Progress in Mathematics

Luca Capogna
Donatella Danielli
Scott D. Pauls
Jeremy T. Tyson

An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem

Birkhäuser

Progress in Mathematics

Volume 259

Luca Capogna
Donatella Danielli
Scott D. Pauls
Jeremy T. Tyson

An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem

Birkhäuser
Basel · Boston · Berlin

Authors:

Luca Capogna
Department of Mathematics
University of Arkansas
Fayetteville, AR 72701
USA
e-mail: lcapogna@uark.edu

Donatella Danielli
Department of Mathematics
Purdue University
West Lafayette, IN 47907-1395
USA
e-mail: danielli@math.purdue.edu

Scott D. Pauls
Department of Mathematics
6188 Kemeny Hall
Dartmouth College
Hanover, NH 03755
USA
e-mail: scott.pauls@dartmouth.edu

Jeremy T. Tyson
Department of Mathematics
University of Illinois at Urbana-Champaign
1409 West Green Street
Urbana, IL 61801
USA
e-mail: tyson@math.uiuc.edu

2000 Mathematics Subject Classification: primary 53C17; secondary 22E30, 30C65, 32T27, 32V15, 43A80, 46E35, 49Q05, 49Q20, 53A35, 53C42, 53C44, 53D10, 70Q05, 92C55, 93C85

Library of Congress Control Number : 2007922258

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 978-3-7643-8132-5 Birkhäuser Verlag AG, Basel - Boston - Berlin

Basel · Boston · Berlin
P.O. Box 133, CH-4010 Basel, Switzerland
Part of Springer Science+Business Media
Printed on acid-free paper produced of chlorine-free pulp. TCF ∞
Printed in Germany
ISBN-10: 3-7643-8132-9 e-ISBN-10: 3-7643-8133-7
ISBN-13: 978-3-7643-8132-5 e-ISBN-13: 978-3-7643-8133-2

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

Dedicated to Nicola Garofalo
on the occasion of his 50th birthday

. . . mercatique solum,
facti de nomine Byrsam,
taurino quantum possent
circumdare tergo.

(*Virgil, Eneid,*
Book I, 367–368)

Contents

List of Figures

Preface

Sub-Riemannian (also known as Carnot–Carathéodory) spaces are spaces whose metric structure may be viewed as a constrained geometry, where motion is only possible along a given set of directions, changing from point to point. They play a central role in the general program of analysis on metric spaces, while simultaneously continuing to figure prominently in applications from other scientific disciplines ranging from robotic control and planning problems to MRI function to new models of neurobiological visual processing and digital image reconstruction. The quintessential example of such a space is the so-called (first) Heisenberg group. For a precise description we refer the reader to Chapter 2; here we merely remark that this is the simplest instance of a sub-Riemannian space which retains many features of the general case.

The Euclidean isoperimetric problem is the premier exemplar of a problem in the geometric theory of the calculus of variations. In Chapter 1 we review the origins of this celebrated problem and present a spectrum of well-known approaches to its solution. This discussion serves as motivation and foundation for the remainder of this survey, which is devoted to the isoperimetric problem in the Heisenberg group. First formulated by Pierre Pansu in 1982 (see (8.2) in Chapter 8 for the precise statement), the isoperimetric problem in the first Heisenberg group is one of the central questions of sub-Riemannian geometric analysis which has resisted sustained efforts by numerous research groups over the past twenty-five years.

Our goals, in writing this survey, are twofold. First, we want to describe the isoperimetric problem in the Heisenberg group, outline recent progress in this field, and introduce a number of techniques which we think may lead to further understanding of the problem. In accomplishing this program we simultaneously provide a concise and detailed introduction to the basics of analysis and geometry in the setting of the Heisenberg group. Rather than present a general, exhaustive introduction to the field of subelliptic equations, Carnot–Carathéodory metrics and sub-Riemannian geometry, as is done (to different extents) in the standard references [32], [100], [103], [130], [243], [203], [255], and in the forthcoming monograph [114], here we focus on the simplest example of the first Heisenberg group. This seems to us a good starting point for a novice who wants to learn some basic techniques and issues in the field without having to face the most general picture first. At present there are no elementary or introductory texts in this area; we are convinced that there is great need for such a text, to motivate young researchers to work in this area or to clarify to mathematicians working in other fields its principal features. While most of the material in this survey has appeared elsewhere, the approach to the horizontal differential geometry of submanifolds via Riemannian approximation is original; we hope it may be helpful for those who wish to further investigate this interesting line of research.

The structure of this survey is as follows:

Chapter 1. We give an abbreviated review of the isoperimetric problem and its solution in Euclidean space, indicating a few proofs for the sharp isoperimetric inequality in the plane arising from diverse areas such as complex analysis, differential geometry, geometric measure theory, nonlinear evolution PDE's (curvature flow), and integral geometry.

Chapters 2, 3. We introduce the first Heisenberg group $\mathbb{H}$ and describe in detail its principal metric, analytic and differential geometric features. Our presentation of the sub-Riemannian structure of $\mathbb{H}$ is somewhat nonstandard, as we first introduce an explicit coordinate system and later define the sub-Riemannian metric by referencing this particular set of coordinates. This "hands-on" approach, while not in the coordinate-free approach of modern geometry, fits well with our basic aim as described above.

In Chapter 3 we present a selection of pure and applied mathematical models which feature aspects of Heisenberg geometry: CR geometry, Gromov hyperbolic spaces, jet spaces, path planning for nonholonomic motion, and the functional structure of the mammalian visual cortex.

Chapter 4. We turn from the global metric structure of the Heisenberg group $\mathbb{H}$ to a study of the geometry of submanifolds. We introduce the concept of horizontal mean curvature, which gives a sub-Riemannian analog for the classical notion of mean curvature. Computations of the sub-Riemannian differential geometric machinery are facilitated by considering $\mathbb{H}$ as a Gromov–Hausdorff limit of Riemannian manifolds. We illustrate this by computing some of the standard machinery of differential geometry in the Riemannian approximants, and identifying the appropriate sub-Riemannian limits. Typical submanifolds in $\mathbb{H}$ contain an exceptional set, the so-called *characteristic set,* where this sub-Riemannian differential geometric machinery breaks down. In Section 4.4 we work through an extended analysis of the limiting behavior of fundamental ingredients of sub-Riemannian submanifold geometry at the characteristic locus. Such an analysis plays a key role in our later discussion of Pansu's isoperimetric conjecture (see Chapter 8).

Chapters 5, 6. Weakening the smoothness requirements of differential geometry leads to the study of geometric measure theory. We give a broad summary of some basic tools of geometric measure theory in $\mathbb{H}$: horizontal Sobolev and BV spaces and the Sobolev embedding theorems, perimeter measure, Hausdorff and Minkowski content and measure, area and co-area formulas, and the Pansu–Rademacher differentiability theorem for Lipschitz functions. This development culminates in Section 6.4, where we present two derivations of the first variation formula for perturbations of the horizontal perimeter. These formulas are essential ingredients in the most recent developments associated with proofs of Pansu's conjecture in certain special cases; our presentation of the first variation formula for the horizontal perimeter is preparatory to our discussion of these developments in Sections 8.5 and 8.6. We conclude Chapter 6 with a brief overview of Mostow's rigidity theorem for cocompact lattices in complex hyperbolic space, emphasizing the role of quasiconformal functions on the Heisenberg group in the proof and building on this to summarize some of the essential aspects of the field of sub-Riemannian geometric function theory which has grown from this application.

Chapters 7, 8. With the above tools in hand, we are prepared to begin our discussion of the sub-Riemannian isoperimetric problem in the Heisenberg group. In Chapter 7 we give two proofs for the isoperimetric inequality in $\mathbb{H}$. Neither proof gives the best constant or identifies the extremal configuration. The first proof relies on the equivalence of the isoperimetric inequality with the geometric Sobolev inequality. The second is Pansu's original proof, which relies on an adaptation of an argument of Croke. Chapter 8 is the heart of the survey. We present Pansu's famous conjecture on the isoperimetry extremals, and discuss the current state of knowledge, including various partial results (requiring *a priori* regularity and/or symmetry), and various Euclidean techniques whose natural analogs have been shown to fail in $\mathbb{H}$.

Chapter 9. In this concluding chapter, we discuss three other analytic "best constant" problems in the Heisenberg group, whose solutions are known.

