1. Introduction and the Heisenberg group

In this set of notes, we explore nilpotent groups from a geometric and geometric group theoretic point of view. We'll start off by introducing the discrete and real Heisenberg groups. Next we'll define free nilpotent groups and Mal'cev coordinates for finitely generated nilpotent groups. These Mal'cev coordinates are useful in proving the decidability of equation problems for nilpotent groups, and they will also provide insight into the proof that nilpotent groups grow polynomially. We will then define Carnot-Carthéodory metrics on certain nilpotent Lie groups, and explore the asymptotic geometry of finitely generated nilpotent groups.

1.1. **The Heisenberg group.** The Heisenberg groups, $H(\mathbb{Z})$ and $H(\mathbb{R})$, are the groups of unipotent upper-triangular 3×3 matrices $\left\{ \begin{pmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix} \right\}$ where the entries are integers and real numbers, respectively. It is easy to see that $H(\mathbb{Z})$ sits as a *uniform lattice* inside $H(\mathbb{R})$; that is, it is a discrete subgroup for which the quotient space is compact. (The cocompactness will become geometrically clear below.) $H(\mathbb{Z})$ is sometimes also denoted H_3 because it has three parameters (and there are *higher Heisenberg groups* with more parameters, as we will see).

Define

$$a = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

An easy direct computation shows that $[a, b] = c$, where $[a, b]$ denotes the *commutator* $aba^{-1}b^{-1}$. Nilpotent groups are defined by their behavior under this commutator operation: note that the commutator of two group elements vanishes precisely if the elements commute, so in $H(\mathbb{Z})$, we can speak of c as "measuring the failure of a and b to commute."

Thus we get $H(\mathbb{Z}) = \langle a, b, c\rangle = \langle a, b\rangle$, because the three elementary matrices suffice to generate the whole group, but c is not needed as a generator since it can be built as the commutator of the other two. But along the way to this we can note the fundamental asymmetry in the matrix coordinates:

$$ab = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \text{ while } \quad ba = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Actually let's notice something else: for unipotent upper-triangular matrices of any size, the multiplication is *additive* in the first non-zero superdiagonal. For instance,

$$\begin{pmatrix} 1 & 0 & x & z \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & x' & z' \\ 0 & 1 & 0 & y' \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x+x' & * \\ 0 & 1 & 0 & y+y' \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

So multiplication in the Heisenberg group is be abelian in the first superdiagonal, i.e., in the a and b coordinates; it is in the c coordinate where the non-commutativity appears.

1.2. **Coordinates.** Next, we observe that $a^A b^B c^C$ is a normal form for group elements, meaning that every group element can be written in exactly one way in this form. (It is not, however, a geodesic normal form—the spellings are unique but not minimal.) To see this, note that $ba = abc^{-1}$ and check that c is central. That means that any word spelled in a, b, c can be rewritten by pulling c letters to the end, then organizing the a and b letters, which only generates more c and c^{-1}. Furthermore if $a^A b^B c^C = a^\alpha b^\beta c^\gamma$, then one checks easily that $A = \alpha, B = \beta, C = \gamma$.

This means we can regard the triple of integers (A, B, C) as giving coordinates on the group. We will call these *Mal'cev coordinates*, and they are excellent for algebraic arguments on the group.

We can write down the group multiplication formula in these coordinates as follows:

$$(A, B, C) \cdot (A', B', C') = (A + A',\ B + B',\ C + C' - A'B).$$

And we have the following matrix correspondence:

$$(A, B, C) \leftrightarrow \begin{pmatrix} 1 & A & C+AB \\ 0 & 1 & B \\ 0 & 0 & 1 \end{pmatrix}.$$

These coordinates allow us to identify $H(\mathbb{Z})$ with $\mathbb{Z}^3$. However, there is a second coordinate system for $H(\mathbb{R})$ that is far better suited to geometric arguments, called *exponential coordinates*. These are the coordinates

$$(x,y,z) \leftrightarrow \begin{pmatrix} 1 & x & z+\frac{1}{2}xy \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix},$$

which have the extremely useful property that $(x,y,z)^n = (nx, ny, nz)$. With these coordinates, $H(\mathbb{R})$ can be identified with $\mathbb{R}^3$ and $H(\mathbb{Z})$ with the lattice obtained by shifting $\mathbb{Z}^3$ vertically by a half-integer over all (odd, odd) points in the xy–plane. (Note that here is where we see $H(\mathbb{Z})$ sitting as a cocompact lattice in $H(\mathbb{R})$, as stated above.)

In exponential coordinates, one derives the group multiplication formula

$$(x,y,z)\cdot(x',y',z') = \left(x+x',\ y+y',\ z+z'+\frac{xy'-yx'}{2}\right).$$

Exercise 1. Write ab in Mal'cev coordinates and in exponential coordinates.

1.3. **Near-similarity.** Consider the maps $\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{\delta_n} \begin{pmatrix} 1 & nx & n^2z \\ 0 & 1 & ny \\ 0 & 0 & 1 \end{pmatrix}$, which you will note are also given in exponential coordinates by $\delta_n(x,y,z) = (nx, ny, n^2z)$ and in Mal'cev coordinates by $\delta_n(A,B,C) = (nA, nB, n^2C)$. We want to show that these are very nearly similarities in $H(\mathbb{Z})$ and derive several interesting consequences. A *similarity* in a metric space is a family of maps $\Delta_t : X \to X$ such that $d(\Delta_t(\mathsf{x}), \Delta_t(\mathsf{y})) = t \cdot d(\mathsf{x},\mathsf{y})$, such as homotheties of normed spaces.

Exercise 2. First, consider the generators $S = \{a,b,c\}^{\pm}$ for and the word metric $|\cdot|_S$ with respect to S in $H(\mathbb{Z})$. Show that

$$\frac{1}{2}\left(|A|+|B|+\sqrt{|C|}\right) \leqslant |(A,B,C)|_S \leqslant |A|+|B|+6\sqrt{|C|}.$$

Find similar upper and lower bounds in the standard generators $\mathsf{std} = \{a,b\}^{\pm}$. Explain why something similar holds for arbitrary generators.

So in general we have that $|(A,B,C)|$ for any S is bounded above and below by multiples of $|A| + |B| + \sqrt{|C|}$. And thus we can derive a comparison of $|\mathsf{w}|$ with $|\delta_n(\mathsf{w})|$ for an arbitrary word w:

$$k_1 n|\mathsf{w}| \leqslant |\delta_n(\mathsf{w})| \leqslant k_2 n|\mathsf{w}|,$$

where k_1, k_2 are constants coming from the generating set. Let's call such a map, which is a similarity up to a bounded multiplicative factor, a *near-similarity*. So a true similarity has $|\delta_n(\mathsf{w})|/n|\mathsf{w}| \equiv 1$, and this near-similarity has that ratio bounded above and below.

1.3.1. *Consequence: Hausdorff dimension 4.* Now consider a metric on $\mathbb{R}^3$ for which $\delta_n(x,y,z) = (nx, ny, n^2z)$ is an similarity. We will define such a metric on the Heisenberg group in Section 4.2. What is the dimension of such a space? *Hausdorff dimension* is defined directly from the distance function, as follows: we let the d–dimensional Hausdorff measure of a set $E \subset X$ be

$$\nu(E) := \lim_{\delta\to 0}\left[\inf \sum \operatorname{diam}(U_i)^d\right],$$

where the infimum is over countable covers $\{U_i\}$ of E with $\operatorname{diam} U_i < \delta$ for all i. Then one proves that there is a critical dimension: if d is too large, the measure of the whole space is zero; if d is too small, the space has infinite measure. The Hausdorff dimension is the threshold where $\nu(X)$ transitions from being infinite to zero.

