Carnot-Carathéodory and Korányi-Geodesics in the Heisenberg Group

Josh Ascher* , Armin Schikorra

Josh Ascher is an undergraduate student studying math and computer science at the University of Pittsburgh. He is in his senior year and plans to attend graduate school where he will conduct research in theoretical computer science. Josh hopes to one day become a professor so that he can continue pursuing research, as well as teaching the next generation of researchers.

Armin Schikorra received his Ph.D. from RWTH Aachen University in 2010 and is currently associate professor at the University of Pittsburgh. He works in partial differential equations which are often motivated from the geometric calculus of variations, one example being harmonic maps between manifolds and various generalizations thereof. In particular he is interested in regularity theory of such local or nonlocal equations.

Abstract

We discuss the Heisenberg group $\mathbb{H}_1$, the three-dimensional space $\mathbb{R}^3$ equipped with one of two equivalent metrics, the Korányi- and Carnot-Carathéodory metric. We show that the notion of length of curves for both metrics coincide, and that shortest curves, so-called geodesics, exist.

1   Introduction

The Heisenberg group $\mathbb{H}_1$ is a subject of intensive study, as a special case of sub-Riemannian manifolds or Carnot groups, see [2] or [1].

From the point of view of Analysis, $\mathbb{H}_1$ consists of all the points $p = (p^1, p^2, p^3) \in \mathbb{R}^3$, where $\mathbb{R}^3$ denotes the usual Euclidean three-dimensional space. However, the distance

*Corresponding author: joa71@pitt.edu

between two points $p, q \in \mathbb{R}^3$ is given by a non-Euclidean metric $d(p,q)$. Actually, there are two typical metrics used in the Heisenberg group $\mathbb{H}_1$, and we begin by describing the first one, the *Carnot-Carathéodory*-metric $d_{cc}(p,q)$ of $\mathbb{H}_1$: Take any (for now continuously differentiable) curve $\gamma \colon [0,1] \to \mathbb{R}^3$ with $\gamma(0) = q$ and $\gamma(1) = p$. From calculus we know that the length of a curve is given by

$$\mathcal{L}(\gamma) = \int_{[0,1]} |\dot{\gamma}(t)| dt, \tag{1}$$

where $\dot{\gamma}$ denotes the derivative of γ. If we consider the minimal possible length of curves $\gamma \colon [0,1] \to \mathbb{R}^3$ that are continuously differentiable and connect p to q in the sense that $\gamma(0) = p$ and $\gamma(1) = q$, then one can show that this minimal length is exactly the Euclidean distance $|p - q|$,

$$|p - q|_{\mathbb{R}^3} = \inf_{\gamma \in X(p,q)} \mathcal{L}(\gamma)$$

where

$$X(p,q) = \left\{ \gamma \colon [0,1] \to \mathbb{R}^3 \colon \quad \text{continuously differentiable, } \gamma(0) = p, \gamma(1) = q \right\}.$$

The Carnot-Carathéodory metric is also the infimum of the lengths of curves connecting p and q, however those curves have to be *horizontal*, meaning that $\dot{\gamma}(t)$ has to belong to the horizontal space $H_{\gamma(t)}\mathbb{H}_1$ for each $t \in (0,1)$, which is spanned by the vectors

$$H_p\mathbb{H}_1 = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 2p^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2p^1 \end{pmatrix} \right\}.$$

That is for each $t \in (0,1)$ there must be some $\lambda_1(t)$ and $\lambda_2(t)$ such that

$$\dot{\gamma}(t) = \lambda_1(t) \begin{pmatrix} 1 \\ 0 \\ 2\gamma^2(t) \end{pmatrix} + \lambda_2(t) \begin{pmatrix} 0 \\ 1 \\ -2\gamma^1(t) \end{pmatrix},$$

or, taking $\lambda_1(t) = \dot{\gamma}^1(t)$ and $\lambda_2(t) = \dot{\gamma}^2(t)$, equivalently,

$$\dot{\gamma}^3(t) = -2\gamma^1(t)\dot{\gamma}^2(t) - \gamma^2(t)\dot{\gamma}^1(t) \quad \forall t \in (0,1). \tag{2}$$

For such curves we define the length

$$\mathcal{L}_{cc}(\gamma) := \int_{[0,1]} \sqrt{|\lambda_1(t)|^2 + |\lambda_2(t)|^2} dt \equiv \int_{[0,1]} \sqrt{|\dot{\gamma}_1(t)|^2 + |\dot{\gamma}_2(t)|^2} dt.$$

The Carnot-Carathéodory length $d_{cc}(p,q)$ is then given by

$$d_{cc}(p,q) = \inf_{\gamma \in Y(p,q)} \mathcal{L}_{cc}(\gamma) \tag{3}$$

where

$$Y(p,q) = \left\{ \gamma \colon [0,1] \to \mathbb{R}^3 \colon \quad \text{continuously differentiable, } \gamma(0) = p, \gamma(1) = q, \text{(2) holds} \right\}.$$

Observe that this is very similar to curves γ into a Riemannian manifold $\mathcal{M} \subset \mathbb{R}^3$: any differentiable curve $\gamma \colon [0,1] \to \mathcal{M}$ satisfies $\gamma(t) \in T_{\gamma(t)}\mathcal{M}$, where $T_p\mathcal{M}$ is the tangent space of the manifold $\mathcal{M}$, and if we want to find the distance between two points p

and q on the manifold, it makes sense to define this distance as the minimal length of curves tangent to the manifold at every point and connecting p and q. So from this perspective, the Heisenberg group is $\mathbb{R}^3$ with a "strange" tangent plane distribution (and since it is strange we call it horizontal plane distribution instead) – the strangeness of the Heisenberg group is that its horizontal plane distribution cannot be written as a tangent space of any manifold $\mathcal{M}$, that is the horizontal plane distribution is not integrable in the sense of the Frobenius' theorem. Here is actually where the "group" of the Heisenberg group enters, the vectors spanning the horizontal space $H_p\mathbb{H}_1$ are left-invariant vector fields for a group structure – but we will not pursue this point of view further here.

It is known that for each $p, q \in \mathbb{R}^3$ the infimum in (3) is attained, i.e. there exists a shortest curve γ, called *geodesic* such that

$$\mathscr{L}_{cc}(\gamma) = d_{cc}(p, q),$$

see e.g. Hajłasz-Zimmerman [3, (1.3)]. In particular between any two points $p, q \in \mathbb{R}^3$ there exist horizontal curves. Let us remark that for more general sub-Riemannian geometry it a very deep result, called Chow–Rashevskii theorem, that $d_{cc}(p, q)$ is even *finite* for all points p, q, cf. [4].

While the above notion of distance $d_{cc}(p, q)$ is attractive from a geometric point of view, it is not easily computable (given p and q we first need to find the shortest curve γ between then, then compute its length).

The other metric we want to consider, the *Korányi-metric*, is much easier to compute. It simply is given by

$$d_K(p, q) := \left(|p_1 - q_1|^2 + |p_2 - q_2|^2{}^2 + |p_3 - q_3 + 2(p_2 q_1 - p_1 q_2)|^2 \right)^{\frac{1}{4}}$$

There is also a more group-theoretic motivation for $d_K(p, q) = \|p^{-1} * q\|_{\mathbb{H}_1}$, but we will also not pursue this aspect further here, we refer the interested reader to [1].

Any metric space naturally is equipped with a notion of length of curves, see Definition 5, which gives us the notion of a Korányi-length $\mathscr{L}_K(\gamma)$.

We will first prove the following result.