We envision this survey as being of use to a variety of audiences and in a variety of ways. Readers who are interested only in obtaining an overview of the general subject area are invited to read Chapters 2–6. These chapters provide a concise introduction to the basic analytic and geometric machinery relevant for the sub-Riemannian metric structure of $\mathbb{H}$. We presuppose a background in Riemannian geometry, PDE and Sobolev spaces (in the Euclidean context), and the basic theory of Lie groups. For those already fluent in sub-Riemannian geometric analysis, Chapters 7 and 8 provide an essentially complete description of the current state of knowledge regarding Pansu's conjecture, and present a wide array of potential avenues for attacks on it and related conjectures. Chapter 9 is essentially independent of the preceding two chapters and can be read immediately following Chapter 6.

We have deliberately aimed at a treatment which is neither comprehensive nor put forth in the most general setting possible, but instead have chosen to work (almost entirely) in the first Heisenberg group, and present those topics and results most closely connected with the isoperimetric problem.

Notable topics which we omit or mention only briefly include:

- The theory of (sub-)Laplacians and the connections between sub-Riemannian geometry, subelliptic PDE and Hörmander's "sums of squares" operators. Similarly, we have very little to say on the subject of potential theory (both linear and nonlinear), apart from some brief results in Chapter 6 connected with the Sobolev embedding theorems.
- Carnot groups as tangent cones of general sub-Riemannian manifolds.
- Further extensions of geometric analysis beyond the sub-Riemannian context, e.g., the emerging theory of "analysis on metric measure spaces".
- *Singular geodesics* in the Martinet (and other sub-Riemannian) distributions.
- Further applications of sub-Riemannian geometry in control theory and nonholonomic mechanics (apart from the discussion in Chapter 3).

These topics are all covered in prior textbooks, which mitigates their omission here. Singular geodesics in sub-Riemannian geometry play a starring role in Montgomery's text [203], and the intricacies of the construction of tangent cones on sub-Riemannian spaces are presented in both [203] and the survey article of Bellaïche [32]. For analysis on metric spaces, the best reference is Heinonen [136]; see also [137]. For nonlinear potential theory (in the Euclidean setting) the principal reference is Heinonen–Kilpeläinen–Martio [139]. In addition to the preceding list, we are also omitting a full discussion of several important recent developments, most notably:

- Rigidity theorems à la Bernstein for minimal surfaces in the Heisenberg group.
- The extraordinary developments in rectifiability and geometric measure theory connected with the extension by Franchi, Serapioni and Serra-Cassano of the structure theorem of de Giorgi to sets of finite perimeter in Carnot groups.

These topics are still very much the subject of active investigation and it is too soon to write their definitive story.

In conclusion, we would be remiss in failing to pay homage to the comprehensive treatise by Gromov [130] on the metric geometry of sub-Riemannian spaces, which provides a wealth of information regarding the structure of these remarkable spaces. Much of the current development in the area represents the working out and elaboration of ideas and notions presented in that work.

Remarks on notation and conventions

With only a few exceptions, we have attempted to keep our discussion of references, citations, etc. limited to the "Further results" and "Notes" sections of each chapter. In certain cases, particularly when we have used without proof some well-known result which can be found in another textbook, we have deviated from this policy. Despite its size, our reference list still represents only a fraction of the work in this area, and should be viewed merely as a guide to the existing literature.

Our notation and terminology is for the most part standard. The Euclidean space of dimension n and its unit sphere are denoted by $\mathbb{R}^n$ and $\mathbb{S}^{n-1}$, respectively. By $H^n_{\mathbb{A}}$ we denote the hyperbolic space over the division algebra $\mathbb{A}$ (either the real field $\mathbb{R}$, the complex field $\mathbb{C}$, the quaternionic division algebra $\mathbb{K}$ or Cayley's octonions $\mathbb{O}$.) We denote by $B(x,r)$ the (open) metric ball with center x and radius r in any metric space (X,d). We write $\operatorname{diam} A$ for the diameter of any bounded set $A \subset X$, and $\operatorname{dist}(A,B)$ for the distance between any two nonempty sets $A, B \subset X$. If the metric needs to be emphasized we may use a notation of the form $B_d(x,r)$, $\operatorname{diam}_d A$, etc. In the case of the Euclidean metric in $\mathbb{R}^n$, we write $B_E(x,r)$, $\operatorname{diam}_E d$, etc. We always reserve the notation $\langle \cdot, \cdot \rangle$ for the standard Euclidean inner product. An alternate family of inner products, associated to a family of degenerating Riemannian metrics g_L on $\mathbb{R}^3$, will be written $\langle \cdot, \cdot \rangle_L$. The latter family of inner products will play an essential role throughout the survey.

We will use both vector notation and complex notation for points in $\mathbb{R}^2$, switching between the two without further discussion. The unit imaginary element in $\mathbb{C}$ will always be written $\mathbf{i}$. For $v = (v_1, v_2) \in \mathbb{R}^2$ we write $v^\perp = (v_2, -v_1)$.

In any dimension n, we write $|A|$ for the Lebesgue measure of a measurable set A. For any domain $\Omega \subset \mathbb{R}^n$, we denote by $W^{k,p}(\Omega)$ the Sobolev space of functions on Ω admitting p-integrable distributional derivatives of order at most k. The surface area measure on a smooth hypersurface S in a Euclidean space $\mathbb{R}^n$ of any dimension will be denoted $d\sigma$. Finally, we write

$$\omega_{n-1} := \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

for the surface area $\sigma(\mathbb{S}^{n-1})$ of the standard unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$.

Acknowledgments

We want to thank Mario Bonk, Giovanna Citti, Bruce Kleiner and Pierre Pansu for encouragement, useful comments and suggestions. This project originated on the occasion of the 2004 New Mexico Analysis Seminar. We are very grateful to the organizers Tiziana Giorgi, Joe Lakey, Cristina Pereyra and Robert Smits for their kind support. Financial support for each of the authors during the preparation of this monograph has been provided by the National Science Foundation. We are deeply grateful to the various referees, whose detailed and penetrating comments led to significant improvements in the presentation and structure of the monograph. Finally, we would like to acknowledge the editorial staff at Birkhäuser for their significant technical assistance in shepherding the manuscript to its final published form.

The work of Nicola Garofalo has influenced and impacted each of our contributions to sub-Riemannian differential geometry and geometric measure theory. We are deeply grateful for his long-standing encouragement and support in his various roles as mentor, collaborator and friend. We dedicate this work to him with sincere admiration and thanks.

This survey is based on the work of a wide array of authors and has benefited from feedback and comments from many colleagues. In particular we are grateful to Thomas Bieske, Gian Paolo Leonardi, Roberto Monti, Manuel Ritoré, César Rosales and Paul Yang for generously devoting time to reading drafts of the manuscript and sharing with us their remarks and corrections. Any remaining errors or omissions are ours alone. We apologize for them in advance and welcome further corrections, which will be incorporated in any later revisions.

L.C. (Fayetteville, Arkansas)
D.D. (West Lafayette, Indiana)
S.D.P. (Hanover, New Hampshire)
J.T.T. (Urbana, Illinois)

October 2006

Chapter 1

The Isoperimetric Problem in Euclidean Space

Figure 1.1: Dido, Queen of Carthage. Engraving by Mathäus Merian the Elder, 1630. Used with permission of the Bayerische Staatsbibliothek, München.

Fleeing the vengeance of her brother, Dido lands on the coast of North Africa and founds the city of Carthage. Within the mythology associated with Virgil's saga lies one of the earliest problems in extremal geometric analysis. For the bargain which Dido agrees to with a local potentate is this: she may have that portion of land which she is able to enclose with the hide of a bull. Legend records Dido's ingenious and elegant solution: cutting the hide into a series of long thin strips, she marks out a vast circumference, forming the eventual line of the walls of ancient Carthage. This problem is a variant of what has become known as the classical *isoperimetric problem.*[1] In more precise terms it may be formulated as follows: *among all bounded, connected open regions in the plane with a fixed perimeter, characterize those regions with the maximal volume.* Needless to say, Dido's solution is correct: the extremal regions are precisely open circular planar discs.