Let C be the standard cube in $\mathbb{R}^3$ with vertices at $(0,0,0)$ and at $(1,1,1)$. Note that $\delta_\epsilon(C)$ has vertices at $(0,0,0)$ and $(\epsilon,\epsilon,\epsilon^2)$. It clearly requires $1/\epsilon^4$ slightly fattened copies of these to cover C. And if the original C has diameter D (in the metric we are studying), then the dilated cube has diameter ϵD by the assumption that δ_n is a similarity. Thus the Hausdorff measure of C is $(\epsilon D)^d/\epsilon^4$, which only has a finite nonzero limit if $d = 4$. (Of course, this is hiding some work that is needed to show that this is the most efficient kind of cover.)

1.3.2. *Consequence: n^4 growth.* Hausdorff dimension doesn't quite make sense in the same way for discrete groups, but nonetheless we can see that the group growth has the same rate as in $\mathbb{Z}^4$, not $\mathbb{Z}^3$.

The standard *growth function* in geometric group theory is the counting function $\beta(n) := |B_n| = \#\{g \in G : |g| \leqslant n\}$ that counts how many group elements can be spelled in n or fewer letters.

Let us say that $f(t) \prec g(t)$ if $\exists K > 0$ such that $f(t) \leqslant K \cdot g(Kt + K) + Kt + K$ for all $t \geqslant 0$, and that $f \asymp g$ if $f \prec g, g \prec f$. Then it is a classic observation in geometric group theory that although β depends on (G, S), it is well-defined up to $\asymp$ depending only on G. (That is, the change from one finite generating set to another induces just an affine change in the metric, which is accounted for by the equivalence relation $\asymp$.)

The near-similarity says that B_{kn} is well-approximated by $\delta_n(B_k)$, so if we can inscribe and circumscribe a nice Riemann-integrable region around B_k, we'll have $\beta(kn) \approx n^4 \cdot \beta(k)$ for some small fixed k, which means $\beta(n) \asymp n^4$.

More precisely: all points (A, B, C) with $|A|, |B| \leqslant r$ and $|C| \leqslant r^2$ have word length at most $8r$ in the $\{a, b, c\}^{\pm}$ generators (by the inequality from Exercise 2), so they are in the ball of radius $8r$ about the identity. But there are more than $8r^4$ many such triples, so we know that $\beta(8r) \geqslant 8r^4$, which means $\beta(n) \geqslant \frac{1}{512}n^4$. On the other hand, to be in the ball of radius n a word must satisfy $|(A, B, C)| \leqslant n$, and in particular this forces $|A|$, $|B|$, and $\sqrt{|C|}$ to all be at most $2n$, so there are no more than $4n + 1$ possible values for A and B and $8n^2 + 1$ possible values for C. This means $\beta(n) \leqslant 129n^4$ for large n. So $\beta(n) \asymp n^4$.

1.4. **Hyperboliclity.** The standard definition of δ–hyperbolicity is that triangles are thin, i.e., if you take any geodesic triangle, then each side is contained in a δ–neighborhood of the union of the other two sides. This bound δ is sometimes called the *insize* of a triangle. As long as some triangles have strictly positive insize, then applying a similarity would produce a geodesic triangle with a larger insize, so a full family of similarities is an obstruction to hyperbolicity (outside of the $\delta = 0$ case, which can be handled with a different argument). Intuitively, it is clear that near-similarities should also be an obstruction to hyperbolicity. However, to actually prove this is tricky because it's not obvious that the δ_n–image of a geodesic would remain geodesic (or even uniformly quasi-geodesic). So to make the obstruction rigorous, we can instead use the less-known 4–point definition of hyperbolicity: consider any four points x, y, z, w in the space and the three *pairsums* $P_1 = d(x, y) + d(z, w)$, $P_2 = d(x, z) + d(y, w)$, and $P_3 = d(x, w) + d(y, z)$. A space is δ–hyperbolic if the largest two of these numbers differ by no more than δ. (This definition was formulated by Gromov and, among length spaces, is equivalent to other definitions of hyperbolicity, though possibly for different values of δ.)

From this definition, it is immediate that if there are any four points for which the top two pairsums are unequal, then a similarity is enough to rule out hyperbolicity for any $\delta > 0$.

Exercise 3. Is a near-similarity enough?

Exercise 4. Find four-tuples in $H(\mathbb{Z})$ which would have arbitrarily large difference of pairsums in any generating set. (One concludes that $H(\mathbb{Z})$ is not a hyperbolic group.)

2. LCS, Mal'cev coordinates

Here we will define various algebraic properties of groups, many of which arise from asking that *subnormal series* satisfy certain criteria. See Druţu and Kapovich's book [4] for a more detailed explanation of these algebraic properties. A (finite ascending) subnormal series of a group G is a sequence of subgroups

$$\{1\} = A_1 \trianglelefteq A_2 \trianglelefteq \ldots \trianglelefteq A_n = G$$

where for $i \geqslant 1$, A_i is a normal subgroup of A_{i+1}. The quotient groups A_{i+1}/A_i are called *factor groups*.

We say that a group G is *solvable* if there exists a subnormal series of G such that all factor groups A_{i+1}/A_i are abelian.

One (possibly infinite) *descending* subnormal series of interest is the *lower central series*. To construct the lower central series, we define $G_1 = G$, and let $G_{i+1} = [G_i, G]$, the subgroup generated by commutators $[x, g]$, with $x \in G_i$ and $g \in G$. We denote by $[a_1, \ldots, a_n]$ the nested commutator of group elements $a_1, \ldots, a_n \in G$. For example,

$$[a_1, a_2, a_3, a_4] = [[[a_1, a_2], a_3], a_4].$$

A *nilpotent group* is a group G whose lower central series terminates in the trivial group $\{1\}$ after finitely many steps.

$$G = G_1 \trianglelefteq G_2 \trianglelefteq \ldots \trianglelefteq G_s \trianglelefteq G_{s+1} = \{1\}$$

If $s + 1$ is the minimum length of a lower central series for G, then we call s (the index of the last nontrivial group) the *nilpotency class of* G. We'll also say that G is *s-step nilpotent*. To be clear, if G is an s-step nilpotent group, the lower central series will have $s + 1$ groups in it, where $G_1 = G$, and $G_{s+1} = \{1\}$. It is worth observing that $[G_i, G_i] \subseteq [G_i, G] = G_{i+1}$. Therefore the factor groups G_i/G_{i+1} are abelian, and hence all nilpotent groups are solvable.

Proposition 1. Although it is not true for general groups, if G is nilpotent and generated by the set S, then the subgroups G_i are generated by the i-step commutators

$$[g_1, g_2, \ldots, g_i], \text{ for } g_k \in S.$$

Consequently, if G is finitely generated, so are the subgroups of the lower central series.

Exercise 5. Compute the lower central series for $H(\mathbb{Z})$.

A group G is said to be *polycyclic* if there exists a subnormal series for G with cyclic factor groups. Since the factor groups in the lower central series of a nilpotent group are abelian, the structure theorem for finitely generated abelian groups tells us that these factor groups G_i/G_{i+1} decompose into products of cyclic groups. Thus the lower central series can be refined to a subnormal series with cyclic factor groups, making all nilpotent groups polycyclic.

A group G is said to be *metabelian* if its commutator subgroup $[G, G]$ is abelian. If G is 2-step nilpotent, then the factor group $G_2/G_3 = [G, G]/\{1\} = [G, G]$. Since all factor groups are abelian, in particular $[G, G]$ is, and so G is metabelian.

2.1. Free nilpotent groups.

The *free nilpotent group* $N_{s,m}$ of step s and rank m is made by first taking the free group on m generators, $J = \langle x_1, \ldots, x_m \rangle$. Let J_{s+1} be the subgroup coming from the lower central series of the free group J, generated by the $(s+1)$-step commutators. We define $N_{s,m}$ to be the quotient

$$N_{s,m} = J/J_{s+1}.$$

Note that in constructing $N_{s,m}$, the group is given the minimal number of relations required to be nilpotent of step s. This endows free nilpotent groups with a universality property, stated below.