Theorem 1. *Let $p, q \in \mathbb{R}^3$. Then there exists a shortest continuous curve (i.e. a geodesic) $\gamma : [0, 1] \to \mathbb{R}^3$, $\gamma(0) = p$, $\gamma(1) = q$ such that*

$$\mathscr{L}_K(\gamma) = \inf_{\tilde{\gamma} \in \tilde{X}(p,q)} \mathscr{L}_K(\tilde{\gamma}),$$

where

$$\tilde{X}(p, q) := \left\{ \gamma : [0, 1] \to \mathbb{R}^3 : \quad \text{continuous}, \ \gamma(0) = p, \ \gamma(1) = q \right\}.$$

Observe the difference to $X(p, q)$ above is that curves do not need to be differentiable.

The above theorem follows from a general principle using the Arzelá-Ascoli theorem and holds true in much more generality.

More specifically to the Heisenberg group we will show that although the metric d_K differs from d_{cc}, the Korányi-length $\mathscr{L}_K$ equals the Carnot-Carathéodory length $\mathscr{L}_{cc}$.

Theorem 2. *Let $\gamma \colon [0,1] \to \mathbb{R}^3$ be twice continuously differentiable. If γ is horizontal (i.e. (2) holds) and $\mathscr{L}_{cc}(\gamma) < \infty$ then $\mathscr{L}_K(\gamma) < \infty$ and we have*

$$\mathscr{L}_K(\gamma) = \mathscr{L}_{cc}(\gamma).$$

From Theorem 2 we actually can conclude that $(\mathbb{R}^3, d_K)$ is not a length space: By the definition of length of a curve in a metric space (X,d), see Definition 5, for any p, q and any curve $\gamma \colon [0,1] \to X$, $\gamma(0) = p$, $\gamma(1) = q$ we have the inequality

$$\mathscr{L}(\gamma) \geq d(p,q).$$

If for any $p, q \in X$ there exists a curve $\gamma \colon [0,1] \to X$, $\gamma(0) = p$, $\gamma(1) = q$ such that we have equality

$$\mathscr{L}(\gamma) = d(p,q),$$

then we call X a length space. The following example shows that $(\mathbb{R}^3, d_K)$ is not a length space (this is in contrast to the Carnot-Carathéodory metric where the corresponding equality holds by definition (3)).

Example 3. *The following is the shortest curve between $p := (0,0,0)$ and $q := (0,0,\frac{1}{4\pi})$*

$$\gamma(t) = \begin{pmatrix} (1 - \cos(2\pi t)) \\ \sin(2\pi t) \\ \frac{1}{4\pi}\left(t - \frac{\sin(2\pi t)}{2\pi}\right) \end{pmatrix}.$$

See [3, Theorem 2.1]. It can be checked by a direct computation that $\mathscr{L}_K(\gamma) = \mathscr{L}_{cc}(\gamma) > d_K(p,q)$

The outline of the remaining part of the paper is as follows: in Section 2 we discuss preliminary results on metric spaces, in particular Arzelá-Ascoli's theorem. In Section 3 we discuss properties of horizontal curves that we need for both theorems. In Section 4 we establish the existence of shortest curves with respect to $\mathscr{L}_K$ in the Heisenberg group. In Section 5 we prove Theorem 2. Let us remark that the results in this work are probably well-known to experts, the purpose of this paper is to provide a detailed account making this exciting field accessible to non-experts, students and early career researchers.

2   Some Preliminary Statements from Analysis: Metric Spaces

Let X be a metric space with metric d. A *curve* γ is simply a continuous map $\gamma \colon I \to X$, where $I = [a,b]$ is any closed finite interval.

We say that a curve $\gamma \colon [a,b] \to X$ connects two points $p, q \in X$ if $\gamma(a) = p$ and $\gamma(b) = q$.

We now want to define the length of a curve $\gamma \colon [a,b] \to X$, however observe that γ may not be differentiable. Indeed, we may not even know what differentiability of γ means since X is not a linear space! So a formula such as (1) does not make sense. But recall from Calculus how we obtained the formula (1), we used polygonal approximation of a curve. We will do the same in metric spaces.

Definition 4 (Partition)**.** Given an interval $[a,b]$, a partition of size n is the set $\{x_0, x_1, \ldots, x_n\}$ where

$$a = x_0 < x_1 < \cdots < x_n = b$$

With the notion of partition we can "approximate" curves by a discrete path through the points $\gamma(a), \gamma(x_1), \ldots, \gamma(b)$. Then we use the metric to define the length of these "polygon"-lines.

Definition 5 (Length of curve)**.** Given a metric space (X,d) and a curve $\gamma : [a,b] \to X$. The length of γ is given by

$$\mathscr{L}(\gamma) = \sup_{p \in P} \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})),$$

where the supremum is taken over all partitions p of $[a,b]$ (i.e. P is the collection of all partitions of $[a,b]$).

Observe that the length of a curve $\mathscr{L}(\gamma)$ is always nonnegative, indeed since $\{a,b\}$ is a partition of $[a,b]$, we have

$$\mathscr{L}(\gamma) \geq d(\gamma(a), \gamma(b)). \tag{4}$$

In general, even if $d(\gamma(a), \gamma(b)) < \infty$ the length $\mathscr{L}(\gamma)$ could be $+\infty$. We call any curve γ with finite length $\mathscr{L}(\gamma) < \infty$ *rectifiable*.

It is worth noting the following

Lemma 6. *Given a metric space* (X,d)*, let* $\gamma : [a,b] \to X$ *be a curve of finite length,* $\mathscr{L}(\gamma) < \infty$*. Then for any* $s_0 \in [a,b]$*, the restricted curves*

$$\gamma\Big|_{[s_0,b]} : [s_0,b] \to X, \quad [s_0,b] \ni t \mapsto \gamma(t)$$

and

$$\gamma\Big|_{[a,s_0]} : [a,s_0] \to X, \quad [a,s_0] \ni t \mapsto \gamma(t)$$

are curves of finite length. Moreover

$$[a,b] \ni s \mapsto \mathscr{L}\left(\gamma\Big|_{[a,s]}\right)$$

and

$$[a,b] \ni s \mapsto \mathscr{L}\left(\gamma\Big|_{[s,b]}\right)$$

are continuous monotone increasing maps.

Proof. Finiteness and monotonicity are easy to obtain from the definition of the curve. For the continuity, we observe that for $a \leq s_1 < s_2 \leq b$

$$\mathscr{L}\left(\gamma\Big|_{[a,s_2]}\right) - \mathscr{L}\left(\gamma\Big|_{[a,s_1]}\right) = \mathscr{L}\left(\gamma\Big|_{[s_1,s_2]}\right)$$

and

$$\mathscr{L}\left(\gamma\Big|_{[s_1,b]}\right) - \mathscr{L}\left(\gamma\Big|_{[s_2,b]}\right) = \mathscr{L}\left(\gamma\Big|_{[s_1,s_2]}\right).$$

So what we need to show is that for any $\varepsilon > 0$ and any $s_1 \in [a,b]$ there exists $\delta > 0$ such that

$$\mathscr{L}\left(\gamma\Big|_{[s_1,s_2]}\right) < \varepsilon \quad \forall s_2 : |s_1 - s_2| < \delta.$$

Fix $\varepsilon > 0$ and $s_1 \in [a,b]$. By continuity of γ we find $\delta_1 > 0$ such that

$$d(\gamma(\tilde{s}), \gamma(\tilde{t})) < \varepsilon \quad \forall |\tilde{s} - s_1|, \ |\tilde{t} - s_1| < \delta_1. \tag{5}$$

Since $\mathscr{L}(\gamma) < \infty$ there exists a partition

$$a = t_0 < t_1 \ldots < t_n = b$$

such that

$$\mathscr{L}(\gamma) - \varepsilon \leq \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})).$$

Set

$$\delta_2 := \inf_{i=1,\ldots,n} |t_i - t_{i-1}|.$$

Set $\delta := \min\{\delta_1, \delta_2\}$ and fix any $s_2 \in [a,b]$ with $|s_1 - s_2| < \frac{\delta}{2}$.