Over the centuries, the isoperimetric problem (in various guises) has served to motivate substantial mathematical research in numerous areas. Indeed, the existence of the entire discipline of geometric measure theory can be attributed to a need to understand the precise setting for the study of classical questions in the calculus of variations such as the isoperimetric problem or Plateau's problem (to determine the surfaces of minimal area spanning a given closed curve in space). A wide array of techniques for the solution to the isoperimetric problem have been obtained from various fields:

- **Geometric measure theory**: The proof of the existence of an isoperimetric profile is based on compactness theorems for the space of functions of bounded variation. Consequently, *a priori* solutions are only guaranteed within the class of Caccioppoli sets (see Chapter 2 for a precise definition).
- **Differential geometry**: (Smooth) isoperimetric solutions are surfaces of constant mean curvature. The classification of such surfaces provides a characterization of isoperimetric profiles.
- **PDE**: The introduction of dynamic algorithms (curvature flow) provides a way to smoothly deform a given region so that the *isoperimetric ratio*

  $$(\text{Perimeter})^{n/n-1}/(\text{Volume})$$

  decreases monotonically. Provided such flows exist for all time (for instance, for special classes of initial data), the deformed regions converge, in a suitable sense, to a solution of the isoperimetric problem. The governing equations for such flows are nonlinear evolution partial differential equations, the most famous example being the volume constrained mean curvature flow.
- **Functional analysis**: Another analytic reformulation of the isoperimetric problem consists in viewing it as a best constant problem for a Sobolev inequality, relating mean values of a given smooth function with those of its derivatives.
- **Geometric function theory**: Symmetrization, in broad terms, refers to operations which replace a given mathematical object or region with one admitting a larger symmetry group. Suitable symmetrization techniques can be employed, in a similar vein to the previous point, to show that discs are isoperimetric solutions.

[1] See also Section 2.3.

To motivate our main topic (the isoperimetric problem in the Heisenberg group), we begin by recalling the classical isoperimetric inequality in Euclidean space.

Theorem 1.1. *For every Borel set Ω in $\mathbb{R}^n$ with finite perimeter $P(\Omega)$,*

$$\min\{|\Omega|^{(n-1)/n}, |\mathbb{R}^n \setminus \Omega|^{(n-1)/n}\} \leq C_{\text{iso}}(\mathbb{R}^n) P(\Omega), \tag{1.1}$$

where

$$C_{\text{iso}}(\mathbb{R}^n) = (n^{1-1/n}\omega_{n-1}^{1/n})^{-1}. \tag{1.2}$$

Equality holds in (1.1) *if and only if $\Omega = B(x, R)$ for some $x \in \mathbb{R}^n$ and $R > 0$.*

If Ω is C^1 smooth, then one can define the surface measure $d\sigma$, and

$$P(\Omega) = \int_{\partial\Omega} d\sigma.$$

For rougher domains one has the notion of perimeter introduced by De Giorgi,

$$\begin{aligned} P(\Omega) &= P(\Omega, \mathbb{R}^n) \\ &= \operatorname{Var}(\chi_\Omega, \mathbb{R}^n), \end{aligned} \tag{1.3}$$

where

$$\operatorname{Var}(f) = \sup_{\substack{G \in C_0^\infty(\mathbb{R}^n, \mathbb{R}^n) \\ |G| \leq 1}} \int_{\mathbb{R}^n} f \sum_{i=1}^n \partial_{x_i} G_i \, dx. \tag{1.4}$$

Roughly speaking, the *isoperimetric problem* consists in finding the "best constant" $C_{\text{iso}}(\mathbb{R}^n)$ and classifying the sets Ω such that the inequality in (1.1) becomes an equality. This problem has two classical, and equivalent, formulations:

- Among all bounded, connected open sets of fixed perimeter L, find one with maximum volume V.
- Among all bounded, connected open sets with fixed volume V, find one with minimum perimeter L.

For example, we recall that in $\mathbb{R}^2$, if we denote by A the area of an open set with finite perimeter and by L its perimeter, then

$$4\pi A \leq L^2, \tag{1.5}$$

where equality is achieved only for the disc. This result is classical and admits a variety of proofs. For the reader's convenience we recall in brief several elegant complex analytic proofs of the planar isoperimetric inequality (1.5), for relatively compact domains $\Omega \subset \mathbb{C}$ with boundary consisting of a single C^1 Jordan curve.

First proof of (1.5). Denoting $z = x_1 + \mathbf{i}\, x_2$ and $dA = dx_1 \wedge dx_2 = \frac{1}{2}\mathbf{i}\, dz \wedge d\overline{z}$ and using Green's theorem and Fubini's theorem together with the fact that the winding number of $\partial\Omega$ about any point in Ω is equal to 1, we find

$$4\pi A = \int_\Omega 2\pi\, \mathbf{i}\ dz \wedge d\overline{z} = \int_\Omega \int_{\partial\Omega} \frac{d\zeta\, dz \wedge d\overline{z}}{\zeta - z} = \int_{\partial\Omega} \int_{\partial\Omega} \frac{\overline{\zeta} - \overline{z}}{\zeta - z}\, dz d\zeta \le L^2$$

as desired. □

Second proof of (1.5). Let $\mathcal{D} = \{z : |z| < 1\}$ and let $f : \mathcal{D} \to \Omega$ be a Riemann map. Since $f' \neq 0$ in $\mathcal{D}$, we may choose an analytic map $g = \sqrt{f'}$. Then $A = \int_{\mathcal{D}} |g(z)|^4\, dA(z)$ and $L = \int_{\partial\mathcal{D}} |g(z)|^2\, |dz|$. Let $g(z) = \sum_{n=0}^\infty a_n z^n$. Expanding the integral representations for A and L and using the orthogonality of the functions $\exp(\mathbf{i}\, n\theta)$, $n \in \mathbb{Z}$, on $[0, 2\pi]$, we find

$$\begin{aligned} A &= \pi \sum_{k,m=0}^\infty \sum_{l=0}^{k+m} \frac{a_k \overline{a_l} a_m \overline{a_{k+m-l}}}{k+m+1} \\ &= \pi \sum_{j=0}^\infty \sum_{k,l=0}^{j} \frac{a_k a_{j-k} \overline{a_l a_{j-l}}}{j+1} \\ &= \pi \sum_{j=0}^\infty \frac{1}{j+1} \left| \sum_{k=0}^{j} a_k a_{j-k} \right|^2 \end{aligned} \tag{1.6}$$

and

$$L = 2\pi \sum_{j=0}^\infty |a_j|^2.$$

An application of the Cauchy–Schwarz inequality in (1.6) gives

$$4\pi A \le 4\pi^2 \sum_{j=0}^\infty \sum_{k=0}^{j} |a_k|^2 |a_{j-k}|^2 = L^2.$$ □

In each of the preceding proofs, the case of equality is easy to analyze.

The interplay between geometric extremal problems (such as the isoperimetric problem) and sharp analytic inequalities is witnessed in the following analytic proof of the planar isoperimetric inequality. The deep connection between the isoperimetric inequality and Sobolev–Poincaré inequalities is developed in detail in Chapters 6, 7 and 9, see especially Section 7.1.

Third proof of (1.5). Let ds denote the element of arc length along a Lipschitz curve $\partial\Omega$ which is the boundary of a domain $\Omega \subset \mathbb{R}^2$. Let x denote the position vector in $\mathbb{R}^2$. Without loss of generality we may assume $\int_{\partial\Omega} x ds = 0$. Using the divergence theorem we obtain

$$\begin{aligned} 2A &= \int_{\Omega} \operatorname{div} x \, dA = \int_{\partial\Omega} \langle x, \vec{n} \rangle \, ds \\ &\leq \int_{\partial\Omega} |x| \, ds \leq \sqrt{L} \left(\int_{\partial\Omega} |x|^2 \, ds \right)^{\frac{1}{2}}. \end{aligned}$$

Let us recall Wirtinger's inequality: if $f \in W^{1,2}([0, 2\pi])$ satisfies $\int_0^{2\pi} f(t) \, dt = 0$, then

$$\int_0^{2\pi} |f(t)|^2 \, dt \leq \int_0^{2\pi} \left| \frac{d}{dt} f \right|^2 \, dt, \tag{1.7}$$

with equality only when $f(t) = c_1 \cos t + c_2 \sin t$. (The Fourier analytic proof of (1.7) is an easy exercise.) Applying this inequality to the coordinate functions x_1, x_2 yields

$$2A \leq \sqrt{L} \left(\int_{\partial\Omega} |x|^2 \, ds \right)^{1/2} \leq \sqrt{L} \left[\left(\frac{L}{2\pi} \right)^2 \int_{\partial\Omega} \left| \frac{dx}{ds} \right|^2 \, ds \right]^{1/2} \leq \frac{L^2}{2\pi}$$

with equality if and only if Ω is a disc. □

One can also approach the isoperimetric problem in the plane through geometric methods. If the existence of an isoperimetric minimizer is assumed, symmetrization arguments may be employed to determine the nature of that minimizer. For example, if Ω is an isoperimetric minimizer, then it must be symmetric with respect to any line $\mathcal{L}$ which cuts it into two pieces of equal area. Suppose this is not true and there exists $\mathcal{L}$ dividing Ω into regions Ω_1, Ω_2 of equal area. Suppose that $P(\Omega_1) \neq P(\Omega_2)$ and, without loss of generality, that $P(\Omega_1) < P(\Omega_2)$. Then form $\tilde{\Omega}$ as the union of Ω_1 and the reflection of Ω_1 over $\mathcal{L}$. The domain $\tilde{\Omega}$ has the same area as Ω but smaller perimeter, violating the assumption that Ω is an isoperimetric minimizer. See Figure 1.2.