Proposition 2. If G is a nilpotent group of step s and has a generating set $\{a_1, \ldots, a_m\}$, then G is a quotient of the free nilpotent group $N_{s,m}$.

Exercise 6. What is the free nilpotent group $N_{2,2}$? What other name does it go by?

2.2. **The torsion subgroup.** An element x of G is said to be *torsion* if there exists a positive integer exponent n such that $x^n = 1$ in G. A group G is a *torsion group* if every element of G is a torsion element. Our goal will be to show that the torsion elements of a nilpotent group form a finite subgroup. Note that this is not a trivial statement. Think, for example, of the free product $(\mathbb{Z}/n\mathbb{Z}) * (\mathbb{Z}/m\mathbb{Z}) = \langle a \rangle * \langle b \rangle$. Although a and b are torsion elements, their product ab has infinite order. In the special case of nilpotent groups, we'll call the subgroup generated by torsion elements the *torsion subgroup*, Tor G.

Before we can show that this torsion subgroup exists, we have to do a bit of work.

Exercise 7. Let $x, y \in G$. and suppose that x and $[x, y]$ commute. Show that for all n, we have $[x, y]^n = [x^n, y]$.

Lemma 3. If a group G is generated by elements of finite order, then for all $i \geqslant 1$, the quotient group G_i/G_{i+1} is torsion.

Proof. We induct on i. If $i = 1$ then $G_i/G_{i+1} = G/[G, G] = G^{\mathrm{ab}}$. Abelian groups generated by torsion elements are torsion.

Assume the statement is true for i. Consider the group $\Gamma = G_{i+1}/G_{i+2}$. From above, we know that all factor groups in the lower central series, in particular Γ, are abelian, and so to show that Γ is torsion, we need only show that it is generated by torsion elements.

Γ is generated by elements of the form $[x, g]$, where $x \in G_i$ and $g \in G$. Using Exercise 7, we see that modulo G_{k+2}, $[x^n, g] \equiv [x, g]^n$ since Γ is abelian. Due to the induction hypothesis on G_i/G_{i+1}, we know there exists $n \in \mathbb{N}$ such that $x^n \in G_{i+1}$.

Combining these two facts, we see that $[x, g]^n = [x^n, g] \in G_{i+2}$, and so $[x, g]^n \equiv 1$ modulo G_{i+2}. Thus Γ is torsion. $\square$

Theorem 4. Let G be a nilpotent group. The set of all finite order elements forms a subgroup of G, denoted by Tor G.

Proof. We induct on the nilpotency class s of G. If $s = 1$, then G is abelian, and the statement is clear.

Assume the statement is true for s-step nilpotent groups, and assume that G is a $(s+1)$-step nilpotent group. Given two torsion elements x and y, without loss of generality we can let G be the subgroup generated by x and y. We want to show that xy is also a torsion element. Note that the quotient group G/G_{s+1} is nilpotent of class s. By the induction hypothesis, there exists a positive integer $n \in \mathbb{N}$ such that $(\bar{x}\bar{y})^n = 1$ in G/G_{s+1}. In other words, $(xy)^n \in G_{s+1}$. By Lemma 3, the factor group $G_{s+1}/G_{s+2} = G_{s+1}/\{1\} = G_{s+1}$ is torsion. Therefore there exists a positive integer $m \in \mathbb{N}$ such that $(xy)^{nm} = 1$. $\square$

Proposition 5. A finitely generated nilpotent torsion group is finite.

Proof. We induct on the nilpotency class s. If $s = 1$, then G is abelian. It is clear that a finitely generated abelian torsion group must be finite.

Assume the statement is true for s-step nilpotent groups, and let G be a $(s+1)$-step nilpotent group. Note that the group G_{s+1} is torsion and abelian by arguments given above. Therefore it is finite. The quotient group G/G_{s+1} is a finitely generated nilpotent group of step s, so by the induction hypothesis it is finite. Therefore G is finite. $\square$

2.3. **Mal'cev coordinates.** In Section 1.2, we defined the Mal'cev coordinates for the Heisenberg group. Similar coordinates can be defined for any finitely generated nilpotent group. The idea behind Mal'cev coordinates is to specify generators adapted to every level of the lower central series. This will allow us to define a normal form for the group. The Mal'cev coordinates are a powerful tool that allow us to solve equations in $H(\mathbb{Z})$, and they will help us see that nilpotent groups have polynomial growth.

First, we'll define Mal'cev coordinates for the free nilpotent group $G = N_{s,m}$. There are m generators that we'll call $a_1, \ldots, a_m$. Take these as our generators adapted to the first level $G_1 = G$. We can think of them as the generators of the abelian group $G_1/G_2 = G/[G, G]$.

The generators adapted to the second level will be the generators of the abelian group G_2/G_3. They are $b_{ij} := [a_i, a_j]$ for $1 \leqslant i < j \leqslant m$. If the step s is equal to 2, then the b_{ij} commute with everything, and we take the set $\{a_i, b_{ij}\}$ as our Mal'cev coordinates. If the step s is greater than two, we continue in the same way.

The generators of the 3rd step will be the generators of the abelian group G_3/G_4. We can call them $\{c_\mu\}$ for the proper indexing μ. The c_μ are the commutators of the b_{ij} with all of the other Mal'cev coordinates we have so far. Continue in this way until we reach the commutators of step s. In the end, take the union of the generators adapted to each step to be our Mal'cev coordinates. It is clear that the Mal'cev coordinates drastically over-generate our group since the a_i would be sufficient. But they have the added feature of providing a normal form for the group. We omit the proof that the Mal'cev coordinates actually do provide a normal form for nilpotent groups and, instead, we will explore some examples.

2.3.1. *Examples of Mal'cev coordinates.* In Section 1.2, we defined at the Mal'cev coordinates of $H(\mathbb{Z})$. To follow the notation described above, we would have rename the generators $a_1 = a$, $a_2 = b$ and $b_{12} = c$ since a and b correspond to the first level of the lower central series, or the first factor group $G_1/[G,G]$. In the remainder of the notes, we'll stick with the notation introduced earlier, using a, b, and c as the Mal'cev coordinates of $H(\mathbb{Z})$.

Let's find the Mal'cev coordinates of the free nilpotent group $G = N_{3,2}$. G is of rank 2 and of step 3. So it is generated by two elements a_1 and a_2 and has relations $[a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4}] = 1$. Therefore at the first level, we will take the two generators of G^{ab}, namely a_1 and a_2. At the second level, we have commutators of a_1 and a_2, of which there is only one (and its inverse). So here we add only one coordinate $b = [a_1, a_2]$. At the third and final level, our generators will be 3-step commutators of the a_i. We end up with two new coordinates, $c_1 = [a_1, a_2, a_1] = [b, a_1]$ and $c_2 = [b, a_2]$. Since all higher step commutators are trivial, the c_i commute with everything, and we are done.

The claim is that every word in $N_{3,2}$ can be written in the form $a_1^{A_1} a_2^{A_2} b^B c_1^{C_1} c_2^{C_2}$. In fact, given a word written in these generators, we can rearrange the letters as we did in the Heisenberg group in Section 1.2. The elements c_1 and c_2 are central and so can always be pushed to the end of the word. Rearranging the a_1, a_2, and b letters will cost us more b, c_1, c_2, and their inverses. But in the end, we arrive at the normal form.