W.l.o.g. $s_1 < s_2$. We then may assume that $t_{i_0-1} < s_1 < t_{i_0} < s_2 < t_{i_0+1}$ for some $i_0 \in \mathbb{N}$ (all other cases follow by an easy adaptation). We now consider the new partition $\tilde{t}_i$,

$$\tilde{t}_i = \begin{cases} t_i & i \leq i_0 - 1 \\ s_1 & i = i_0 \\ t_{i_0} & i = i_0 + 1 \\ s_2 & i = i_0 + 2 \\ t_{i-2} & i \geq i_0 + 3. \end{cases}$$

Then, by triangular inequality,

$$\mathscr{L}(\gamma) - \varepsilon \leq \sum_{i=1}^{n+2} d(\gamma(\tilde{t}_i), \gamma(\tilde{t}_{i-1})). \tag{6}$$

Now let $s_1 = r_0 < r_1 < \ldots = r_m = s_2$ be any partition of $[s_1, s_2]$. Then

$$\sum_{j=0}^{m} d(\gamma(r_j), \gamma(r_{j-1})) = \sum_{i \neq i_0+1, i_0+2} d(\gamma(\tilde{t}_i), \gamma(\tilde{t}_{i-1})) + \sum_{j=0}^{m} d(\gamma(r_j), \gamma(r_{j-1}))$$

$$- \sum_{i=1}^{n+2} d(\gamma(\tilde{t}_i), \gamma(\tilde{t}_{i-1}))$$

$$+ d(\gamma(\tilde{t}_{i_0+1}), \gamma(\tilde{t}_{i_0})) + d(\gamma(\tilde{t}_{i_0+2}), \gamma(\tilde{t}_{i_0+1}))$$

Since we can combine the partitions $\tilde{t}_i$, $i \neq i_0 + 1, i_0 + 2$ with r_j to obtain a partition of $[a,b]$, we have by the definition of length,

$$\sum_{i \neq i_0+1, i_0+2} d(\gamma(\tilde{t}_i), \gamma(\tilde{t}_{i-1})) + \sum_{j=0}^{m} d(\gamma(r_j), \gamma(r_{j-1})) \leq \mathscr{L}(\gamma).$$

By (6) we have

$$- \sum_{i=1}^{n+2} d(\gamma(\tilde{t}_i), \gamma(\tilde{t}_{i-1})) \leq -\mathscr{L}(\gamma) + \varepsilon.$$

By (5) which we can apply since $s_1 < t_{i_0} < s_2$ and thus $|s_1 - s_2|, |t_{i_0} - s_2| < \delta_1$,

$$d(\gamma(\tilde{t}_{i_0+1}), \gamma(\tilde{t}_{i_0})) + d(\gamma(\tilde{t}_{i_0+2}), \gamma(\tilde{t}_{i_0+1})) = d(\gamma(t_{i_0}), \gamma(s_1)) + d(\gamma(s_2), \gamma(t_{i_0})) \leq 2\varepsilon.$$

So we have shown

$$\sum_{j=0}^{m} d(\gamma(r_j), \gamma(r_{j-1})) \leq 3\varepsilon.$$

This holds for any partition (r_j) of $[s_1, s_2]$ and thus

$$\mathscr{L}\left(\gamma\Big|_{[s_1, s_2]}\right) < 3\varepsilon.$$

Since ε was arbitrary, we can conclude. $\square$

For simplicity, we will often restrict our attention to curves defined on $I = [0, 1]$, which we can do without loss of generality. Indeed any curve

$$\gamma : [a, b] \to X$$

can be reparametrized to a curve

$$\tilde{\gamma} : [0, 1] \to X$$

by simply setting

$$\tilde{\gamma}(t) := \gamma(tb + (1-t)a).$$

Similarly any curve $\gamma : [0, 1] \to X$ can be reparametrized to a curve $\tilde{\gamma} : [a, b] \to X$. The length of the curve γ and $\tilde{\gamma}$ above are the same, $\mathscr{L}(\gamma) = \mathscr{L}(\tilde{\gamma})$. Indeed, the length of curves is invariant under reparametrization.

[Reparametrization] Let $\gamma : [a, b] \to X$ be a curve. Let $\tau : [c, d] \to [a, b]$ be a continuous bijection with continuous inverse (i.e. a homeomorphism) such that $\tau(c) = a$ and $\tau(d) = b$. Then, τ is a reparametrization of γ.

Lemma 7. *Let $\gamma : [a, b] \to X$ be a curve and $\tau : [c, d] \to [a, b]$ be a reparametrization. Then if we set $\tilde{\gamma}(t) := \gamma(\tau(t))$ we get that $\tilde{\gamma} : [c, d] \to X$ is a curve and*

$$\mathscr{L}(\gamma) = \mathscr{L}(\tilde{\gamma})$$

We leave the proof as an exercise, but observe that τ maps any partition for $[c, d]$ into a partition of $[a, b]$, and τ^{-1} maps any partition of $[a, b]$ into a partition of $[c, d]$.

Now we want to find geodesics, i.e. shortest curves between two points p and q in X. A curve $\gamma : I \to X$ is called the *shortest curve* or (minimizing) geodesic from p to q if it connects p and q and for any other curve $\tilde{\gamma} : \tilde{I} \to X$ which connects p and q we have we have $\mathscr{L}(\gamma) \leq \mathscr{L}(\tilde{\gamma})$.

In general metric spaces X there is no reason that there exists such a shortest curve γ. As a side-note a shortest curve in general is not unique: think of the many shortest curves connecting the north pole and the south pole of a sphere. In order to conduct in the following chapters our analysis of the Heisenberg group, we conclude this section with a few important notions and facts on maps (possibly) on metric spaces.

The first result from Analysis is the Arzelá-Ascoli theorem – the proof can be found in essentially all Advanced Calculus books. Recall that a set $E \subset X$ is compact, if any sequence $(x_n)_{n \in \mathbb{N}} \subset E$ has a subsequence $(x_{n_i})_{i \in \mathbb{N}}$ and a point $x \in E$ such that $d(x_{n_i}, x) \xrightarrow{i \to \infty} 0$.

Theorem 8 (Arzelá-Ascoli). *Let (X,d) be a metric space and $E \subset X$ be compact. Assume there is a sequence of maps $\gamma_k : [0,1] \to E$ which are* equicontinuous, *i.e. for any $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_{k \in \mathbb{N}} d(\gamma_k(t), \gamma_k(s)) < \varepsilon$ for all $s,t \in [0,1]$ with $|s-t| < \delta$.*

Then, there exists a subsequence $(\gamma_{k_i})_{i \in \mathbb{N}}$ and a continuous limit function $\gamma : [0,1] \to X$ such that γ_{k_i} uniformly converge to γ in the sense that

$$\sup_{t \in [0,1]} d(\gamma_{k_i}(t), \gamma(t)) \xrightarrow{i \to \infty} 0.$$

We will use later that uniform Lipschitz continuity implies equicontinuity. Namely if there exists $\Lambda > 0$ such that

$$\sup_{k \in \mathbb{N}} d(\gamma_k(s), \gamma_k(t)) \leq \Lambda |s-t| \quad \text{for all } s,t \in [0,1]$$

then the equicontinuity condition in Theorem 8 is satisfied.

We now show that any curve with finite length can be parametrized so that it is Lipschitz continuous (so curves with uniformly bounded length are uniformly Lipschitz continuous, and thus equicontinuous).

