}^3$ is isometrically embedded in $\mathcal{H}^{2n+1}$. It is surprising that efficient spanning of loops in the isometrically embedded $\mathcal{H}^3$ *requires* disks that do not lie in $\mathcal{H}^3$. Finally, by proving theorem 4.1, we have justified the claim in [3] that the groups $\mathcal{H}_{\mathbb{Z}}^{2n+1}$ for $n > 1$ provide examples of finitely presented groups that satisfy quadratic isoperimetric inequalities while not admitting automatic structures.

Another important reason to study the Heisenberg groups is that they occur as horospheres in complex hyperbolic space $\mathbb{C}H^n$. Another family of nilpotent Lie groups, the quaternionic Heisenberg groups $\mathbb{H}\mathcal{H}^{4n-1}$, appear as the horospheres in quaternionic hyperbolic space $\mathbb{H}H^n$ (see [11]), and if $n > 4$ then the techniques of §§ 2–4 can be applied directly to yield quadratic isoperimetric inequalities for them. In the case $n = 4$, the techniques can also be applied, but more care is required. If a group G acts (meaning that it acts discretely and isometrically) on $\mathbb{C}H^n$ or $\mathbb{H}H^n$ with noncompact finite volume quotient, then each of its cusp groups acts cocompactly on a copy of $\mathcal{H}^{2n-1}$ or $\mathbb{H}\mathcal{H}^{4n-1}$. In his work on relatively hyperbolic groups, Farb [4] has shown that under these conditions, G and its cusp groups satisfy isoperimetric inequalities of the same degrees. This proves

Theorem 5.1. *A group G which acts on $\mathbb{C}H^n$ ($n > 2$) or $\mathbb{H}H^n$ ($n > 3$) with noncompact finite-volume quotient satisfies a quadratic (but no subquadratic) isoperimetric inequality.*

Remark: The optimality follows from work of Gromov [8, p. 104] and Olshanskii [13], that any group satisfying a subquadratic isoperimetric inequality is word hyperbolic. Since a word hyperbolic group cannot contain a subgroup isomorphic to $\mathbb{Z}^2$, and G contains Heisenberg groups and hence copies of $\mathbb{Z}^2$, G cannot satisfy a subquadratic inequality. If a group acts as in the theorem on real hyperbolic space H^n then it satisfies a quadratic inequality (and a linear one if $n = 1$), and if on $\mathbb{C}H^2$ it satisfies a cubic inequality. Again these are optimal. The only hyperbolic spaces for which such precise results are not yet known are $\mathbb{H}H^2$, $\mathbb{H}H^3$, and the hyperbolic plane $\mathbb{O}H^2$ defined over the alternative field of octaves. (See [11] for a description of these spaces.)

We hope the reader will find merit in the following comments upon the proofs. The heart of the proof of theorem 4.1 is contained in its first paragraph. It is known that $\mathcal{H}^{2n+1}$ is quasi-isometric to the metric space $\mathcal{H}^{2n+1}_{\text{Carnot}}$, in which the distance between two points is the infimum of the lengths of *horizontal* paths joining them. Since quasi-isometric spaces tend to satisfy isoperimetric inequalities of the same degrees (see, e.g., [1]), one might hope to deduce the isoperimetric inequality for $\mathcal{H}^{2n+1}$ from this, but it's not clear what 'area' means in $\mathcal{H}^{2n+1}_{\text{Carnot}}$.

We began this work by trying to fill loops in a Cayley graph for $\mathcal{H}_{\mathbb{Z}}^{2n+1}$, working combinatorially. After a while, it became clear that two words in the generators x_i, y_i commute exactly when the parallelogram they span in the central quotient $\mathbb{Z}^{2n}$ of $\mathcal{H}_{\mathbb{Z}}^{2n+1}$ encloses zero symplectic area. Then it seemed more natural to work with polygonal paths in $\mathbb{R}^{2n}$, and it was in this setting that the proof was completed. It is something of a bonus that the technique applies to smooth loops, and in fact is phrased most naturally in terms of them. Most of our constructions have analogues in $\mathcal{H}_{\mathbb{Z}}^{2n+1}$. The process used in the proof of lemma 4.2 corresponds to the operation on words in $\mathcal{H}_{\mathbb{Z}}^{2n+1}$ of commuting each z all the way to the far end. Lemma 4.3 performs in our smooth setting the same sort of service as the combinatorial operation of replacing each z by $x_1 y_1 x_1^{-1} y_1^{-1}$. The proof of lemma 2.1 uses a smooth version of the process "commute all the x_1's and y_1's to the beginning of a word." Only lemma 2.2 seems to have no neat combinatorial analogue. In fact, even if the given loops are polygonal, it produces a homotopy between them that isn't polygonal: the tracks of the homotopy are elliptical arcs. Anyone wishing to devise a combinatorial algorithm for contracting loops in $H_{\mathbb{Z}}^{2n+1}$ might first describe how to span a polygonal loop in $\mathbb{R}^{2n}$ (that encloses zero symplectic area) with a polygonal isotropic disk of appropriately small area.