Understanding the multiplication formula in these coordinates will give intuition into why nilpotent groups have polynomial growth. Even in the relatively simple case of $G = N_{3,2}$ the multiplication formula becomes quite complicated. We have

$$\begin{pmatrix} A_1 \\ A_2 \\ B \\ C_1 \\ C_2 \end{pmatrix} \cdot \begin{pmatrix} A_1' \\ A_2' \\ B' \\ C_1' \\ C_2' \end{pmatrix} = \begin{pmatrix} A_1 + A_1' \\ A_2 + A_2' \\ B + B' - A_2 A_1' \\ C_1 + C_1' + BA_1' - A_2\left(\frac{A_1'(A_1'+1)}{2}\right) \\ C_2 + C_2' + BA_2' - A_1'\left(A_2 A_2' + \left(\frac{A_2(A_2+1)}{2}\right)\right) \end{pmatrix}.$$

Observe that in the first two coordinates (corresponding to the first level of the Mal'cev coordinates), the multiplication formula is linear. In B, coming from the second level of coordinates, the formula is quadratic. In the last two coordinates (corresponding to the third level), multiplication is cubic. This pattern, which generalizes to the Mal'cev coordinates of all finitely generated nilpotent groups, will be crucial when exploring the growth of nilpotent groups.

Exercise 8. Mal'cev coordinates can help us to better understand higher Heisenberg groups. For example, consider the Heisenberg group $H_5(\mathbb{Z})$ of dimension 5. An element of $H_5(\mathbb{Z})$ is of the form

$$\begin{pmatrix} 1 & * & * & * \\ 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

with integer entries. What are the Mal'cev coordinates of $H_5(\mathbb{Z})$? What does group multiplcation look like in these coordinates?

In the examples above, the Mal'cev coordinates created a correspondence between $H(\mathbb{Z})$ and $\mathbb{Z}^3$, and between $N_{3,2}$ and $\mathbb{Z}^5$, where $\mathbb{Z}^3$ and $\mathbb{Z}^5$ had multiplication laws that were polynomial in each coordinate. In general, the Mal'cev coordinates of a finitely generated nilpotent group G will create a correspondence between G and $\mathbb{Z}^n \times \mathbb{Z}_{\ell_1} \times \cdots \times \mathbb{Z}_{\ell_t}$, where $\mathbb{Z}_{\ell_1} = \mathbb{Z}/\ell_1\mathbb{Z}$. Again the group multiplication on this product will be polynomial in each coordinate. This pattern is, in fact, characteristic of nilpotent groups, as we will see below

2.4. **Mal'cev completion.** As stated above, the Mal'cev coordinates for a finitely generated nilpotent group G provide a group structure on $\mathbb{Z}^n \times \mathbb{Z}_{\ell_1} \times \cdots \times \mathbb{Z}_{\ell_t}$. Let $\mathbb{R}_{\ell_i} = \mathbb{R}/\ell_1\mathbb{Z}$. We can endow

$$\mathbb{R}^n \times \mathbb{R}_{\ell_1} \times \cdots \times \mathbb{R}_{\ell_t}$$

with the same group multiplication from the Mal'cev coordinates. Doing so gives us a simply-connected, nilpotent Lie group, called the *Mal'cev completion*, in which our group G embeds cocompactly.

2.5. **Application: Solving single equations.** Mal'cev coordinates can be used to get results on the decidability of equation problems in groups. The equation problem studies the solvability of equations of the form $w = 1$, where w is a word written as a product of constants (group elements) and variables that can range over group elements. If the problem is decidable, then there exists an algorithm for taking any single equation and answer either YES or NO to the question of whether solutions exist.

Using Mal'cev coordinates, it can be shown that in $H(\mathbb{Z})$, the equation problem is decidable. This can be generalized to all 2-step nilpotent groups with rank-one commutators. One can also ask if it is possible to solve a system of equations simultaneously. As long as $s, m \geqslant 2$ (making our group is non-abelian), the problem of solving systems of equations in $N_{s,m}$ is not decidable.

How can we show that the problem for single equations is decidable? If we want a word w to be equal to 1, first we can put w in it's Mal'cev normal form. Then we ask each of the exponents to be equal to (or congruent to) 0. This will give us a finite set of linear equations, a finite set of congruences, and one quadratic equation. This type of system of equations with congruences is decidable.

See [5] for more details.

3. Polynomial growth

3.1. **Distortion of a subgroup in a group.** Let (G, S) be a finitely generated group with generating set S and $(H, T) \leqslant (G, S)$ a finitely generated subgroup with generating set T. We can assume that $T \subset S$, and hence the Cayley graph of (H, T) is a subgraph of the Cayley graph of (G, S). Let d_S and d_T be the respective word metrics on G and H.

Define the *distortion* of H in G to be

$$\Delta_G^H(n) = \max\{d_T(1, h) \mid h \in H, d_S(1, h) \leqslant n\}.$$

The subgroup H is said to be *undistorted* in G if $\Delta_G^H(n) \asymp n$.

3.1.1. *Subgroup distortion in the Heisenberg group:* Consider the Heisenberg group $H(\mathbb{Z})$. We showed above that $H(\mathbb{Z})$ has the normal form $a^A b^B c^C$ via the Mal'cev coordinates. Since $[H(\mathbb{Z}), H(\mathbb{Z})] = \langle c \rangle$, the lower central series is

$$H(\mathbb{Z}) \trianglelefteq \langle c \rangle \trianglelefteq \{1\}.$$

Let's look at the distortion of $G_2 = \langle c \rangle$ in $H(\mathbb{Z})$. Let $S = \{a, b, c\}$ be the generating set of $H(\mathbb{Z})$ and $T = \{c\}$ the generating set of $\langle c \rangle$. A simple exercise shows that

$$c^{n^2} = [a, b]^{n^2} = [a^n, b^n].$$

Thus $d_S(c^{n^2}) = 4n$, while $d_T(c^{n^2}) = n^2$. This implies that $\Delta_{H(\mathbb{Z})}^{\langle c \rangle}(n) \geq n^2$. Now we just have to show that the distortion can be no worse than quadratic. Here we can use the bounds on word length from Section 1.3. We said that for any generating set S, the word length of a word in Mal'cev coordinates was bounded above and below by multiples of

$$|A| + |B| + \sqrt{|C|}.$$

Thus if a sequence of word grows linearly in $|\cdot|_S$, then A and B can grow at most linearly, and C can grow at most quadratically. A sequence of words in $G_2 = \langle c \rangle$ can, therefore, grow at most quadratically while geodesic word length grows linearly. This proves that the subgroup distortion is quadratic.

We won't prove it here, but the example of the Heisenberg group follows a pattern that generalizes to all nilpotent groups. In general, we have that

$$\Delta_G^{G_i}(n) \asymp n^i.$$

So if, for example, your group is 3-step nilpotent like $G = N_{3,2}$, then $G_2 = [G, G]$ is quadratically distorted. Meanwhile, G_3 is cubically distorted in G.

3.2. **Growth function.** Let G be a group generated by the finite symmetric set S (i.e., $S = S^{-1}$) and $|\cdot|_S$ the word length. As in the first section of the notes, we define the *growth function* of G to be

$$\beta(n) = |B_n| = \#\{g \in G \bigm| |g|_S \leqslant n\}.$$

In this section we want to study the asymptotic behavior of this growth function for nilpotent groups. If $\beta(n)$ grows polynomially (or exponentially), we say that the group G has polynomial (or exponential) growth.

In Section 1.3.2, we showed that for $H(\mathbb{Z})$, $\beta(n)$ grows like n^4. To do so, we used the Mal'cev coordinates and bounds on word length by multiples of $|A| + |B| + \sqrt{|C|}$. Now we'll show that the polynomial growth is characteristic of nilpotent groups.

3.3. **History of polynomial growth.** Let $d(G) = \sum_{i \geqslant 1} i \operatorname{rk}(G_i/G_{i+1})$ denote the *homogeneous degree* of the nilpotent group G. Wolf showed in 1968 [**?**, wolf1968]hat there are constants K_1, $K_2 > 0$ such that

$$K_1 n^d \leqslant \beta(n) \leqslant K_2 n^e,$$

for all $n \gg 1$, where $d = d(G)$ and $e = \sum_{i \geqslant 1} 2^{i-1} \operatorname{rk}(G_i/G_{i+1})$. This tells us that the growth function is polynomial but does not specify the degree of polynomial growth.

Then in the early 1970s, Bass [1] and Guivarc'h [7] separately found that it suffices to let $e = d$. Thus we get the following proposition.