Proposition 9 (Monotone Reparametrization). *Let $\gamma : [a,b] \to X$ be a curve of finite length, $\mathscr{L}(\gamma) < \infty$.*

Then γ admits a Lipschitz reparameterization in the following sense.

There exists $\tilde{\gamma} : [0,1] \to X$ with the following properties

- $\tilde{\gamma}(0) = \gamma(a)$ *and* $\tilde{\gamma}(1) = \gamma(b)$

- $\tilde{\gamma}([0,1]) = \gamma([a,b])$ *(in the sense of sets in X)*

- $\mathscr{L}(\gamma) = \mathscr{L}(\tilde{\gamma})$,

- $|\tilde{\gamma}(s) - \tilde{\gamma}(t)| \leq \mathscr{L}(\gamma)|s-t| \quad \forall s,t \in [0,1]$.

Proof. Without loss of generality, $[a,b] = [0,1]$. Let $\gamma : [0,1] \to (X,d)$ be a curve of finite length.

Define $\tau(t) := \mathscr{L}(\gamma|_{[0,t]}) : [0,1] \to [0, \mathscr{L}(\gamma)]$, which by Lemma 6 is continuous and monotone increasing.

We would like to set $\hat{\gamma} := \gamma \circ \tau^{-1} : [0, \mathscr{L}(\gamma)] \to X$. The issue is that τ may not be strictly monotone, so τ may not be invertible.

However $\hat{\gamma}$ is still well-defined. Observe that if for some $0 \leq r \leq \tilde{r} \leq 1$ we have $\tau(r) = \tau(\tilde{r})$, then

$$0 = \mathscr{L}(\gamma|_{[0,\tilde{r}]}) - \mathscr{L}(\gamma|_{[0,r]}) = \mathscr{L}(\gamma|_{[r,\tilde{r}]}),$$

that is $\mathscr{L}(\gamma|_{[r,\tilde{r}]}) = 0$ and from the definition of the length $\mathscr{L}$ we conclude that $d(\gamma(s), \gamma(t)) = 0$ for all $s,t \in [r,\tilde{r}]$.

That is $\tau(r) = \tau(\tilde{r})$ implies that γ is constant on $[\tilde{r}, r]$, in particular $\gamma(r) = \gamma(\tilde{r})$.

So we can still define $\hat{\gamma} := \gamma \circ \tau^{-1}$ in the following sense: for a given $t \in [0, \mathscr{L}(\gamma)]$ take any $r \in [0, 1]$ such that $\tau(r) = t$. Such a r exists by the intermediate value theorem since τ is continuous, $\tau(0) = 0$ and $\tau(1) = \mathscr{L}(\gamma)$. Then we set

$$\hat{\gamma}(t) := \gamma(r).$$

If we were to pick any other $\tilde{r}$ with $\tau(\tilde{r}) = t$ then by the above observation we have $\gamma(r) = \gamma(\tilde{r})$ and $\hat{\gamma}(t)$ still has the same value.

We now claim that $\hat{\gamma}$ is continuous. Fix $t_0 \in [0, \mathscr{L}(\gamma)]$ and $\varepsilon > 0$. Take $R \subset [0, 1]$ such that $\tau(r) = t_0$ for all $r \in R$. By the above observation, whenever $r, \tilde{r} \in R$ we have $[r, \tilde{r}] \subset R$. On the other hand if $(r_k)_{k \in \mathbb{N}} \subset [0, 1]$ such that $\tau(r_k) = t_0$ for all $k \in \mathbb{N}$ then if $r = \lim_{k \to \infty} r_k$ we have $\tau(r) = t_0$, by continuity of τ. Combining this with monotonicity of τ we find that for some $r_0 \leq r_1$

$$R = [r_0, r_1], \quad \text{and} \quad \tau(r) < t_0 \quad \text{if } r < r_0, \quad \text{and} \quad \tau(r) > t_0 \quad \text{if } r > r_1.$$

By continuity of γ, there exists an $\delta_1 > 0$ such that $|\gamma(r) - \hat{\gamma}(t_0)| < \varepsilon$ whenever $r \in (r_0 - \delta_1, r_1 + \delta_1)$. Let now $\delta_2 := \min\{\tau(r_0) - \tau(r_0 - \delta_1), \tau(r_1 + \delta_1) - \tau(r_0)) > 0$. Recall that $t_0 = \tau(r_0) = \tau(r_1)$. So whenever t satisfies $|t - t_0| < \delta_2$ then we have $t \in (\tau(r_0 - \delta_1), \tau(r_1 + \delta_1))$, and thus by monotonicity, $t \in \tau(r_0 - \delta_1, r_1 + \delta_1)$ which implies that $|\hat{\gamma}(t) - \hat{\gamma}(t_0)| < \varepsilon$. That is, we have shown continuity of $\hat{\gamma}$.

With the same observation as above, it is now not too difficult to show that $\mathscr{L}(\gamma) = \mathscr{L}(\hat{\gamma})$ – since the only points where τ is not invertible are points where no length is added. Indeed, let $0 = r_0 < r_1 < \ldots < r_n = 1$ be a partition of $[0, 1]$. Set $t_0 = 0$ and $t_n = \mathscr{L}(\gamma)$ and set $t_i := \tau(r_i)$ for $i = 1, \ldots, n - 1$. Then $\hat{\gamma}(t_i) = \gamma(r_i)$. By monotonicity of τ we have $0 = t_1 \leq t_2 \leq \ldots \leq t_n = \mathscr{L}(\gamma)$. It might happen that we have equality $t_i = t_{i-1}$ but then $\tau(r_i) = \tau(r_{i-1})$ which by the argument above means $\hat{\gamma}(t_i) = \hat{\gamma}(t_{i-1})$ and thus $d(\hat{\gamma}(t_i), \hat{\gamma}(t_{i-1})) = 0$. Consequently we have

$$\sum_{i=1}^{n} d(\gamma(r_i), \gamma(r_{i-1})) = \sum_{i=1}^{n} d(\hat{\gamma}(t_i), \hat{\gamma}(t_{i-1})) \leq \mathscr{L}(\hat{\gamma}).$$

Taking the supremum of all partitions of $[0, 1]$ we have

$$\mathscr{L}(\gamma) \leq \mathscr{L}(\hat{\gamma}). \tag{7}$$

For the other direction let $0 = t_0 < t_1 < \ldots < t_n = \mathscr{L}(\gamma)$ be any partition of $[0, \mathscr{L}(\gamma)]$. We now create a new partition $0 = r_0 < \ldots < r_i < \ldots < r_n = 1$ such that $\tau(r_i) = t_i$, and thus by the definition of $\hat{\gamma}$, $\gamma(r_i) = \hat{\gamma}(t_i)$. We set $r_0 := 0$ and $r_n := 1$. We define r_i to be any $r_i \in (0, 1)$ such that $\tau(r_i) = t_i$, this choice of r_i may not be unique but from the intermediate value theorem at least one such r_i must exists. Since $t_{i-1} < t_i$ for all i, from the monotonicity of τ we conclude that $r_{i-1} < r_i$ for all i, and thus $0 = r_0 < r_1 < \ldots < r_n = 1$ is the desired new partition of $[0, 1]$. We then have

$$\sum_{i=1}^{n} d(\hat{\gamma}(t_i), \hat{\gamma}(t_{i-1})) = \sum_{i=1}^{n} d(\gamma(r_i), \gamma(r_{i-1})) \leq \mathscr{L}(\gamma).$$

Taking the supremum over all partitions of $[0, \mathscr{L}(\gamma)]$ we conclude

$$\mathscr{L}(\hat{\gamma}) \leq \mathscr{L}(\gamma). \tag{8}$$

Together, (7) and (8) imply

$$\mathscr{L}(\hat{\gamma}) = \mathscr{L}(\gamma).$$

Next, we observe that the definition of the length of a curve implies

$$d(\gamma(t), \gamma(s)) \overset{(4)}{\leq} \mathscr{L}\left(\gamma\mid_{[s,t]}\right) = \left|\mathscr{L}\left(\gamma\mid_{[0,t]}\right) - \mathscr{L}\left(\gamma\mid_{[0,s]}\right)\right| = |\tau(t) - \tau(s)|.$$

Let $\hat{s}, \hat{t} \in [0, \mathscr{L}(\gamma)]$ and take any $s, t \in [0,1]$ such that $\tau(s) = \hat{s}$, $\tau(t) = \hat{t}$. Then

$$d(\hat{\gamma}(\hat{t}), \hat{\gamma}(\hat{s}))d(\gamma(t), \gamma(s)) \leq |\tau(t) - \tau(s)| = |\hat{t} - \hat{s}|.$$

Thus, $\hat{\gamma}$ is Lipschitz continuous, albeit with the wrong constant, which is easy to fix.