Furthermore, every isoperimetric minimizer is necessarily convex. Indeed, if Ω is not convex, then we can construct a line tangent to Ω at two points as in Figure 1.3. By reflecting a portion of the curve over this line, we would create a new domain $\tilde{\Omega}$ with greater area than Ω but the same perimeter. Thus Ω is not an isoperimetric minimizer.

Carrying out the preceding two operations repeatedly, we eventually deduce that Ω is convex and that, relative to a suitable coordinate system in $\mathbb{R}^2$, the tangent line is orthogonal to the position vector at every point of differentiability of $\partial\Omega$. Together with the convexity, this easily implies that $\partial\Omega$ is a circle.

One could propose yet another approach to the planar isoperimetric problem: beginning with any bounded, open, and connected set $\Omega_0 \subset \mathbb{C}$, deform Ω_0

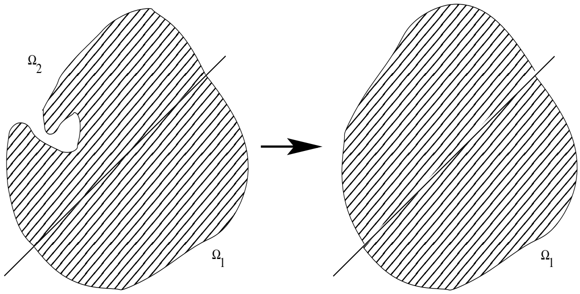


Figure 1.2: Isoperimetric sets in $\mathbb{R}^2$ have symmetry.

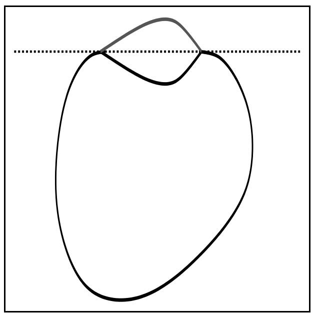


Figure 1.3: Isoperimetric sets in $\mathbb{R}^2$ are convex.

continuously in a flow Ω_t, $t > 0$, so that the Area(Ω_t) is constant and the perimeter $P(\Omega_t) = \text{Length}(\partial\Omega_t)$ is a decreasing function of t. Provided such a flow is well defined for all positive times t, one expects that the sets Ω_t will converge, as $t \to \infty$, to the disc of area Area(0). The most efficient way to reduce the length of the contour is to choose a velocity field $\vec{V}$ so that, if the boundary of Ω_t is represented by a curve $c_t : [0,1] \to \mathbb{R}^2$ and every point on the curve moves with velocity $\vec{V}$:

$$c_t' = \vec{V}(c_t), \quad \text{then} \quad \frac{d}{dt}\,\text{Length}(c_t)$$

is minimal among all choices of $\vec{V}$. More precisely, if the curve c_t is C^2 then

$$\frac{d}{dt}\,\text{Length}(c_t)|_{t=0} = \int_0^1 \frac{\langle c_t', c_t''\rangle}{|c_t'|} = -\int_0^1 \langle k_t\vec{n}_t, \vec{V}(c_t)\rangle,$$

where $k_t\vec{n}_t$ denotes the time-dependent curvature vector. The obvious choice is $\vec{V}(c_t) = k_t\vec{n}_t$. The resulting PDE is the famous *curve shrinking* flow equation

$$c_t' = k_t\vec{n}_t. \tag{1.8}$$

Solutions of the system (1.8) with initial data given by a smooth embedded curve are defined for all $t > 0$, remain smooth and embedded, and evolve asymptotically into a shrinking circle. In particular, the flow (1.8) clearly does not preserve the area enclosed by the curve. Imposing the area constraint introduces a non-local perturbation term in (1.8):

$$c_t' = \left(k_t - \frac{\int k_t}{\mathrm{Length}(c_t)}\right)\vec{n}_t. \tag{1.9}$$

A simple computation shows that

$$\frac{d}{dt}\,\mathrm{Area}(\Omega_t)\bigg|_{t=0} = \int_{c_t} \langle \vec{n}_t, c_t' \rangle \, ds = 0.$$

Any convex, closed embedded curve evolving by (1.9) stays convex and embedded for all times, and becomes circular asymptotically. Non-convex initial data gives rise to singularities in the flow and, in view of the non-local term in (1.9), most of the available techniques to study the flow past singular points cannot be applied. It is not clear if this asymptotic behavior persists for more general initial data, so this "geometric flow" approach only gives a partial answer to the isoperimetric problem in the plane, namely, *among all convex, bounded, connected simply connected, open sets in the plane with fixed area, the one with minimum perimeter is the disc.*

An integral geometric proof of the planar isoperimetric inequality. Integral geometry (also known as geometric probability) is the study of the measure-theoretic properties of random sets of geometric objects. The oldest problem of the subject is the famous Buffon needle problem. During the twentieth century, the relation between integral geometry and the theory of Lie groups and homogeneous spaces was formalized, and the field was further enriched by the reformulation of its classical concepts and methods in terms of stochastic processes.

To employ the techniques of integral geometry to prove the planar isoperimetric inequality (1.5) we introduce a measure $d\ell = dp\,d\phi$ on the space $\mathcal{L}$ of lines in $\mathbb{R}^2$ via the parametrization

$$(p,\phi) \in \mathbb{R} \times [0,\pi] \mapsto \ell = \ell_{p,\phi} \in \mathcal{L} \,:\, x\cos\phi + y\sin\phi - p = 0.$$

It is easy to verify that $d\ell$ is invariant under the action of Euclidean rigid motions of $\mathbb{R}^2$ on the space $\mathcal{L}$.

As discussed above, it suffices to establish (1.5) for convex sets, so let $\Omega \subset \mathbb{R}^2$ be a convex domain with area A and perimeter L. Fix a reference point $(x_0, y_0) \in \partial\Omega$ and parameterize $\partial\Omega$ by arc length from (x_0, y_0) in a specific (say, counterclockwise) direction.

For $\ell = \ell_{p,\phi} \in \mathcal{L}$, denote by $\sigma = \sigma(p)$ the length (possibly zero) of the chord $\ell_{p,\theta} \cap \Omega$. An easy application of Cavalieri's principle (e.g., Fubini's theorem) gives $\int_{\{p:\ell_{p,\phi}\cap\Omega\neq\emptyset\}} \sigma\, dp = A$ for each $\phi \in [0,\pi]$, whence

$$\int_{\mathcal{L}} \sigma\, d\ell = \pi A. \tag{1.10}$$

On the other hand,

$$\int_0^\pi \sigma^2 \, d\phi = 2 \int_0^\pi \int_0^\sigma r \, dr \, d\phi = 2A \tag{1.11}$$

since Ω is convex.

A generic line $\ell \in \mathcal{L}$ which intersects Ω has two points of intersection with $\partial\Omega$. Let (x, y) be one of these points, denote by s the arc length coordinates of (x, y), denote by θ the angle formed by ℓ with the support line to $\partial\Omega$ at (x, y) (note that almost every point in $\partial\Omega$ has a unique such support line and that we will only use the quantity $\sin\theta$ in computations) and denote by σ the length of the chord $\ell \cap \Omega$. For $\ell_i \in \mathcal{L}$, $i = 1, 2$, denote this data by (x_i, y_i), s_i, θ_i and σ_i. An easy calculation gives

$$2d\ell_i = \sin\theta_i \, ds_i \, d\theta_i \tag{1.12}$$

(the factor of 2 occurs due to the cardinality of $\ell_i \cap \partial\Omega$).

Consider the integral

$$I := \int_{\mathcal{L}\times\mathcal{L}} \frac{(\sigma_1 \sin\theta_2 - \sigma_2 \sin\theta_1)^2}{\sin\theta_1 \sin\theta_2} \, d\ell_1 \, d\ell_2.$$

Expanding and using (1.12) gives

$$I = \frac{1}{2} L^2 \int_0^\pi \sigma_1^2 \, d\theta_1 \int_0^\pi \sin^2\theta_2 \, d\theta_2 - 2 \left(\int_{\mathcal{L}} \sigma \, d\ell \right)^2.$$

Using (1.10) and (1.11), we obtain

$$I = \frac{\pi}{2} L^2 A - 2\pi^2 A^2 = \frac{\pi}{2} A(L^2 - 4\pi A).$$

Since $I \geq 0$, we conclude that $L^2 \geq 4\pi A$, as desired.