Proposition 6. (Bass-Guivarc'h formula) The degree of polynomial growth of a finitely generated nilpotent group is equal to

$$d(G) = \sum_{i \geqslant 1} i \operatorname{rk}(G_i/G_{i+1}).$$

So all finitely generated nilpotent groups have polynomial growth. Is the converse true? The answer turns out to be yes (in a sense). In 1981 Gromov [6] proved the following theorem.

Theorem 7. If a finitely generated group G has polynomial growth, then G is *virtually nilpotent*, i.e., G contains a nilpotent subgroup of finite index.

So polynomial growth, in a sense, is characteristic of nilpotent groups.

3.4. **Polynomial growth via distortion.** Subgroup distortion allows us to sketch a proof of the polynomial growth of nilpotent groups. Let G be a nilpotent group of class s. Then we lower central series of G is of the form

$$G = G_1 \trianglelefteq G_2 \trianglelefteq \ldots \trianglelefteq G_k \trianglelefteq G_{s+1} = \{1\}$$

Recall from above that we claimed to know the order of distortion of G_i inside of G. This is intimately related to the fact that the multiplication for Mal'cev coordinates adapted to the i^{th} level of the lower central series is polynomial of degree i. We claimed that

$$\Delta_G^{G_i}(n) \asymp n^i.$$

This pattern allows us to place bounds on word length using the Mal'cev coordinates as we did in Exercise 2. Recall from the construction of the Mal'cev coordinates that at level i we added as coordinates the generators of the abelian group G_i/G_{i+1}. Asymptotically, the generators of the torsion parts of this abelian group will not affect the growth function, so for our purposes, we need only consider the generators of the nontorsion part. There are $\operatorname{rk}(G_i/G_{i+1})$ nontorsion coordinates added at this i^{th} level of the lower central series. If we believe that G_i is n^i-distorted in G, then the "directions" associated to each of the $\operatorname{rk}(G_i/G_{i+1})$ generators will be distorted to this degree. This implies that word length of a group element will be bounded between linear combinations of the i^{th}-roots of the Mal'cev coordinates from level i. So as the geodesic word length grows linearly, these coordinates can grow like n^i. The growth function, therefore, grows polynomially with degree equal to the homogenous dimension of G, $\sum_{i \geqslant 1} i \operatorname{rk}(G_i/G_{i+1})$.

4. Carnot-Carathéodory metrics and coarse vs fine geometry

In this section we will define Carnot-Carthéodory metrics on $H(\mathbb{R})$. These constructions generalize to all *Carnot groups*: stratified, nilpotent, homogeneous Lie groups. Here, though, we will focus on the Heisenberg group. For further reading, we suggest [3], [10].

4.1. **Horizontal subspace.** Above we mentioned that $H(\mathbb{R})$ is a nilpotent Lie group, but we never defined what that meant. A *nilpotent Lie group* is a connected Lie group G whose Lie algebra $\mathfrak{g}$ is nilpotent. That is, its Lie algebra lower central series

$$\mathfrak{g}_1 = \mathfrak{g},\ \mathfrak{g}_2 = [\mathfrak{g},\mathfrak{g}]\ ,\ldots,\ \mathfrak{g}_{i+1} = [\mathfrak{g}_i,\mathfrak{g}],\ \ldots$$

eventually terminates in $\mathfrak{g}_{s+1} = 0$. (Note the similarities between this Lie algebra lower central series and the group lower central series.) This lower central series allows us to associate to $\mathfrak{g}$ a graded Lie algebra,

$$\mathfrak{g}_\infty = \oplus_{i\geqslant 1}\, \mathfrak{g}_i/\mathfrak{g}_{i+1}.$$

We endow this algebra with the unique Lie bracket $[\cdot,\cdot]_\infty$ with the property that for $X \in \mathfrak{g}_i$ and $Y \in \mathfrak{g}_j$, the bracket is defined modulo $\mathfrak{g}_{i+j+1}$, as

$$[X,Y]_\infty = \overline{[X,Y]}.$$

Example 8. Let's take another look at our favorite example, the Heisenberg group. The Lie algebra, i.e., the tangent space at the identity, is generated by the three vectors

$$X_1 = \begin{pmatrix} 0&1&0\\0&0&0\\0&0&0\end{pmatrix},\quad X_2 = \begin{pmatrix} 0&0&0\\0&0&1\\0&0&0\end{pmatrix},\quad X_3 = \begin{pmatrix} 0&0&1\\0&0&0\\0&0&0\end{pmatrix}.$$

One can check that $[X_1, X_2] = X_3$ and $[X_1, X_3] = [X_2, X_3] = [X_3, X_3] = 0$. Therefore, we have the grading

$$\mathfrak{g}_\infty = \langle X_1, X_2\rangle \oplus \langle X_3\rangle.$$

The first factor of this grading plays a special role, and we'll call it the *horizontal subspace*, denoted by $\mathfrak{m}$. We'll abuse notation and think of $\mathfrak{m}$ as both the horizontal subspace in the Lie algebra as well as the horizontal xy-plane in $H(\mathbb{R}) \cong \mathbb{R}^3$. Let $\pi_\mathfrak{m}(x_1, x_2, x_3) = (x_1, x_2)$ denote the orthogonal projection to $\mathfrak{m}$. Since we are in a Lie group we can push this subspace $\mathfrak{m}$ at the identity to the tangent space at any point via the differential. For each $p \in H(\mathbb{R})$, we can push $\mathfrak{m}$ to $\mathfrak{m}_p = \langle dL_p(X_1), dL_p(X_2)\rangle$, giving us a plane field in $H(\mathbb{R})$.

A curve $\gamma \in H(\mathbb{Z})$ is called *admissible* if for all t,

$$\gamma'(t) \in \mathfrak{m}_{\gamma(t)}.$$

In words, a curve is admissible if its tangent vectors lie in the plane field for all time t.

Lemma 9. (Balayage lemma) Given a path $\gamma = (\gamma_1, \gamma_2)$ from $(0, 0)$ to (x, y) in the horizontal plane $\mathfrak{m}$, there is a unique admissible curve (a *Legendrian lift*) $\overline{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ starting at the origin that projects down to γ under $\pi_\mathfrak{m}$. The lifted curve connects $(0, 0, 0)$ to the point (x, y, z) where z is equal to the signed area of the region enclosed by γ and a straight chord from $(0, 0)$ to (x, y) in the horizontal plane.

Proof. Let $\gamma = (\gamma_1, \gamma_2) : [0, 1] \to \mathfrak{m}$ be a curve in the horizontal subspace. Use a straight line to connect $\gamma(1)$ to the origin, and call the closed loop $\hat{\gamma}$. Let S be the region enclosed by the curve $\hat{\gamma}$. If we want $\overline{\gamma}$ to be admissible and if we want $\pi_\mathfrak{m}(\overline{\gamma}) = \gamma$, its derivative must be of the form

$$\overline{\gamma}' = \langle \gamma_1', \gamma_2', \frac{1}{2}\left(\gamma_1\gamma_2' - \gamma_2\gamma_1'\right)\rangle.$$

A direct computation shows that the straight line we used to close off γ does not affect the integral of $\gamma_3' = \frac{1}{2}(\gamma_1\gamma_2' - \gamma_2\gamma_1')$. Therefore, by Green's Theorem we have

$$\begin{aligned}\gamma_3 &= \int_0^1 \gamma_3'(t)dt \\ &= \frac{1}{2}\int_0^1 (\gamma_1\gamma_2' - \gamma_2\gamma_1')(t)dt \\ &= \frac{1}{2}\int_{\hat{\gamma}} x_1dx_2 - x_2dx_1 \\ &= \int\int_S dxdy.\end{aligned}$$

□

Note that since the third coordinate of an admissible path is given by the signed area enclosed, almost all of the data of the curve is preserved if we project any path in $H(\mathbb{R})$ via $\pi_{\mathfrak{m}} : H(\mathbb{R}) \to \mathfrak{m}$ to the horizontal plane $\mathfrak{m}$. What we lose is the initial height of the path since the plane field $\mathfrak{m}_p$ is invariant under vertical translations. Still, it will often be convenient to perform calculations on these "shadows" of paths instead of on the paths themselves.