Set

$$\tilde{\gamma}(s) := \hat{\gamma}(\mathscr{L}(\gamma)s), \quad s \in [0,1].$$

Then we have

$$d(\tilde{\gamma}(s), \tilde{\gamma}(t)) \leq \mathscr{L}(\gamma)|s - t| \quad \forall s, t \in [0,1].$$

$$\square$$

The Arzelá-Ascoli theorem, Theorem 8, will play a crucial role in constructing a *candidate* for a shortest curve in the Heisenberg group. Another important ingredient is the following lower semicontinuity of the length.

Proposition 10 (Lower semicontinuity of the length functional). *Let (X,d) be a metric space, and $\{\gamma_n\}_{n \in \mathbb{N}}$ be a sequence of curves into X. If γ_n converges pointwise to a curve, γ, in X, then*

$$\mathscr{L}(\gamma) \leq \liminf_{n \to \infty} \mathscr{L}(\gamma_n)$$

Proof. As discussed above, without loss of generality we can assume that all curves $\gamma_n : [0,1] \to X$.

Let $\varepsilon > 0$ be arbitrary. Since

$$\mathscr{L}(\gamma) = \sup_{p \in P} \sum_{i \geq 1} (d(\gamma(t_i), \gamma(t_{i-1})))$$

where P is the set of partitions of $[0,1]$, we can find a specific partition, $\mu = (t_0, t_1, \ldots, t_m)$, such that

$$\mathscr{L}(\gamma) < \left(\sum_{t_i \in \mu, i \geq 1} d(\gamma(t_i), \gamma(t_{i-1})) \right) + \frac{\varepsilon}{2}.$$

By pointwise convergence $\gamma_n(t) \xrightarrow{n \to \infty} \gamma(t)$ for each fixed t, we can find $N \in \mathbb{N}$ such that

$$d(\gamma_n(t_i), \gamma(t_i)) < \frac{\varepsilon}{4m} \quad \forall i = 0, \ldots, m, \quad \forall n \geq N.$$

Then,

$$d(\gamma(t_i), \gamma(t_{i-1}) \leq d(\gamma(t_i), \gamma_n(t_i)) + d(\gamma_n(t_i), \gamma_n(t_{i-1})) + d(\gamma_n(t_{i-1}), \gamma(t_{i-1}))$$

$$< \frac{\varepsilon}{4m} + d(\gamma_n(t_i), \gamma_n(t_{i-1})) + \frac{\varepsilon}{4m}$$

$$= d(\gamma_n(t_i), \gamma_n(t_{i-1})) + \frac{\varepsilon}{2m}.$$

Thus,

$$\mathscr{L}(\gamma) < \left(\sum_{t_i \in \mu, i \geq 1} d(\gamma_n(t_i), \gamma_n(t_{i-1})) \right) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

Finally, since

$$\mathscr{L}(\gamma_n) = \sup_{\mu \in \mathscr{P}} \sum_{t_i \in \mu, i \geq 1} (d(\gamma_n(t_i), \gamma_n(t_{i-1})))$$

we have

$$\sum_{t_i \in \mu, i \geq 1} d(\gamma_n(t_i), \gamma_n(t_{i-1})) \leq \mathscr{L}(\gamma_n).$$

Thus we have shown,

$$\mathscr{L}(\gamma) < \mathscr{L}(\gamma_n) + \varepsilon, \quad \forall n \geq N.$$

In particular

$$\mathscr{L}(\gamma) < \liminf_{n \to \infty} \mathscr{L}(\gamma_n) + \varepsilon.$$

This holds for any $\varepsilon > 0$, letting $\varepsilon \to 0$ we conclude

$$\mathscr{L}(\gamma) \leq \liminf_{n \to \infty} \mathscr{L}(\gamma_n).$$

$\square$

From Arzelá-Ascoli theorem, Theorem 8, and the observations above we obtain the existence of shortest curves in the following sense.

Theorem 11. *Let (X, d) be any* complete *metric space and $E \subset X$ be a compact set. Let $p \neq q \in E$ such that there exists a continuous curve $\gamma_0 : [0, 1] \to E$ of finite length $\mathscr{L}(\gamma_0) < \infty$ and $\gamma_0(0) = p$ and $\gamma_0(1) = q$. Then there exists a geodesic between p and q, i.e. a curve $\gamma : [0, 1] \to E$ such that $\gamma(0) = p$ and $\gamma(1) = q$ and such that*

$$\mathscr{L}(\gamma) = \inf_{\tilde{\gamma}} \mathscr{L}(\tilde{\gamma})$$

where the infimum is taken over all continuous curves $\tilde{\gamma} : [0, 1] \to E$ with $\tilde{\gamma}(0) = p$ and $\tilde{\gamma}(1) = q$.

It is important to note that above the notion of "shortest curve" is with respect to E not with respect to X, and this might lead to a different notion of what is a shortest curve. Take for example a compact banana-shaped set E in $\mathbb{R}^3$. The straight line from top to bottom of the banana E is likely to not lie within E, so it is not the shortest curve in E!

Proof of Theorem 11. For simplicity we assume $X = E$. Since there exists one curve connecting p and q with finite length we have

$$I := \inf_{\tilde{\gamma}} \mathscr{L}(\tilde{\gamma}) \in [0, \infty).$$

Since there exists one curve connecting p and q there also must be a "minimizing sequence"

$$\gamma_k : [0, 1] \to X \quad \text{of finite length, } \mathscr{L}(\gamma_0) < \infty, \text{ and } \gamma_k(0) = p \text{ and } \gamma_k(1) = q$$

such that

$$\mathscr{L}(\gamma_k) \xrightarrow{k \to \infty} I.$$

We may even assume that

$$I \leq \mathscr{L}(\gamma_k) \leq I + \frac{1}{k} \quad \forall k.$$

By Proposition 9 we may assume without loss of generality (otherwise use $\tilde{\gamma}_k$ instead of γ_k)

$$|\gamma_k(x) - \gamma_k(y)| \leq \left(I + \frac{1}{k}\right)|x - y| \quad \forall x, y \in [0, 1], \quad k \in \mathbb{N}.$$

By Arzelá-Ascoli, Theorem 8, we may assume that we have uniform convergence to some continuous $\gamma : [0, 1] \to X$, otherwise we could pass yet again to a subsequence.

Then, by lower semicontinuity of the length, Proposition 10, we have

$$\mathscr{L}(\gamma) \leq \liminf_{k \to \infty} \mathscr{L}(\gamma_k)$$

This means

$$I \leq \mathscr{L}(\gamma) \leq \liminf_{k \to \infty} \mathscr{L}(\gamma_k) = I.$$

So γ is a shortest curve. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3   Horizontal Curves in the Heisenberg Group

Recall that a differentiable curve $\gamma : [0, 1] \to \mathbb{R}^3$ is called horizontal if (2) holds. In this section we compute important properties of horizontal curves that we will use in the proofs of both our main theorems.