Observe that equality holds in the isoperimetric inequality if and only if $I = 0$, or equivalently, if $\sigma / \sin\theta$ is constant over all lines $\ell \in \mathcal{L}$ intersecting Ω. An easy geometric argument yields that Ω is a disc in this case.

1.1 Notes

The Euclidean (in particular the two-dimensional) isoperimetric problem has a long and interesting history. The book by Chavel [59] is a good reference, and most of the arguments in this chapter can be found there as well. The proof of the planar isoperimetric inequality via series estimates for the parametrizing Riemann map is due to Carleman; see Duren [89, Chapter 1]. For Wirtinger's inequality, see [134]. Other excellent references for the wide-ranging sphere of work motivated by the isoperimetric problem are the 1978 survey article of Osserman [216] and the books of Pólya–Szegö [229], and Bandle [27].

For applications of the machinery of geometric measure theory to verify higher regularity of minimizers arising in the calculus of variations, see [95], [125] and [195]. Key references for the functional analytic approach to variational problems are [97] and [197]. For symmetrization as an avenue to extremal variational problems in geometry and analysis, see [245], [17] and [229]. We also point out that the first proof of the isoperimetric inequality in the generality of Theorem 1.1 is due to De Giorgi [85].

Curvature flow has become an increasingly important topic in recent years after Perelman's work on the Ricci flow, with applications to Thurston's Geometrization Conjecture and the Poincaré conjecture [225, 226]. The approach to the Poincaré conjecture through the Ricci flow had been proposed by Hamilton as early as 1982, see, e.g., [133]. Curve shortening flow was intensively studied in the 1980s in the celebrated papers [110], [112] and [128]. The higher-dimensional analogue is the mean curvature flow, studied by Huisken [152].

The integral geometric proof of the planar isoperimetric inequality is a standard exercise which appears in all of the basic texts on the subject, see, e.g., [237, Section I.3.4]. The proof is generally attributed to Blaschke.

Chapter 2

The Heisenberg Group and Sub-Riemannian Geometry

In this chapter we provide a detailed description of the sub-Riemannian geometry of the first Heisenberg group. We describe its algebraic structure, introduce the horizontal subbundle (which we think of as *constraints*) and present the Carnot–Carathéodory metric as the least time required to travel between two given points at unit speed along horizontal paths. Subsequently we introduce the notion of sub-Riemannian metric and show how it arises from degenerating families of Riemannian metrics. For use in later chapters we compute some of the standard differential geometric apparatus in these Riemannian approximants.

2.1 The first Heisenberg group $\mathbb{H}$

We begin with a matrix model for the first Heisenberg group,[1] namely, the following subgroup of the group of three by three upper triangular matrices equipped with the usual matrix product:

$$\mathbb{H} = \left\{ \begin{pmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix} \in GL(3,\mathbb{R}) : x_1, x_2, x_3 \in \mathbb{R} \right\}. \tag{2.1}$$

The Heisenberg group $\mathbb{H}$ is an analytic Lie group of dimension 3.[2] Its Lie algebra $\mathfrak{h}$ can be equivalently defined either as the tangent space at the identity, $T_I\mathbb{H}$, or as the set of all left invariant tangent vectors. Clearly $\mathfrak{h}$ is a three-dimensional vector

[1] Also known as the *polarized Heisenberg group*.

[2] In other words, it is a group which is also an analytic manifold and for which the map $(x, y) \to xy^{-1}$ is an analytic transformation of $\mathbb{H} \times \mathbb{H}$ into $\mathbb{H}$.

space and we can identify it by explicitly computing the tangent spaces of $\mathbb{H}$. To do this, we left translate a fixed element

$$\begin{pmatrix} 1 & u_1 & u_3 \\ 0 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix},$$

which we will denote by the triplet (u_1, u_2, u_3), by one parameter families of matrices to form curves in $\mathbb{H}$, and then take derivatives along those curves to determine the tangent vectors. For example, let

$$U_1 = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \begin{pmatrix} 1 & \epsilon & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & u_1 & u_3 \\ 0 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & u_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$U_2 = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \epsilon \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & u_1 & u_3 \\ 0 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$U_3 = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \begin{pmatrix} 1 & 0 & \epsilon \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & u_1 & u_3 \\ 0 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The tangent space to $\mathbb{H}$ at (u_1, u_2, u_3) is spanned by the left invariant vector fields $\{U_1, U_2, U_3\}$. Computing the Lie brackets of these vector fields, we observe that $[U_1, U_2] = U_1U_2 - U_2U_1 = U_3$ while all other brackets are zero. We introduce a system of coordinates (generally known as *polarized coordinates* or *canonical coordinates of the second kind*) by identifying (u_1, u_2, u_3) with the matrix product

$$\exp(u_3U_3|_I)\exp(u_2U_2|_I)\exp(u_1U_1|_I),$$

here we have denoted by $\exp(U) = I + U + \frac{1}{2}U^2 + \cdots$ and by $U|_I$ the vector field U evaluated at the identity. We note explicitly the group law in polarized coordinates

$$(u_1, u_2, u_3)(v_1, v_2, v_3) = (u_1 + v_1, u_2 + v_2, u_3 + v_3 + u_1v_2).$$

The Heisenberg group $\mathbb{H}$ is the unique analytic, nilpotent Lie group whose background manifold is $\mathbb{R}^3$ and whose Lie algebra $\mathfrak{h}$ has the following properties:

- $\mathfrak{h} = V_1 \oplus V_2$, where V_1 has dimension 2 and V_2 has dimension 1, and
- $[V_1, V_1] = V_2$, $[V_1, V_2] = 0$ and $[V_2, V_2] = 0$.

The matrix presentation (2.1) is simply one way of realizing this general structure. We now turn to another, more intrinsic, presentation of $\mathbb{H}$ via a different system of coordinates. First, note that since $\mathfrak{h}$ is nilpotent the exponential map $\exp : \mathfrak{h} \to \mathbb{H}$

is a diffeomorphism. Fix an arbitrary basis X_1, X_2 of V_1 and let $X_3 = [X_1, X_2] \in V_2$. Recall the form which the Baker–Campbell–Hausdorff formula[3] takes in this (nilpotent step two) setting:

$$\exp^{-1}(\exp(x)\exp(y)) = x + y + \frac{1}{2}[x, y].$$

Here we have denoted by $x = x_1X_1 + x_2X_2 + x_3X_3 = (x_1, x_2, x_3)$ a generic point in $\mathfrak{h}$. Using the commutation relation $X_3 = [X_1, X_2]$ we obtain

$$\begin{aligned} x + y + \frac{1}{2}[x, y] &= (x_1 + y_1)X_1 + (x_2 + y_2)X_2 + (x_3 + y_3)X_3 \\ &\quad + \frac{1}{2}\Big\{x_1y_1[X_1, X_1] + (x_1y_2 - x_2y_1)[X_1, X_2] + x_2y_2[X_2, X_2]\Big\} \\ &= \left(x_1 + y_1, x_2 + y_2, x_3 + y_3 + \frac{1}{2}(x_1y_2 - x_2y_1)\right). \end{aligned}$$

We identify $\mathbb{H}$ with $\mathbb{C}\times\mathbb{R}$ by identifying (x_1, x_2, x_3) with $\exp(x_1X_1+x_2X_2+x_3X_3)$. The coordinates (x_1, x_2, x_3) are called *canonical coordinates of the first kind* or simply *exponential* coordinates and we will use the notation[4] $x = (x_1, x_2, x_3) = (z, x_3) \in \mathbb{H}$, with $z = x_1 + \mathbf{i}\, x_2 \in \mathbb{C}$ and $x_3 \in \mathbb{R}$. Using these coordinates the group law reads

$$(z, x_3)(w, y_3) = (z + w, x_3 + y_3 - \frac{1}{2}\operatorname{Im}(z\bar{w})). \tag{2.2}$$

In the remainder of this monograph we will almost invariably work with this model of the Heisenberg group, where the group law is given as in (2.2).[5] The group identity is $o = (0, 0, 0)$ while $x^{-1} = (-x_1, -x_2, -x_3)$. The group has a homogeneous structure given by the non-isotropic dilations $\delta_s(x) = (sx_1, sx_2, s^2x_3)$. An isomorphism between this latter model of $\mathbb{H}$ and the polarized Heisenberg group defined in (2.1) is obtained by mapping the element $\exp(x_1X_1 + x_2X_2 + x_3X_3)$ to the matrix

$$\begin{pmatrix} 1 & x_1 & x_3 + \frac{1}{2}x_1x_2 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

By moving in a left invariant fashion the frame X_1, X_2 and X_3 we obtain the explicit representation

$$X_1 = \partial_{x_1} - \frac{1}{2}x_2\partial_{x_3}, \quad X_2 = \partial_{x_2} + \frac{1}{2}x_1\partial_{x_3} \quad \text{and} \quad X_3 = \partial_{x_3}. \tag{2.3}$$

[3]See, for example, [72] for the statement of the full Baker–Campbell–Hausdorff formula.