It is also important to note that we can construct an admissible curve that connects any two points, say $(0, 0, 0)$ and $p = (p_1, p_2, p_3)$. Projecting to $\mathfrak{m}$, we need only connect $(0, 0)$ to (p_1, p_2) via a circular arc enclosing signed area equal to p_3. Then the admissible lift of this curve will connect the origin to our point p. See Figure 1 for a few examples. Observe that the red signed area A_1 is *negative* due to the orientation of the curve, so in this case p_3 would be negative. Meanwhile the paths enclosing A_2 and A_3 are positively oriented, making p_3 positive.

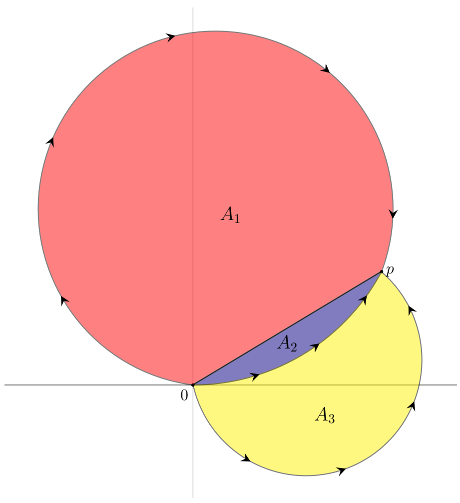


FIGURE 1. How to connect any two points with an admissible curve using circular arcs.

Exercise 9. The push-forward (differential) dL_p of left multiplication L_p in $H(\mathbb{R})$ is derived as follows:

$$L_p(x) = px = (p_1 + x_1, p_2 + x_2, p_3 + x_3 + \frac{1}{2}(p_1x_2 - p_2x_1)); \qquad (dL_p)_x = \begin{pmatrix} 1 & & \\ & 1 & \\ -\frac{p_2}{2} & \frac{p_1}{2} & 1 \end{pmatrix}$$

Let X_1, X_2 be the tangent vectors in the x, y directions at the identity. Then we get a vector field by

$$X_1(p) = dL_p(X_1) = \begin{pmatrix} 1 & & \\ & 1 & \\ -\frac{p_2}{2} & \frac{p_1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -\frac{p_2}{2} \end{pmatrix}. \qquad \text{Similarly,} \qquad X_2(p) = \begin{pmatrix} 0 \\ 1 \\ \frac{p_1}{2} \end{pmatrix}.$$

(Note that these are linearly independent but not orthogonal.) Sketch part of this plane field. Choose two different examples of smooth curves γ in $\mathfrak{m}$. Find their admissible lifts to $H(\mathbb{R})$ using the balayage lemma. For a point along each curve, check explicitly that the tangent vector $\gamma'(t)$ is in the plane $\mathfrak{m}_{\gamma(t)}$.

4.2. **Carnot-Carthéodory metric.** We are now ready construct our Carnot-Carthéodory metrics on $H(\mathbb{R})$. We'll start with any norm $||\cdot||$ on the horizontal subspace $\mathfrak{m}$. Given an admissible path $\gamma = (\gamma_1, \gamma_2, \gamma_3) : [0,1] \to H(\mathbb{R})$, we can calculate its *Minkowski length*

$$L(\gamma) = \int_0^1 ||\pi_{\mathfrak{m}}(\gamma'(t))||dt = \int_0^1 ||(\gamma_1'(t), \gamma_2'(t))||dt.$$

Again, we recall that by projecting an admissible curve to $\mathfrak{m}$ via $\pi_{\mathfrak{m}}$ we do not lose information about the length of the curve. Then we define the distance $d(p,q)$ between p and q to be the infimum of the Minkowski lengths of admissible paths connecting p to q. This CC metric depends only on the norm $||\cdot||$ chosen on $\mathfrak{m}$. It is a sub-Finsler metricnsince non-admissible paths are not given lengths, and lengths are defined using a norm which may or may not be induced by an inner product as in the sub-Riemannian case.

Exercise 10. Let $\mathfrak{m} = \langle X_1, X_2 \rangle$ and $\mathfrak{m}_p = \langle X_1(p), X_2(p) \rangle$. Note that $\mathfrak{m}_p$ sits in the $\mathbb{R}^3$ space coordinatizing $H(\mathbb{R})$ as $p + \langle X_1(p), X_2(p) \rangle$. Check that the length of an arbitrary vector in $\mathfrak{m}$ is the same as the length of the projection to $\mathfrak{m}$ of its push-forward in $\mathfrak{m}_p$. Explain how to interpret this as saying that admissible curves in $H(\mathbb{R})$ for any CC metric can have their lengths computed just from their shadows in the xy-plane.

4.3. **The Heisenberg similarity.** Recall from the first section that we have a dilation $\delta_n : H(\mathbb{Z}) \to H(\mathbb{Z})$ taking $(x, y, z) \mapsto (nx, ny, n^2 z)$. While for the word metric on $H(\mathbb{Z})$ this was a near-similarity, in the real Heisenberg group with a CC metric, the map δ_n is a similarity.

Proposition 10. For $H(\mathbb{R})$ with any CC metric, the maps δ_n are similarities. In other words,

$$d_{\mathrm{CC}}(\delta_n(p), \delta_n(q)) = n \cdot d_{\mathrm{CC}}(p, q).$$

To prove this proposition, we'll first need a lemma.

Lemma 11. If γ is an admissible curve between p and q, then $\delta_n(\gamma)$ is an admissible curve between $\delta_n(p)$ and $\delta_n(q)$.

Proof. Let $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ be an admissible curve. Then $\delta_n(\gamma) = (n\gamma_1, n\gamma_2, n^2\gamma_3)$. The differential $d\delta_n$ acts on γ' as follows:

$$d\delta_n(\gamma') = (n\gamma_1', n\gamma_2', n^2\gamma_3').$$

We showed above that admissible curves γ satisfy

$$\gamma_3' = \frac{1}{2}(\gamma_1\gamma_2' - \gamma_2\gamma_1'). \tag{1}$$

Since γ is admissible, it follows that

$$\begin{aligned} n^2\gamma_3' &= n^2\frac{1}{2}(\gamma_1\gamma_2' - \gamma_2\gamma_1') \\ &= \frac{1}{2}((n\gamma_1)(n\gamma_2') - (n\gamma_2)(n\gamma_1')). \end{aligned}$$

Thus $d\delta_n(\gamma')$ also satisfies equation (1), making $\delta_n(\gamma)$ admissible. □

Proof. (of Proposition 10). Suppose γ is an admissible curve. Lemma 11 tells us that $\delta_n(\gamma)$ is an admissible curve between $\delta_n(p)$ and $\delta_n(q)$. If the Minkowski length of $\delta_n\gamma$ is $nL(\gamma)$, then we are done. Indeed,

$$\begin{aligned} L(\delta_n(\gamma)) &= \int_0^1 ||(n\gamma_1'(t), n\gamma_2'(t))||dt \\ &= \int_0^1 n \cdot ||(\gamma_1'(t), \gamma_2'(t))||dt \\ &= nL(\gamma). \end{aligned}$$

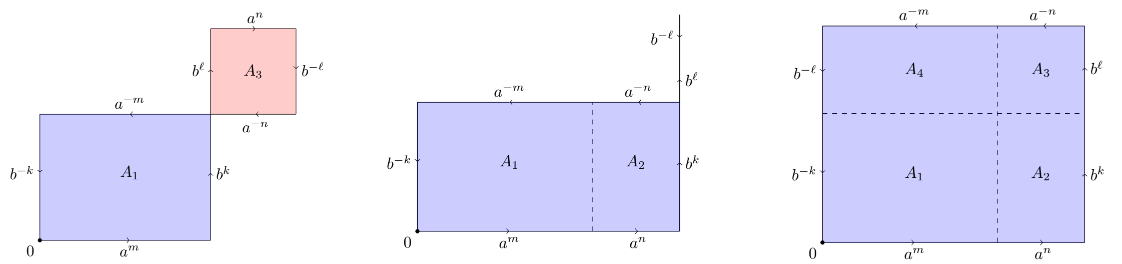
□

4.4. **Commutator calculus and area.** Here we will take a look at how the so-called "commutator calculus" relates to area in the Heisenberg group $H(\mathbb{Z})$. Recall (or check) that in $H(\mathbb{Z})$,

$$[a^m, b^n] = a^m b^n a^{-m} b^{-n} = c^{mn}.$$

We can use this relationship to adjust the projections of paths to the horizontal plane without losing information about height of the lifted path.