Proposition 12. *If $\gamma \in C^2([0, 1])$ and (2) holds. Then*

$$\lim_{s \to t} \frac{\frac{\gamma^3(t) - \gamma^3(s)}{t-s} + 2\left(\frac{(\gamma^2(t) - \gamma^2(s))}{t-s}\gamma^1(s) - \frac{(\gamma^1(t) - \gamma^1(s))}{t-s}\gamma^2(s)\right)}{t - s} = 0.$$

The convergence rate is uniform in t.

Proof. Since γ is C^2, we have

$$\gamma(s) = \gamma(t) + (s - t)\dot{\gamma}(t) + \frac{1}{2}\ddot{\gamma}(t)(s - t)^2 + o(|t - s|^2).$$

and o is uniform in the domain of γ.

Then,

$$\frac{\frac{\gamma^3(t) - \gamma^3(s)}{t-s} + 2\left(\frac{(\gamma^2(t) - \gamma^2(s))}{t-s}\gamma^1(s) - \frac{(\gamma^1(t) - \gamma^1(s))}{t-s}\gamma^2(s)\right)}{t - s}$$

$$= \frac{\dot{\gamma}^3(t) - \frac{1}{2}\ddot{\gamma}^3(t)(t - s) + 2\left(\left(\dot{\gamma}^2(t) - \frac{1}{2}\ddot{\gamma}^2(t)(t - s)\right)\gamma^1(s) - \left(\dot{\gamma}^1(t) - \frac{1}{2}\ddot{\gamma}^1(t)(t - s)\right)\gamma^2(s)\right)}{t - s}$$

$$+ o(1)$$

$$= \frac{-\frac{1}{2}\ddot{\gamma}^3(t)(t-s) + 2\left(\left(-\frac{1}{2}\ddot{\gamma}^2(t)(t-s)\right)\gamma^1(s) - \left(-\frac{1}{2}\ddot{\gamma}^1(t)(t-s)\right)\gamma^2(s)\right)}{t-s}$$

$$+ \frac{\dot{\gamma}^3(t) + 2\left(\left(\dot{\gamma}^2(t)\right)\gamma^1(s) - \left(\dot{\gamma}^1(t)\right)\gamma^2(s)\right)}{t-s}$$

$$+ o(1)$$

$$= -\frac{1}{2}\ddot{\gamma}^3(t) + 2\left(-\frac{1}{2}\ddot{\gamma}^2(t)\gamma^1(t) + \frac{1}{2}\ddot{\gamma}^1(t)\gamma^2(t)\right) + o(1)$$

$$+ \frac{\dot{\gamma}^3(t) + 2\left(\left(\dot{\gamma}^2(t)\right)\gamma^1(s) - \left(\dot{\gamma}^1(t)\right)\gamma^2(s)\right)}{t-s}$$

$$+ o(1)$$

We define

$$f(s) := \dot{\gamma}^3(t) + 2\left(\left(\dot{\gamma}^2(t)\right)\gamma^1(s) - \left(\dot{\gamma}^1(t)\right)\gamma^2(s)\right)$$

and, we observe that by horizontality, $f(t) = 0$. Thus

$$\frac{f(s)}{t-s} = -\frac{f(s) - f(t)}{s-t} = -f'(t) + o(1)$$

Then, we have

$$f'(s) = 2\left(\left(\dot{\gamma}^2(t)\right)\dot{\gamma}^1(s) - \left(\dot{\gamma}^1(t)\right)\dot{\gamma}^2(s)\right)$$

So

$$f'(t) = 2\left(\left(\dot{\gamma}^2(t)\right)\dot{\gamma}^1(t) - \left(\dot{\gamma}^1(t)\right)\dot{\gamma}^2(t)\right)$$

Consequently,

$$\frac{\frac{\gamma^3(t)-\gamma^3(s)}{t-s} + 2\left(\frac{\left(\gamma^2(t)-\gamma^2(s)\right)}{t-s}\gamma^1(s) - \frac{\left(\gamma^1(t)-\gamma^1(s)\right)}{t-s}\gamma^2(s)\right)}{t-s}$$

$$= -\frac{1}{2}\ddot{\gamma}^3(t) + 2\left(-\frac{1}{2}\ddot{\gamma}^2(t)\gamma^1(t) + \frac{1}{2}\ddot{\gamma}^1(t)\gamma^2(t)\right) + o(1)$$

$$- \left(2\left(\left(\dot{\gamma}^2(t)\right)\dot{\gamma}^1(t) - \left(\dot{\gamma}^1(t)\right)\dot{\gamma}^2(t)\right)\right) + o(1)$$

$$+ o(1)$$

$$= -\frac{1}{2}\frac{d}{dt}\dot{\gamma}^3(t) + 2\left(\dot{\gamma}^2(t)\gamma^1(t) - \dot{\gamma}^1(t)\gamma^2(t)\right)$$

$$- \frac{1}{2} - 2\left(\dot{\gamma}^2(t)\dot{\gamma}^1(t) - \dot{\gamma}^1(t)\dot{\gamma}^2(t)\right)$$

$$- \left(+2\left(\left(\dot{\gamma}^2(t)\right)\dot{\gamma}^1(t) - \left(\dot{\gamma}^1(t)\right)\dot{\gamma}^2(t)\right)\right)$$

$$+ o(1)$$

$$= 0$$

$$- \left(+1\left(\left(\dot{\gamma}^2(t)\right)\dot{\gamma}^1(t) - \left(\dot{\gamma}^1(t)\right)\dot{\gamma}^2(t)\right)\right)$$

$$+ o(1)$$

$$= o(1).$$

Then

$$\lim_{s \to t} \frac{\frac{\gamma^3(t)-\gamma^3(s)}{t-s} + 2\left(\frac{(\gamma^2(t)-\gamma^2(s))}{t-s}\gamma^1(s) - \frac{(\gamma^1(t)-\gamma^1(s))}{t-s}\gamma^2(s)\right)}{t-s} = 0.$$

as desired.

Then,

$$\lim_{s \to t} \frac{\left[|\gamma^1(t)-\gamma^1(s)|^2 + |\gamma^2(t)-\gamma^2(s)|^2 + \left|\gamma^3(t)-\gamma^3(s) + 2\left((\gamma^2(t)-\gamma^2(s))\gamma^1(s) - (\gamma^1(t)-\gamma^1(s))\gamma^2(s)\right)\right|\right]^2}{|t-s|^4}$$

$$= \lim_{s \to t} \frac{|\gamma^1(t)-\gamma^1(s)|^2 + |\gamma^2(t)-\gamma^2(s)|^{2^2}}{|t-s|^4}$$

$$= \lim_{s \to t} \frac{|\gamma^1(t)-\gamma^1(s)|^4 + 2|\gamma^1(t)-\gamma^1(s)|^2|\gamma^2(t)-\gamma^2(s)|^2 + |\gamma^2(t)-\gamma^2(s)|^4}{|t-s|^4}$$

$$= \lim_{s \to t} \left(\frac{|\gamma^1(t)-\gamma^1(s)|}{|t-s|}\right)^4 + 2\left(\frac{|\gamma^1(t)-\gamma^1(s)|}{|t-s|}\right)^2\left(\frac{|\gamma^2(t)-\gamma^2(s)|}{|t-s|}\right)^2 + \left(\frac{|\gamma^2(t)-\gamma^2(s)|}{|t-s|}\right)^4$$

$$= \dot{\gamma}^1(t)^4 + 2\dot{\gamma}^1(t)^2\dot{\gamma}^2(t)^2 + \dot{\gamma}^4(t) = \left(\dot{\gamma}^1(t)^2 + \dot{\gamma}^2(t)^2\right)^2$$

and the convergence is uniformly in t by the above considerations. $\square$

From Proposition 12 we readily obtain

Corollary 13. *If $\gamma \in C^2([0,1],\mathbb{R}^3)$ and (2) holds*

$$\frac{d_K(\gamma(t),\gamma(s))}{|t-s|} \xrightarrow{s \to t} \sqrt{\dot{\gamma}^1(t)^2 + \dot{\gamma}^2(t)^2}$$

The convergence is uniform in t. In particular we have

$$\mathscr{L}_K(\gamma) < \infty.$$

4 Existence of Shortest Curves in the Heisenberg Group

In this section we want to show Theorem 1.