[4]Here and in the following we make a slight abuse of notation and denote by x both the point in the group and the corresponding point $\exp^{-1} x$ in $\mathfrak{h}$.

[5]With some exceptions, however; note the use of the matrix model (2.1) in the application of Heisenberg geometry to jet spaces in Section 3.1.

Indeed, observe that the operation of left translation, $L_y(x) = yx$, has differential

$$dL_y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2}y_2 & \frac{1}{2}y_1 & 1 \end{pmatrix}. \tag{2.4}$$

A simple computation yields $X_i(xy) = dL_y X_i(x)$.

The Haar measure in $\mathbb{H}$ is simply the Lebesgue measure in $\mathbb{R}^3$. To see this, observe that the Euclidean volume form $dx_1 \wedge dx_2 \wedge dx_3$ at the origin is invariant under pull-back via left translation: $(L_y)_* dx_1 \wedge dx_2 \wedge dx_3 = dx_1 \wedge dx_2 \wedge dx_3$. Throughout the paper we will denote the measure of a Borel set $\Omega \subset \mathbb{H}$ as $|\Omega|$.

2.1.1 The horizontal distribution in $\mathbb{H}$

The left invariant frame X_1, X_2 is a basis for the horizontal fibration $H(x) = \mathrm{Ker}[dx_3 - \frac{1}{2}(x_1 dx_2 - x_2 dx_1)]$. Note that

$$\omega = dx_3 - \frac{1}{2}(x_1 dx_2 - x_2 dx_1) \tag{2.5}$$

is a contact form in $\mathbb{R}^3$, i.e., $\omega \wedge d\omega = -dx_1 \wedge dx_2 \wedge dx_3$.[6] According to the Darboux theorem, modulo local change of variables, ω is the only contact form in $\mathbb{R}^3$. See Figure 2.1 for a picture of the horizontal planes $H(x)$, first along the x_1 axis and second passing through various points in the z-plane.

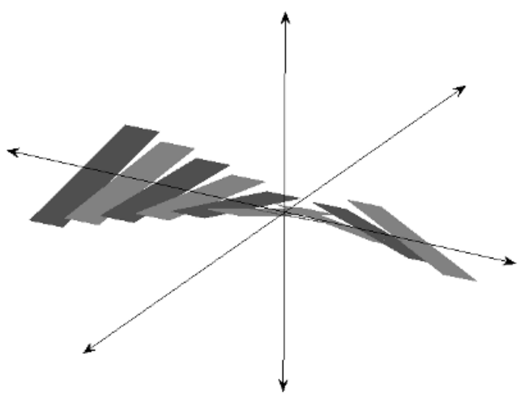


Figure 2.1: Examples of horizontal planes at different points.

The vector fields X_1 and X_2 are left invariant, first-order differential operators, homogeneous of order 1 with respect to the dilations δ_s. For any C^1 function ϕ

[6] An equivalent definition of contact form is that $d\omega$ restricted to $\mathrm{Ker}(\omega)$ is nondegenerate.

defined in an open set of $\mathbb{H}$ we denote by

$$\nabla_0 \phi = X_1 \phi X_1 + X_2 \phi X_2 \tag{2.6}$$

its *horizontal gradient.*

2.1.2 Higher-dimensional Heisenberg groups $\mathbb{H}^n$

Higher-dimensional analogs of the Heisenberg group are given by the Lie groups $\mathbb{H}^n$ which have as background manifold $\mathbb{C}^n \times \mathbb{R}$, and whose Lie algebra has a step two stratification $\mathfrak{h}^n = V_1 \oplus V_2$, where V_1 has dimension $2n$, V_2 has dimension 1, and $[V_1, V_1] = V_2$, $[V_1, V_2] = 0$ and $[V_2, V_2] = 0$. Using exponential coordinates

$$x = (z_1, \dots, z_n, x_{2n+1}) = (x_1 + \mathbf{i}\, x_{n+1}, \dots, x_n + \mathbf{i}\, x_{2n}, x_{2n+1})$$

we may express the group law as

$$xy = (z_1 + w_1, \dots, z_n + w_n, x_{2n+1} + y_{2n+1} - \frac{1}{2} \sum_{i=1}^{n} \operatorname{Im}(z_n \overline{w_n})),$$

where $x = (z_1, \dots, z_n, x_{2n+1})$ and $y = (w_1, \dots, w_n, y_{2n+1})$. The left invariant translates of the canonical basis at the identity are given by the vector fields $X_i = \partial/\partial x_i - \frac{1}{2} x_{i+n} \partial/\partial x_{2n+1}$, $X_{i+n} = \partial/\partial x_{i+n} + \frac{1}{2} x_i \partial/\partial x_{2n+1}$, $i = 1, \dots, n$, and $X_{2n+1} = \partial/\partial x_{2n+1}$. The first $2n$ vector fields span the horizontal distribution in $\mathbb{H}^n$; the corresponding homogeneous structure is provided by the parabolic dilations $\delta_s(x) = (sx_1, \dots, sx_{2n}, s^2 x_{2n+1})$.

2.1.3 Carnot groups

The Heisenberg groups are a particular example of a wide class of nilpotent, homogeneous, stratified Lie groups sometimes, referred to as *Carnot groups* in the literature. The Lie algebra $\mathfrak{g}$ of a Carnot group $\mathbb{G}$ has a stratification (or grading) $\mathfrak{g} = V_1 \oplus \cdots \oplus V_r$ satisfying:

- $[V_1, V_i] = V_{i+1}$, for $i = 1, \dots, r-1$, and
- $[V_j, V_r] = 0$, $j = 1, \dots, r$.

Elements in $\mathfrak{g}$ can be viewed either as tangent vectors to $\mathbb{G}$ at the identity element o, or as left invariant vector fields on $\mathbb{G}$. Following the notation in $\mathbb{H}$, we write L_y for the operation of left translation by $y \in \mathbb{G}$.

Choose a Riemannian metric with respect to which the V_i are mutually orthogonal. For $i = 1, \dots, r$ let $m_i = \dim(V_i)$ and denote by $\{X_{ij}\}$, $j = 1, \dots, m_i$ an orthonormal basis of V_i. Canonical coordinates of the second kind are given by

$$x = (x_{11}, x_{12}, \dots, x_{rm_r}) \leftrightarrow \exp(\sum_{i,j} x_{ij} X_{ij}). \tag{2.7}$$

A homogeneous structure on $\mathbb{G}$ is obtained by defining the dilations $[\delta_s(x)]_{ij} = s^i x_{ij}$. The *homogeneous dimension* of $\mathbb{G}$ is

$$Q = \sum_{i=1}^{r} i m_i.$$

Observe that the homogeneous dimension of $\mathbb{H}^n$ is $2n+2$. We will see the role of the homogeneous dimension in the metric geometry of $\mathbb{G}$ in Section 2.2.3.

The Haar measure on $\mathbb{G}$ coincides with the push-forward of the Lebesgue measure on the Lie algebra $\mathfrak{g}$ under the exponential map. It is easy to verify that the Jacobian determinant of the dilation $\delta_s : \mathbb{G} \to \mathbb{G}$ is constant, equal to s^Q.

As with the Heisenberg group, we define the *horizontal gradient* of a C^1 function $f : \mathbb{G} \to \mathbb{R}$ by

$$\nabla_0 f = \sum_{j=1}^{m_1} X_{1j} f \, X_{1j}.$$

At various points in this survey we will work in this general setting to emphasize the fact that certain results do not depend on the special structure of $\mathbb{H}$. However, we make no systematic attempt to present all results in the most general framework possible.

To conclude this section, we define the notion of a linear map between Carnot groups.

Definition 2.1. Given two Carnot groups $\mathbb{G}_1$, $\mathbb{G}_2$ with dilations δ_s^1 and δ_s^2, a map $L : \mathbb{G}_1 \to \mathbb{G}_2$ is a *horizontal linear map* if L is a group homomorphism which respects the dilations: $L(\delta_s^1 x) = \delta_s^2 L(x)$.

Example 2.2. Each horizontal linear map $L : \mathbb{H} \to \mathbb{H}$ takes the form $L(x) = Ax$, where the matrix A takes the form

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & ad - bc \end{pmatrix}$$

for some $a, b, c, d \in \mathbb{R}$. This is easy to verify from the definition.