Let a and b be the standard generators of $H(\mathbb{Z})$. Starting at the origin 0, follow the path represented by the word $w = a^m b^{k+\ell} a^n b^{-\ell} a^{-(m+n)} b^{-k}$ as in Figure 2a. We can view this path as living in $\mathfrak{m} = \mathbb{R}^2 \subset H(\mathbb{R})$, which we'll endow with the L^1 norm. We know that this path has a unique admissible lift and that the height of the lift at any point should be given by signed area enclosed by the shadow so far. Therefore, although we have a closed loop in the shadow, the lift in general is *not* a loop. The balayage lemma tells us that at the end of the path when we arrive again at the origin in the shadow, the z-coordinate of the lift should be the difference of the areas $A_1 - A_3$, or $mk - nl$. (Note that we travel around A_3 clockwise, and so its signed area is *negative*.)



(a) Start with a curve in the xy-plane.

(b) Replace $b^{k+\ell}a^n$ with $a^n b^{k+\ell} c^{-n(k+\ell)}$.

(c) Replace $b^{-\ell}a^{-(m+n)}$ with $a^{-(m+n)}b^{-\ell}c^{-\ell(m+n)}$.

FIGURE 2. How commutator algebra relates to area in the plane. Blue corresponds to positive signed area, and red is negative.

We can use commutator algebra to change our path from Figure 2(a) into something more familiar. Our goal is to get the rectangular path in Figure 2(c), since we know that

$$a^{m+n} b^{k+\ell} a^{-(m+n)} b^{-(k+\ell)} = [a^{m+n}, b^{k+\ell}] = c^{(m+n)(k+\ell)}.$$

So traveling around the lift of the rectangular loop is the same as traveling vertically $(m+n)(k+\ell)$ units. In this rectangular case, we can forget about the CC metric and area, instead using the commutator relation to algebraically compute the height of our curve. To morph w into the rectangular path, we'll use the same manipulations we did when putting words into their Mal'cev coordinates, remembering that $ba = abc^{-1}$. First we'll move a^n across $b^{k+\ell}$. This move is not free, though! It costs us $c^{-n(k+\ell)}$ to make this move. Making this switch is telling our path to do a^n first and then $b^{k+\ell}$. Figure 2(b) shows what this new path looks like. (Note that in the shadow we don't see the effect of c.) Now our word is

$$w = a^{m+n} b^{k+\ell} b^{-\ell} a^{-(m+n)} b^{-k} c^{-n(k+\ell)}.$$

Next, we'll want to do travel along $a^{-(m+n)}$ before $b^{-\ell}$. Making this switch algebraically costs us $c^{-\ell(m+n)}$. So we see that w is

$$\begin{aligned} w &= a^{m+n}b^{k+\ell}a^{-(m+n)}b^{-(k+\ell)}c^{-n(k+\ell)-\ell(m+n)} \\ &= [a^{m+n}, b^{k+\ell}]c^{-n(k+\ell)-\ell(m+n)} \\ &= c^{(m+n)(k+\ell)}c^{-n(k+\ell)-\ell(m+n)} \\ &= c^{mk-n\ell}. \end{aligned}$$

Note that we get back the height/area we computed above, $mk - n\ell$! Between Figures 2(a) and 2(b), we added $A_2 = nk$ and $A_3 = n\ell$ to the signed area, and to compensate for it, we had to multiply our word by $c^{-nk-n\ell}$. Likewise between Figures 2(b) and 2(c), we added $A_3 = n\ell$ and $A_4 = mk$ to the signed area and in return had to multiply our word by c^{-nl-mk}.

In the end, the algebraic manipulation may not be the easier way to compute the height of an admissible life, but this example allows us to see how the commutator, the algebra, and the geometry are related.

5. Asymptotic geometry

5.1. **Asymptotic cone, Pansu's theorem, Krat's theorem.** In this section, we ask about the large-scale geometry of finitely generated nilpotent groups. Let Γ be a finitely generated nilpotent group with symmetric generating set $S = S^{-1}$. As we zoom out on the group and view its Cayley graph as a finer mesh, what does it look like? Pansu showed in [9] that the Cayley graphs of Γ with word metrics scaled by $\frac{1}{n}$ converge in the pointed Gromov-Hausdorff topology to a connected, simply-connected nilpotent Lie group with a left-invariant Carnot-Carthéodory metric (G_∞, d_∞). Before precisely stating the theorem, we should say something about the construction of (G_∞, d_∞), the *asymptotic cone* of a finitely generated torsion-free nilpotent group.

Let Γ be a finitely generated, torsion-free nilpotent group with symmetric generating set S. In Section 2.4, we introduced the Mal'cev completion, the simply-connected nilpotent Lie group G in which Γ embeds cocompactly. Let $\pi_{\mathfrak{m}} : G \to \mathfrak{m} = \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ be the projection to the horizontal subspace as defined in the previous section. Since Γ embeds cocompactly in G, it must be the case that $\pi_{\mathfrak{m}}(\Gamma)$ forms a discrete full-rank lattice inside of $\mathfrak{m}$. Let P be the polyhedron formed from the convex hull of $\pi_{\mathfrak{m}}(S)$. Since S is symmetric, P is symmetric about the origin. Also, it has full dimension inside $\mathfrak{m}$ since it generates the lattice formed by $\pi_{\mathfrak{m}}(\Gamma)$. Therefore we can define a norm $||\cdot||$ on $\mathfrak{m}$ by taking P to be the unit ball.

Recall that $\mathfrak{m} = \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ is the first summand in the graded Lie algebra $\mathfrak{g}_\infty$ constructed from $\mathfrak{g}$. Let G_∞ be the group associated to $\mathfrak{g}_\infty$. Then we can use the norm $||\cdot||$ to put a sub-Finsler Carnot-Carthèodory metric d_∞ on G_∞. We'll call this space the asymptotic cone of Γ.

Example 12. Let $\Gamma = H(\mathbb{Z})$ with generating set $S = \{a, b, a^{-1}, b^{-1}\}$. Then the Mal'cev completion of Γ is $G = H(\mathbb{R})$. The horizontal subspace $\mathfrak{m}$ at the origin is the xy-plane, as shown in Example 8. The projection map $\pi_{\mathfrak{m}}$ sends a to $X_1 = (1, 0)$, sends b to $X_2 = (0, 1)$, and sends their inverses to $-X_1$ and $-X_2$, respectively. The convex hull, therefore is the unit diamond. Therefore, the asymptotic cone of $H(\mathbb{Z})$ with the standard generating set is $H(\mathbb{R})$ with the CC metric induced by the L^1 norm on $\mathbb{R}^2$.

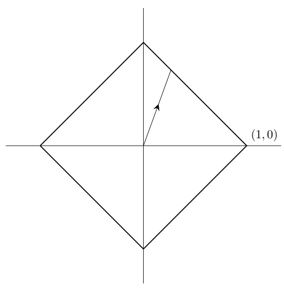


FIGURE 3. The L^1 unit circle.