Of course we would like to apply Theorem 11, however we need to be careful with the compactness assumption in that theorem, since $\mathbb{H}_1$ is not compact. However, one could justifiably believe that any curve $\gamma : [0,1] \to \mathbb{H}_1$ which goes too far away from p and q is not a good candidate for shortest curve. We need to quantify this and for this we compare the Korányi metric locally with the Euclidean metric.

Lemma 14. *Let $K \subset \mathbb{R}^3$ be compact (in the sense of the Euclidean metric). Then $K \subset \mathbb{H}_1$ is compact (in the sense of the Korányi metric).*

Proof. Since K is compact as Euclidean set $\mathbb{R}^3$ it is bounded and thus there must be some $\Lambda > 0$ such that

$$\max\{|p_1|,|p_2|,|p_2|\} < \Lambda \quad \forall p = (p_1,p_2,p_3) \in K.$$

Using repeatedly Young's inequality $2ab \le a^2 + b^2$ we find that for $p,q \in K$

$$d_K(q,p) = (|p_1 - q_1|^2 + |p_2 - q_2|^{2^2} + |p_3 - q_3 + 2(p_2 q_1 - p_1 q_2)|^2)^{\frac{1}{4}}$$

$$\le (|p_1 - q_1|^2 + |p_2 - q_2|^{2^2} + 2|p_3 - q_3|^2 + 2|2(p_2 q_1 - p_1 q_2)|^2)^{\frac{1}{4}}$$

$$= (|p_1 - q_1|^2 + |p_2 - q_2|^{2^2} + 2|p_3 - q_3|^2 + 2|2(p_2 - q_2)q_1 + (q_1 - p_1)q_2)|^2)^{\frac{1}{4}}$$

$$\le (|p_1 - q_1|^2 + |p_2 - q_2|^{2^2} + 2|p_3 - q_3|^2 + 8(|p_2 - q_2|\Lambda + |q_1 - p_1|\Lambda))^2)^{\frac{1}{4}}$$

We conclude that for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $p,q \in K$ and $|p - q| < \delta$ (in the Euclidean sense) then $d_K(p,q) < \varepsilon$.

In particular any (Euclidean) converging sequence in K also converges in the sense of the Korányi metric d_K. Thus K is also compact in the Korányi sense.   □

The following lemma shows that "far away" in the Euclidean sense implies "far away" in the Korányi sense.

Lemma 15. *Fix $q \in \mathbb{R}^3$. For any $\Lambda > 0$ there exists $\Theta > 0$ such that the following is true: if for some $p \in \mathbb{R}^3$ we have*

$$|p - q| > \Theta$$

then

$$d_K(p,q) > \Lambda.$$

Proof. Observe that for any $p,q \in \mathbb{R}^3$

$$d_K(p,q)^2 \ge |p_3 - q_3 + 2(p_2 q_1 - p_1 q_2)| = |p_3 - q_3 + 2((p_2 - q_2)q_1 + q_1 q_2 - (p_1 - q_1)q_2$$
$$- q_1 q_2)|$$
$$= |p_3 - q_3 + 2((p_2 - q_2)q_1 - (p_1 - q_1)q_2|$$
$$\ge (|p_3 - q_3| - 2|q_1(p_2 - q_2) - q_2(p_1 - q_1)|)$$
$$\ge (|p_3 - q_3| - 2(|q_1||p_2 - q_2| + |q_2||p_1 - q_1|))$$
$$\ge (|p_3 - q_3| - 2(|q_1||p_2 - q_2| + |q_2||p_1 - q_1|))$$

Now fix $q = (q_1, q_2, q_3) \in \mathbb{R}^3$ and $\Lambda > 0$ and set

$$\Gamma := |q_1| + |q_2|.$$

Take $\Theta > 0$ so that the following conditions are satisfied: $\Theta > \sqrt{3}\Lambda$ and $\frac{1}{\sqrt{3}}\Theta - 2\Gamma\Lambda > \Lambda^2$.

Now take $p = (p_1, p_2, p_3) \in \mathbb{R}^3$ such that

$$|p - q| > \Theta.$$

Then

$$\max\{|p_1 - q_1|, |p_2 - q_2|, |p_3 - q_3|\} > \frac{1}{\sqrt{3}}\Theta.$$

Then either

$$\max\{|p_1 - q_1|, |p_2 - q_2|\} > \Lambda$$

or

$$|p_3 - q_3| > \frac{1}{\sqrt{3}}\Theta.$$

From the above estimates we have

$$d_K(p,q) \geq \max\left\{|p_1-q_1|, |p_2-q_2|, (|p_3-q_3| - 2(|q_1||p_2-q_2| + |q_2||p_1-q_1|))^{\frac{1}{2}}\right\}$$

In the case that $\max\{|p_1-q_1|, |p_2-q_2|\} > \Lambda$ we conclude that

$$d_K(p,q) > \Lambda,$$

and we are done. If on the other hand both $|p_1-q_1|$ or $|p_2-q_2| < \Lambda$ then we have $|p_3-q_3| > \frac{1}{\sqrt{3}}\Theta$ and thus

$$d_K(p,q)^2 \geq |p_3-q_3| - 2(|q_1||p_2-q_2| + |q_2||p_1-q_1|)$$

$$\geq \frac{1}{\sqrt{3}}\Theta - 2\Gamma\Lambda$$

Again in this case, by the choice of Θ we find that

$$d_K(p,q)^2 > \Lambda^2,$$

and we can conclude $d_K(p,q) > \Lambda$ as desired. $\square$

Proof of Theorem 1. Fix $p,q \in \mathbb{R}^3$. There exists a smooth horizontal curve $\tilde{\gamma}$ connecting p and q, take for example the $\mathscr{L}_{cc}$-geodesic from [3], and in view of Corollary 13 $\tilde{\gamma}$ has finite length: $\mathscr{L}_K(\tilde{\gamma}) < \infty$.

Let $R > 0$ such that for any $r \in \mathbb{R}^3$ with $|p-r| > R$ we have in view of Lemma 15

$$d_K(p,r) > \mathscr{L}_K(\tilde{\gamma}).$$

This implies that any continuous curve $\gamma : [0,1] \to \mathbb{R}^3$ with $\gamma(0) = p$ and $\gamma(1) = q$ and $|\gamma(t) - p| > R$ for any $t \in (0,1)$ we have

$$\mathscr{L}_K(\gamma) > \mathscr{L}_K(\tilde{\gamma}).$$

Set $E := \{r \in \mathbb{R}^3 : |r-p| \leq R\}$ which is a compact set in the Euclidean sense, and thus in view of Lemma 14 also in the Korányi sense. Then we have shown that

$$\inf_{\gamma:[0,1]\to E} \mathscr{L}_K(\gamma) = \inf_{\gamma:[0,1]\to\mathbb{R}^3} \mathscr{L}_K(\gamma),$$

where both infima are taken over continuous curves γ with $\gamma(0) = p$ and $\gamma(1) = q$. Now we can finally apply Theorem 11. Thus, there is a shortest curve between p and q. $\square$

5   Length of Curves in the Heisenberg Group – Proof of Theorem 2

In this section we show that

$$\mathscr{L}_{cc}(\gamma) = \mathscr{L}_K(\gamma),$$

whenever $\gamma \in C^2$ is a horizontal curve, i.e. whenever γ satisfies (2).