2.2 Carnot–Carathéodory distance

2.2.1 CC distance I: *Constrained dynamics*

Let x and y be points in $\mathbb{H}$. For $\delta > 0$ we define the class $C(\delta)$ of absolutely continuous paths $\gamma : [0,1] \to \mathbb{R}^3$ with endpoints $\gamma(0) = x$ and $\gamma(1) = y$, so that

$$\gamma'(t) = a(t) X_1|_{\gamma(t)} + b(t) X_2|_{\gamma(t)} \tag{2.8}$$

and

$$a(t)^2 + b(t)^2 \le \delta^2 \tag{2.9}$$

for a.e. $t \in [0,1]$. Paths satisfying (2.8) are called *horizontal* or *Legendrian* paths. Note that (2.8) is equivalent with the statement

$$\omega(\gamma') = \gamma_3' - \frac{1}{2}(\gamma_1\gamma_2' - \gamma_2\gamma_1') = 0 \tag{2.10}$$

a.e., where ω is the contact form on $\mathbb{R}^3$ given in (2.5) and $\gamma = (\gamma_1, \gamma_2, \gamma_3)$.

Remark 2.3. Let $\pi : \mathbb{H} \to \mathbb{C}$ denote the projection $\pi(x) = x_1 + \mathbf{i}\, x_2$. Given any absolutely continuous planar curve $\alpha : [0,1] \to \mathbb{C}$ and a point $x = (\alpha(0), h) \in \mathbb{H}$ it is possible to *lift* α to a Legendrian path $\gamma : [0,1] \to \mathbb{H}$ starting at x satisfying $\pi(\gamma) = \alpha$. To accomplish this we let $\gamma_1(t) = \alpha_1(t), \gamma_2(t) = \alpha_2(t)$ and

$$\gamma_3(t) = h + \frac{1}{2}\int_0^t (\gamma_1\gamma_2' - \gamma_2\gamma_1')(\sigma)\, d\sigma.$$

It is easy to see that for any choice of $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$, the set $C(\delta)$ is nonempty for sufficiently large δ.

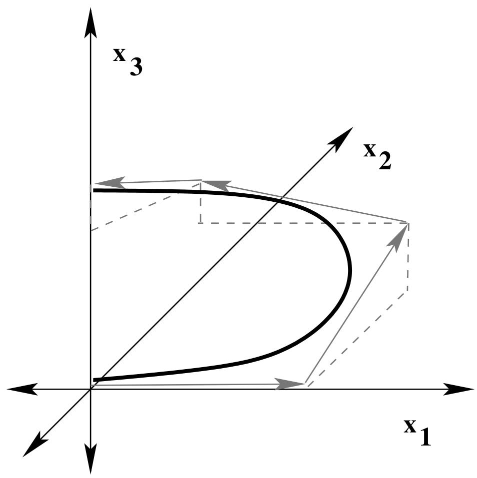


Figure 2.2: Horizontal paths connecting points in $\mathbb{H}$.

In Figure 2.2, we illustrate this fact by connecting the origin to the point $(0,0,1)$. First, we travel in the X_1 direction; as we begin at the origin, this is simply travel along the x_1 axis. From the point $(1,0,0)$, we travel in the X_2 direction to the point $\left(1,1,\frac{1}{2}\right)$. We then travel from this point in the $-X_1$ direction to the point $(0,1,1)$. Finally, we travel in the $-X_2$ direction, arriving at the terminus $(0,0,1)$. The smooth curve illustrated in Figure 2.2 which winds around and up the x_3 axis is a smooth horizontal curve that approximates this approach.

We define the Carnot–Carathéodory (CC) metric

$$d(x,y) = \inf\{\delta \text{ such that } C(\delta) \neq \emptyset\}.$$

A dual formulation is

$$d(x,y) = \inf \left\{ T : \begin{array}{c} \exists \gamma : [0,T] \to \mathbb{R}^3, \gamma(0) = x, \gamma(T) = y, \\ \text{and } \gamma' = aX_1|_\gamma + bX_2|_\gamma \text{ with } a^2 + b^2 \le 1 \text{ a.e.} \end{array} \right\},$$

that is, $d(x,y)$ is the shortest time that it takes to go from x to y, travelling at unit speed along horizontal paths. Since the vector fields X_1 and X_2 are left invariant, left translates of horizontal curves are still horizontal and it is easy to verify that $d(x,y) = d(y^{-1}x, 0)$.

Note that if γ is a horizontal curve, then so is its dilation $\delta_s\gamma$. In fact, if

$$\gamma'(t) = \sum_{i=1}^{2} \gamma_i'(t) X_i|_{\gamma(t)}$$

then

$$(\delta_s\gamma)' = \left(s\gamma_1', s\gamma_2', \frac{1}{2} s^2 (\gamma_1\gamma_2' - \gamma_2\gamma_1') \right) = \sum_{i=1}^{2} s\gamma_i' X_i|_{\delta_s\gamma}.$$

Moreover, if $\gamma \in C(\delta)$ then $\delta_s\gamma \in C(s\delta)$ (observe that the endpoints must be dilated as well). Consequently

$$d(\delta_s(x), \delta_s(y)) = s\delta(x,y),$$

in particular, this implies continuity of $x \mapsto d(x,0)$.

The Korányi gauge and metric. An equivalent distance on $\mathbb{H}$ is defined by the so-called Korányi metric

$$d_{\mathbb{H}}(x,y) = ||y^{-1}x||_{\mathbb{H}}$$

and Korányi gauge

$$||x||_{\mathbb{H}}^4 = (x_1^2 + x_2^2)^2 + 16x_3^2. \tag{2.11}$$

To verify that $d_{\mathbb{H}}$ is a metric, one needs to prove the triangle inequality:

$$d_{\mathbb{H}}(x,y) \le d_{\mathbb{H}}(x,z) + d_{\mathbb{H}}(z,y). \tag{2.12}$$

This can be done by direct computation as we now recall.

Proof of (2.12). By replacing $z^{-1}x$ with x and $y^{-1}z$ with y, it suffices to prove (2.12) in the case when $z = o = (0,0,0)$ is the identity element, i.e., to show that

$$||xy||_{\mathbb{H}} \le ||x||_{\mathbb{H}} + ||y||_{\mathbb{H}}. \tag{2.13}$$

Writing $x = (z, x_3)$ and $y = (w, y_3)$ and using the group law (2.2), we find

$$\begin{aligned} ||xy||_{\mathbb{H}}^4 &= |z+w|^4 + 16(x_3 + y_3 - \frac{1}{2}\operatorname{Im}(z\bar{w}))^2 \\ &= \left| |z+w|^2 + 4\,\mathbf{i}\,(x_3 + y_3 - \frac{1}{2}\operatorname{Im}(z\bar{w}) \right|^2 \\ &= \left| |z|^2 + 4\,\mathbf{i}\,x_3 + 2\bar{z}w + |w|^2 + 4\,\mathbf{i}\,y_3 \right|^2 \\ &\le \left(||x||_{\mathbb{H}}^2 + 2|z||w| + ||y||_{\mathbb{H}}^2 \right)^2 \le (||x||_{\mathbb{H}} + ||y||_{\mathbb{H}})^4. \end{aligned}$$

□

The lack of isotropy of the distance $d_{\mathbb{H}}$ follows precisely the lack of isotropy of the CC metric d. In particular, both behave like the Euclidean distance in horizontal directions (X_1 and X_2), and behave like the square root of the Euclidean distance in the missing direction (X_3). Clearly, $d_{\mathbb{H}}$ is homogeneous of order 1 with respect to the dilations (δ_s): $||\delta_s x||_{\mathbb{H}} = s||x||_{\mathbb{H}}$. Consequently, there exist constants $C_1, C_2 > 0$ so that

$$C_1||x||_{\mathbb{H}} \le d(x,0) \le C_2||x||_{\mathbb{H}}$$

for any $x \in \mathbb{H}$. This follows immediately from compactness of the Korányi unit sphere $\{x \in \mathbb{H} : ||x||_{\mathbb{H}} = 1\}$ and continuity of $x \mapsto d(x,0)$.