Gromov [10] has developed his arguments of [9] to give another proof of theorem 4.1. He derives the result from his theorem 3.5D, the "disk extension theorem", a version of which we reproduce here.

Theorem 5.2. *If V is a simply connected compact Riemannian contact manifold of dimension $2n+1 \geq 5$ then there exists $C > 0$ such that the following holds. For every Lipschitz function $f_0 : S^1 \to V$ there exists a Lipschitz extension $f : D^2 \to V$ of f_0 such that the Lipschitz constant of f is bounded by C times that of f_0.*

(A contact $(2n+1)$-manifold is a manifold equipped with a hyperplane field locally equivalent to that of $\mathcal{H}^{2n+1}$, and the term Lipschitz refers to the Carnot metric on V induced by the Riemannian metric and the hyperplane distribution. A smooth map is called horizontal if its derivative takes values in the distribution, and smooth horizontal maps from S^1 are automatically Lipschitz.)

This immediately implies that a horizontal loop γ in V of length L spans a horizontal disk of area $\leq C^2L^2/4\pi$, which yields a quadratic isoperimetric inequality for horizontal loops in V. Despite the hypothesis that V be compact, one may apply the theorem to prove our theorem 4.1. One takes V to be a simply connected compact manifold-with-boundary neighborhood of $1 \in \mathcal{H}^{2n+1}$. If γ is any horizontal loop in $\mathcal{H}^{2n+1}$ of length L then let γ' be the image of γ under the "dilation" $x_i \mapsto tx_i$, $y_i \mapsto ty_i$, $z \mapsto t^2z$ for t small enough so that $\gamma' \subseteq V$. Because the dilation preserves the field of horizontal hyperplanes, γ' is horizontal and has length tL. By the theorem, it spans a horizontal disk of area $\leq C^2(tL)^2/4\pi$. Then by taking the image of the disk under the inverse of the dilation we see that γ spans a disk of area $\leq C^2L^2/4\pi$. This argument does not address the issue of non-horizontal loops in $\mathcal{H}^{2n+1}$, but one is generally not interested in such things when thinking about Carnot geometry. In any case, to prove that any loop bounds a disk of Riemannian area quadratic in the loop's length, one can reduce to the horizontal case and apply the argument above. The reduction to the horizontal case is easy (say, our lemmas 4.2 and 4.3.)

Gromov's arguments for theorem 5.2 are part of a systematic study of Lipschitz maps to Carnot-Carathéodory manifolds (a generalization of a contact manifold). Among other things, he treats the approximation of continuous functions by horizontal Lipschitz functions, which can be used to provide the reduction to the case of horizontal loops. He also treats the problem of extending a piecewise smooth and horizontal map satisfying bounds on its higher derivatives by functions satisfying similar conditions.

Our methods differ from Gromov's primarily in that the explicit homotopies we employ use the global structure of $\mathcal{H}^{2n+1}$ rather than just the local contact structure. (Our homotopies may wander far from the original curves.) The proofs of lemmas 4.2 and 4.3 could be rewritten to show that for any curve γ in $\mathcal{H}^{2n+1}$ there exists a horizontal curve of approximately the same length within (say) distance 1 of γ. However, the key to our construction, lemma 2.2, seems to require a "global homotopy". As a consequence, it seems unlikely that the argument could be modified to work for general Carnot-Carathéodory manifolds. On the other hand, our explicit approach has the advantage of directness and provides a uniform isoperimetric inequality for all the $\mathcal{H}^{2n+1}$ with $n > 1$.