Proof of Theorem 2. From (2) in particular,

$$\frac{d}{dt}\left(\dot{\gamma}^3(t) + 2\left(\dot{\gamma}^2(t)\gamma^1(t) - \dot{\gamma}^1(t)\gamma^2(t)\right)\right) = 0$$

We apply Proposition 12 and obtain

$$
\lim_{s\to t}\left(\frac{d_K(\gamma(t),\gamma(s))}{|t-s|}\right)^4 = \lim_{s\to t}\frac{\left(|\gamma^1(t)-\gamma^1(s)|^2+|\gamma^2(t)-\gamma^2(s)|^2\right)^2}{|t-s|^4}
$$

$$
=\lim_{s\to t}\frac{|\gamma^1(t)-\gamma^1(s)|^4+2|\gamma^1(t)-\gamma^1(s)|^2|\gamma^2(t)-\gamma^2(s)|^2+|\gamma^2(t)-\gamma^2(s)|^4}{|t-s|^4}
$$

$$
=\lim_{s\to t}\left(\frac{|\gamma^1(t)-\gamma^1(s)|}{|t-s|}\right)^4+2\left(\frac{|\gamma^1(t)-\gamma^1(s)|}{|t-s|}\right)^2\left(\frac{|\gamma^2(t)-\gamma^2(s)|}{|t-s|}\right)^2+\left(\frac{|\gamma^2(t)-\gamma^2(s)|}{|t-s|}\right)^4
$$

$$
=\dot\gamma^1(t)^4+2\dot\gamma^1(t)^2\dot\gamma^2(t)^2+\dot\gamma^4(t)=\left(\dot\gamma^1(t)^2+\dot\gamma^2(t)^2\right)^2
$$

Then, taking the fourth root,

$$
\lim_{s\to t}\frac{d_K(\gamma(t),\gamma(s))}{|t-s|}=\sqrt{\dot\gamma^1(t)^2+\dot\gamma^2(t)^2}
$$

<div style="text-align:right">□</div>

Proof of $\mathscr{L}_{cc}(\gamma)=\mathscr{L}_K(\gamma)$ if γ is horizontal. From Corollary 19, we see that

$$
\lim_{s\to t}\frac{d_K(\gamma(t),\gamma(s))}{|t-s|}=\sqrt{\dot\gamma^1(t)^2+\dot\gamma^2(t)^2}
$$

uniformly in t. Then, from the above limit, given some $\varepsilon>0$ choose $\delta>0$, such that when $|t_i-t_{i-1}|<\delta$, we have

$$
\left|\frac{d(\gamma(t_i),\gamma(t_{i-1}))}{|t_i-t_{i-1}|}-\sqrt{|\dot\gamma^1(t_i)|^2+|\dot\gamma^2(t_i)|^2}\right|<\varepsilon.
$$

Multiplying by $|t_i-t_{i-1}|$,

$$
\left|\frac{d(\gamma(t_i),\gamma(t_{i-1}))}{|t_i-t_{i-1}|}|t_i-t_{i-1}|-\sqrt{|\dot\gamma^1(t_i)|^2+|\dot\gamma^2(t_i)|^2}|t_i-t_{i-1}|\right|<\varepsilon|t_i-t_{i-1}|.
$$

Now, let $\mathscr{P}$ be the set of partitions of $[0,1]$ such that for any $\mu\in\mathscr{P}$, we have $|t_i-t_{i-1}|<\delta$ for each t_i in μ. Then, for a given $\mu\in\mathscr{P}$,

$$
\sum_{t_i\in\mu,i\geq1}\left|\frac{d(\gamma(t_i),\gamma(t_{i-1}))}{|t_i-t_{i-1}|}|t_i-t_{i-1}|-\sqrt{|\dot\gamma^1(t_i)|^2+|\dot\gamma^2(t_i)|^2}|t_i-t_{i-1}|\right|<\varepsilon\underbrace{\sum_{t_i\in\mu,i\geq1}|t_i-t_{i-1}|}_{=1}\ =\varepsilon
$$

So, we get

$$
\left|\sum_{t_i\in\mu,i\geq1}\frac{d(\gamma(t_i),\gamma(t_{i-1}))}{|t_i-t_{i-1}|}|t_i-t_{i-1}|-\sum_{t_i\in\mu,i\geq1}\sqrt{|\dot\gamma^1(t_i)|^2|\dot\gamma^2(t_i)|^2}|t_i-t_{i-1}|\right|
$$

$$
=\left|\sum_{t_i\in\mu,i\geq1}d(\gamma(t_i),\gamma(t_{i-1}))-\sum_{t_i\in\mu,i\geq1}\sqrt{|\dot\gamma^1(t_i)|^2+|\dot\gamma^2(t_i)|^2}|t_i-t_{i-1}|\right|<\varepsilon
$$

Then,

$$\varepsilon > \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} d(\gamma(t_i), \gamma(t_{i-1})) - \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right|$$

$$\geq \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} d(\gamma(t_i), \gamma(t_{i-1})) \right| - \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right|$$

$$= \mathscr{L}_K(\gamma) - \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right|$$

Similarly,

$$\varepsilon > \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} d(\gamma(t_i), \gamma(t_{i-1})) - \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right|$$

$$\geq \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right| - \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} d(\gamma(t_i), \gamma(t_{i-1})) \right|$$

$$= \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right| - \mathscr{L}_K(\gamma)$$

That is,

$$\varepsilon > \left| \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right| - \mathscr{L}_K(\gamma) \right|$$

So, we have

$$\left| \mathscr{L}_K(\gamma) - \sup_{\mu \in \mathscr{P}} \left| \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right| \right|$$

$$\leq \left| \mathscr{L}_K(\gamma) - \sup_{\mu \in \mathscr{P}} \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| \right| < \varepsilon$$

Note that, since $\dot{\gamma}$ is continuous, the function

$$\sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2}$$

is continuous and hence integrable. So,

$$\sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}|$$

is a Riemann Sum, and

$$\sup_{\mu \in \mathscr{P}} \sum_{t_i \in \mu, i \geq 1} \sqrt{|\dot{\gamma}^1(t_i)|^2 + |\dot{\gamma}^2(t_i)|^2} |t_i - t_{i-1}| = \int_0^1 \sqrt{|\dot{\gamma}^1(t)|^2 + |\dot{\gamma}^2(t)|^2} dt$$

Then,

$$\left| \mathscr{L}_K(\gamma) - \int_0^1 \sqrt{|\dot{\gamma}^1(t)|^2 + |\dot{\gamma}^2(t)|^2} dt \right| < \varepsilon$$

This holds for any $\varepsilon > 0$, so letting $\varepsilon \to 0$ we conclude

$$\mathscr{L}_K(\gamma) = \int_0^1 \sqrt{|\dot{\gamma}^1(t)|^2 + |\dot{\gamma}^2(t)|^2} dt.$$

This proves $\mathscr{L}_{cc}(\gamma) = \mathscr{L}_K(\gamma)$ which in particular implies Theorem 2. $\square$

Acknowledgement

Bibliography

[1] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, volume 259 of Progress in Mathematics. Birkh¨auser Verlag, Basel, 2007.

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

[3] P. Hajłasz and S. Zimmerman. Geodesics in the Heisenberg group. *Anal. Geom. Metr. Spaces,* 3(1):325–337, 2015.

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