Now we are ready to state Pansu's theorem.

Theorem 13. (Pansu) Let Γ be a finitely generated, torsion-free nilpotent group with generating set S. Let (G_∞, d_∞) be the asymptotic cone of Γ as defined above. Then

$$\lim_{|x|\to\infty} \frac{d_S(1, x)}{d_\infty(1, x)} = 1.$$

Restricting to the Heisenberg group, we can get an even stronger statement from Krat [8].

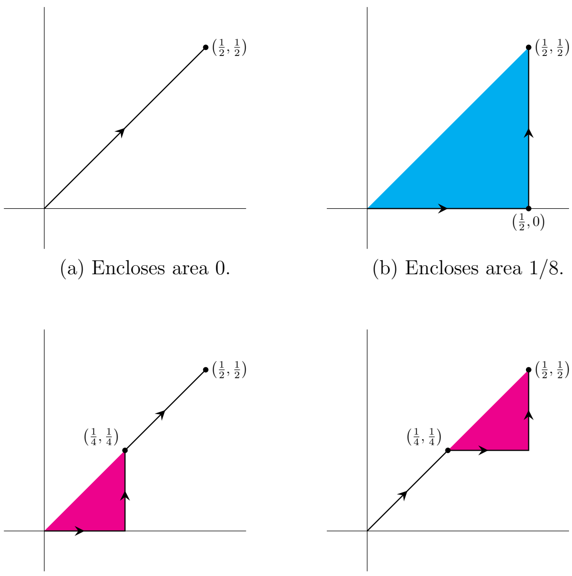
Theorem 14. (Krat) For the Heisenberg group $H(\mathbb{Z})$ and any generating set S, there exists a constant C dependent on S such that for all $x \in H(\mathbb{Z})$,

$$\left| \, d_S(1, x) - d_\infty(1, x) \, \right| \leqslant C.$$

Krat's theorem says not only that the distances have similar asymptotic growth, but that they must stay bounded distance from one another for all x. Krat proved this is true for $H(\mathbb{Z})$ and for word hyperbolic groups. It is not true for nilpotent groups in general. In particular, Breuillard and Le Donne show in [2] show in that Krat's Theorem does not hold for $H(\mathbb{Z}) \times \mathbb{Z}$.

5.2. **The unit ball and the isoperimetric problem.** Let's consider what the unit ball in $H(\mathbb{R})$ is with the CC metric induced by the standard generators $\{a, b\}^{\pm}$. Pansu's theorem tells us that this generating set induces the L^1 norm on $\mathfrak{m}$. For starters, we have the L^1 unit circle (diamond) in the horizontal plane. The straight lines from the origin to the unit diamond have length 1, do not enclose any area, and hence their lifts in $H(\mathbb{R})$ remain in the horizontal plane.

But we know that these straight lines are not the only geodesics with respect to the L^1 norm. In the horizontal plane, as long as we do not backtrack in any direction, we remain geodesic. In Figure 4, we see four different geodesics from $(0, 0)$ to $(1/2, 1/2)$. The first three curves (a), (b), and (c) enclose different signed areas, and so their lifts to $H(\mathbb{R})$ will have different endpoints at different heights. The final two curves, (c) and (d), enclose the same amount of area, and so they are distinct geodesics between two points.



(a) Encloses area 0.

(b) Encloses area 1/8.

(c) Encloses area 1/32.

(d) Encloses area 1/32.

FIGURE 4. Various heights of the unit ball above the point $(1/2, 1/2)$.

By building staircases between $(0, 0)$ and $(1/2, 1/2)$ in different ways, we can achieve any height between 0 and 1/8. This tells us that the unit sphere in $(H(\mathbb{R}), d_{\mathrm{CC}})$ contains a vertical line segment through $(1/2, 1/2, 0)$. In fact, the unit ball contains vertical faces along each edge of the unit diamond. We can ask, though, what the maximal height above $(1/2.1/2)$ is. Of our examples in Figure 4, 1/8 is the maximal height. Is it possible to travel from $(0, 0)$ to $(1/2, 1/2)$ geodesically while enclosing area *more than* 1/8?

We're fixing the length of our line and asking how much area we can enclose. Suddenly we're talking about the isoperimetric problem for L^1! Fortunately, it is known that the square is the isoperimetrix for the L^1 norm in $\mathbb{R}^2$. So a geodesic curve of maximal height in $H(\mathbb{R})$ will be one that follows the shape of a square. The curve in 4(b) does just that. Therefore, 1/8 is indeed the maximal height of the unit ball above $(1/2, 1/2)$.

Now we can look at different points within the L^1 unit ball and ask what height (or heights) the unit ball achieves above it (worrying only about positive heights since the ball will be symmetric across the

xy-plane). In Figure 5, we draw one geodesic of unit length between $(0,0)$ and $(\frac{1}{2},0)$. This is, indeed, the *only* unit-length geodesic connecting these two points. Any other positive area less than $\frac{1}{8}$ can be enclosed by a rectangular curve of length less than 1. So $(\frac{1}{2},0,\frac{1}{8})$ is the unique point on the unit ball above $(\frac{1}{2},0)$.

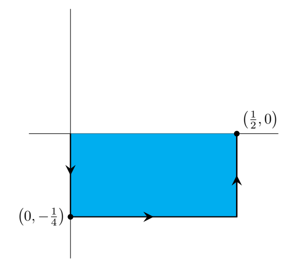


FIGURE 5. Finding the height of the unit ball above the point $(1/2,0)$.

Exercise 11. Let's explore the large-scale geometry in the standard generators on $H(\mathbb{Z})$, which induces the L^1 norm on $\mathfrak{m}$. For the three points $(0,0)$, $(1/3,1/3)$, and $(2/3,0)$ in the normed plane $(\mathfrak{m},L^1)$, compute the height of the CCF unit sphere over each point. Above we showed that the maximum height over $(1/2,1/2)$ was $1/8$ and that the unique height over $(1/2,0)$ was also $1/8$. Using this and the Heisenberg similarity, find the distance from the origin to $(0,0,z)$ in the CC metric. Compare that to the word length of a long central word in the discrete group. If Pansu's theorem is right, they should be very close to the same. (Hint: it's right.)

Exercise 12. Which geodesics based at 0 are prolongable in this CCF metric? Which are infinitely prolongable (i.e., what are the geodesic *rays* based at the origin)? Which geodesics are uniquely prolongable? For what $p \in H(\mathbb{R})$ is there a unique geodesic $\overline{0p}$? (Note these are four different questions with quite different levels of difficulty, so say what you can.)

References

[1] H. Bass. The degree of polynomial growth of finitely generated nilpotent groups. *Proc. London Math. Soc. (3)*, 25:603–614, 1972.

[2] Emmanuel Breuillard and Enrico Le Donne. On the rate of convergence to the asymptotic cone for nilpotent groups and subfinsler geometry. *Proceedings of the National Academy of Sciences*, 110(48):19220–19226, 2013.

[3] Luca Capogna, Donatella Danielli, Scott D Pauls, and Jeremy Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, volume 259. Springer Science & Business Media, 2007.

[4] Cornelia Druţu and Michael Kapovich. *Geometric group theory*, volume 63. American Mathematical Soc., 2018.

[5] Moon Duchin, Hao Liang, and Michael Shapiro. Equations in nilpotent groups. *Proceedings of the American Mathematical Society*, 143(11):4723–4731, 2015.

[6] Mikhael Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.

[7] Yves Guivarc'h. Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, 101:333–379, 1973.

[8] S. A. Krat. Asymptotic properties of the Heisenberg group. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 261(Geom. i Topol. 4):125–154, 268, 1999.

[9] Pierre Pansu. Croissance des boules et des géodésiques fermées dans les nilvariétés. *Ergodic Theory Dynam. Systems*, 3(3):415–445, 1983.

[10] Ayla Sánchez. *A theory of sub-Finsler surface area in the Heisenberg group*. Tufts University, PhD Thesis, 2017.