Yng-Ing Lee [12] has also proven a version of theorem 2.3. Namely, that there is a constant $c > 0$ such that if γ is a smoothly embedded loop of length L in $\mathbb{R}^4$ such that (i) the center of mass of γ is the origin, (ii) γ encloses zero symplectic area, and (iii) γ satisfies condition **(H)** below, then γ bounds a smoothly embedded disk in $\mathbb{R}^4$ of area $\leq cL^2$. We say that γ satisfies condition **(H)** if there is a smooth homotopy through smoothly embedded loops γ_t, each enclosing zero symplectic area, that carries $\gamma = \gamma_1$ to $\gamma_0 = C$, a standard circle of circumference L in the (x_1, x_2)-plane. Below, we compare this with her original definition. Her construction is completely explicit (given the homotopy) and has the advantage of yielding an embedded spanning disk. However, her method only works for smoothly embedded loops satisfying **(H)**. This is not a serious drawback: she sketches an argument that any loop enclosing zero symplectic area may be approximated by a loop satisfying **(H)**.

Her argument is more in the style of pure symplectic geometry than ours: it proceeds by realizing the given homotopy as the track of γ under the flow of a suitable time-dependent Hamiltonian vector field V_t. Reversing the flow carries the obvious disk D spanning C to a disk bounded by γ. Since the flow of a Hamiltonian vector field is a symplectomorphism, this yields an isotropic disk spanned by γ. The key to her argmument is that one can choose the vector fields V_t in such a way that they and their first derivatives with respect to the x_i and y_i are all bounded by some universal constant. This lets one estimate how much the disk D is distorted as it flows along the V_t, and thus yields a bound on the area of the disk spanning γ. She chooses the V_t by explicitly writing down a time-dependent Hamiltonian function, and then checking that is satisfies the various required properties.

We close by indicating why our version of condition**(H)** implies hers. The additional criteria she imposes are that all the γ_t have the same length and that there be $\delta > 0$ such that there is an embedded δ-tubular neighborhood of γ_t for all t. If γ satisfies our version of **(H)**, with the length of γ_t being L_t, then replacing each γ_t with the scaled loop $\frac{L}{L_t}\gamma_t$, we recover her first extra condition. The existence of the tubular neighborhoods for some δ follows from the facts that each γ_t is smoothly embedded and that there is a uniform bound on the extrinsic curvature of the paths γ_t. (The latter claim follows from the compactness of the homotopy.)

References

[1] J. Alonso. Inegalities isoperimtriques et quasi-isometries. *C.R.A.S., Série A*, 311:761–4, 1991.

[2] J. Burillo. Equivalence of geometric and combinatorial Dehn functions. preprint, 1995.

[3] D. B. A. Epstein, J. W. Cannon, S. V. F. Levy, M. S. Paterson, and W. P. Thurston. *Word Processing in Groups*. Jones and Bartlett, Boston, 1992.

[4] B. Farb. *Relatively hyperbolic groups, with applications to negatively curved manifolds*. PhD thesis, Princeton University, 1995.

[5] S. Gersten. Dehn functions and ℓ_1-norms of finite presentations. In G. Baumslag. and Miller, editors, *Algorithms and Classification in Combinatorial Group Theory*, pages 195–224. Springer-Verlag, 1992.

[6] S. Gersten. Isoperimetric and isodiametric inequalities in finite presentations. preprint, 1992.

[7] W. M. Goldman. Complex hyperbolic geometry. preprint, 1993.

[8] M. Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in Group Theory*, volume 8 of *MSRI Publications*, pages 75–263. Springer-Verlag, 1987.

[9] M. Gromov. Asymptotic invariants for infinite groups. In Niblo and Roller, editors, *Geometric Group Theory, vol. 2*, number 182 in London Mathematical Societies Lecture Notes. Cambridge University Press, 1993.

[10] M. Gromov. Carnot-Carathéodory spaces seen from within. *Progress in Mathematics*, 144:79–323, 1996.

[11] M. Gromov et al., editors. *Geometric Topology: Recent Developments*, number 1504 in Lecture Notes in Mathematics. Springer-Verlag, 1991.

[12] Y.-I. Lee. *The Metric Properties of Lagrangian Surfaces*. PhD thesis, Stanford University, 1992.

[13] A. Y. Olshanskii. Hyperbolicity of groups with subquadratic isoperimetric inequality. *IJAC*, 1:281–289, 1991.

[14] P. Scott. The geometries of 3-manifolds. *Bulletin of the London Mathematical Society*, 15:401–87, 1983.