The Heisenberg group admits a conformal inversion in the Korányi unit sphere analogous to the classical Euclidean inversion $j(x) = x/|x|^2$ in $\mathbb{R}^n$. For $x \in \mathbb{H} \setminus \{o\}$, let

$$j_{\mathbb{H}}(x) = \left(\frac{-z}{|z|^2 + 4\,\mathbf{i}\,x_3}, \frac{-x_3}{|z|^4 + 16x_3^2} \right). \tag{2.14}$$

Since $||j_{\mathbb{H}}(x)||_{\mathbb{H}} = ||x||_{\mathbb{H}}^{-1}$, $j_{\mathbb{H}}$ preserves the Korányi unit sphere. The dilation property $j_{\mathbb{H}}(\delta_s x) = \delta_{1/s} x$ is also self-evident. Less obvious is the following Heisenberg analog of a classical Euclidean inversion relation:

$$d_{\mathbb{H}}(j_{\mathbb{H}}(x), j_{\mathbb{H}}(y)) = \frac{d_{\mathbb{H}}(x,y)}{||x||_{\mathbb{H}}||y||_{\mathbb{H}}}. \tag{2.15}$$

Proof of (2.15). As in the proof of (2.12), we write $x = (z, x_3)$ and $y = (w, y_3)$ and use the group law (2.2) to compute

$$\begin{aligned}
&d_{\mathbb{H}}(j_{\mathbb{H}}(x), j_{\mathbb{H}}(y))^4 \\
&= \Bigg| \left| \frac{z}{|z|^2 + 4\,\mathbf{i}\,x_3} - \frac{w}{|w|^2 + 4\,\mathbf{i}\,y_3} \right|^2 \\
&\qquad + 4\,\mathbf{i} \left(\frac{x_3}{|z|^4 + 16x_3^2} - \frac{y_3}{|w|^4 + 16y_3^2} + \frac{1}{2}\operatorname{Im}\left(\frac{z\overline{w}}{(|z|^2 + 4\,\mathbf{i}\,x_3)(|w|^2 - 4\,\mathbf{i}\,y_3)} \right) \right) \Bigg|^2 \\
&= \left| \frac{|z|^2 + 4\,\mathbf{i}\,x_3}{|z|^4 + 16x_3^2} - 2\frac{\overline{z}w}{(|z|^2 - 4\,\mathbf{i}\,x_3)(|w|^2 + 4\,\mathbf{i}\,y_3)} + \frac{|w|^2 - 4\,\mathbf{i}\,y_3}{|w|^4 + 16y_3^2} \right|^2 \\
&= \left| \frac{|w|^2 + 4\,\mathbf{i}\,y_3 - 2\overline{z}w + |z|^2 - 4\,\mathbf{i}\,x_3}{(|z|^2 - 4\,\mathbf{i}\,x_3)(|w|^2 + 4\,\mathbf{i}\,y_3)} \right|^2 = \frac{d_{\mathbb{H}}(x,y)^4}{||x||_{\mathbb{H}}^4 ||y||_{\mathbb{H}}^4}.
\end{aligned}$$

□

The relation between the Korányi gauge $||\cdot||_{\mathbb{H}}$ and Korányi inversion $j_{\mathbb{H}}$ will be pursued further in Sections 3.3 and 3.4, where we discuss the connections between the Heisenberg group, CR geometry, and Gromov hyperbolic geometry.

2.2.2 CC distance II: *Sub-Riemannian structure*

A sub-Riemannian metric on $\mathbb{H}$ is determined by any choice of inner product on the horizontal subbundle of the Lie algebra. Starting from this datum one may define the length of horizontal curves and equip $\mathbb{H}$ with the structure of a metric *length space*, which turns out to agree with the Carnot–Carathéodory metric. Since we have already made an arbitrary choice of coordinates to present $\mathbb{H}$, we may with no loss of generality assume that X_1 and X_2 form an orthonormal basis of each horizontal space $H(x)$ relative to this inner product. We extend this inner product to an inner product defined on the full tangent space, i.e., a Riemannnian metric, by requiring that the two layers in the stratification of the Lie algebra are orthogonal and that X_1, X_2 and X_3 form an orthonormal system. We denote this extended inner product by g_1 or $\langle\cdot,\cdot\rangle_1$ as dictation by specific situations.

Accordingly, we define the *horizontal length* of γ to be

$$\mathrm{Length}_{\mathbb{H},CC}(\gamma)=\int_0^1\sqrt{\langle\gamma'(t),X_1|_{\gamma(t)}\rangle_1^2+\langle\gamma'(t),X_2|_{\gamma(t)}\rangle_1^2}\,dt \tag{2.16}$$

and claim that

$$d(x,y)=\inf_\gamma \mathrm{Length}_{\mathbb{H},CC}(\gamma), \tag{2.17}$$

where the infimum is taken over all horizontal curves joining x and y. Note that if $\pi:\mathbb{H}\to\mathbb{C}$ denotes the projection $\pi(x)=x_1+\mathbf{i}\,x_2$, $d\pi:\mathfrak{h}\to\mathbb{C}$ denotes its differential, and $\mathrm{Length}_{\mathbb{C},\mathbf{Eucl}}(\cdot)$ denotes Euclidean length in the plane, then

$$\mathrm{Length}_{\mathbb{H},CC}(\gamma)=\mathrm{Length}_{\mathbb{C},\mathbf{Eucl}}(\pi(\gamma)). \tag{2.18}$$

To prove (2.17) we fix $x,y\in\mathbb{H}$ and let $\bar d=\inf_\gamma \mathrm{Length}_{\mathbb{H},CC}(\gamma)$. For any $\delta>d(x,y)$ we consider a curve $\gamma\in C(\delta)$ and note that

$$\bar d\le \mathrm{Length}_{\mathbb{C},\mathbf{Eucl}}(\pi(\gamma))\le\int_0^1\sqrt{a^2+b^2}\,dt\le\delta$$

by (2.9). Thus $\bar d\le d(x,y)$.

To prove the opposite inequality, let $\epsilon>0$ and choose a curve $\gamma:[0,1]\to\mathbb{H}$ connecting x and y so that $\mathrm{Length}_{\mathbb{C},\mathbf{Eucl}}(\pi(\gamma))=\bar d+\epsilon$. If we reparametrize γ to have constant velocity $|d\pi(\gamma')|=\bar d+\epsilon$, then $\gamma\in C(\bar d+\epsilon)$. Hence $\bar d+\epsilon\ge d(x,y)$. Since $\epsilon>0$ was arbitrary, $\bar d=d(x,y)$.

The next lemma shows that the Korányi and CC metrics generate the same infinitesimal structure.

Lemma 2.4. *If $\gamma:[0,1]\to\mathbb{R}$ is a C^1 curve and $t_i=i/n$, $i=1,\dots,n$, is a partition of $[0,1]$, then*

$$\limsup_{n\to\infty}\sum_{i=1}^n d_{\mathbb{H}}(\gamma(t_i),\gamma(t_{i-1}))=\begin{cases}\mathrm{Length}_{\mathbb{H},CC}(\gamma) & \textit{if } \gamma \textit{ is horizontal,}\\ \infty & \textit{otherwise.}\end{cases}$$

Proof. Set $\gamma(t) = (\gamma^1(t), \gamma^2(t), \gamma^3(t))$, $\gamma_i = \gamma(t_i) = (\gamma_i^1, \gamma_i^2, \gamma_i^3)$, and $\dot\gamma_i^j = (d\gamma^j/dt)(t_i)$. Then

$$\begin{aligned}
d_{\mathbb{H}}(\gamma(t_i), \gamma(t_{i-1})) &= ||\gamma_{i-1}^{-1}\gamma_i||_{\mathbb{H}} \\
&= \left\{ \left[(\gamma_i^1 - \gamma_{i-1}^1)^2 + (\gamma_i^2 - \gamma_{i-1}^2)^2 \right]^2 \right. \\
&\quad \left. + \left(\gamma_i^3 - \gamma_{i-1}^3 - \frac{1}{2}[\gamma_i^1(\gamma_i^2 - \gamma_{i-1}^2) - \gamma_i^2(\gamma_i^1 - \gamma_{i-1}^1)] \right)^2 \right\}^{\frac{1}{4}} \\
&= \frac{1}{n} \left\{ \left[\left(\dot\gamma_i^1 + o(1)\right)^2 + \left(\dot\gamma_i^2 + o(1)\right)^2 \right]^2 \right. \\
&\quad \left. + n^2 \left(\dot\gamma_i^3 - \frac{1}{2}[\gamma_i^1(\dot\gamma_i^2 + o(1)) - \gamma_i^2(\dot\gamma_i^1 + o(1))] \right)^2 \right\}^{\frac{1}{4}}.
\end{aligned}$$

The proof follows immediately from this derivation together with (2.16), (2.18) and (2.10). □

2.2.3 CC distance III: *Carnot groups*

The definition of a Carnot–Carathéodory distance can be extended easily to higher-dimensional Heisenberg groups and to general Carnot groups. Consider a Carnot group $\mathbb{G}$ with graded Lie algebra $\mathfrak{g} = V_1 \oplus \cdots \oplus V_r$, homogeneous dimension Q, and let $\langle \cdot, \cdot \rangle_{\mathbb{G}}$ be a left invariant inner product on V_1. Let $\{X_{1,j}\}_{j=1}^{m_1}$ be an orthonormal basis for V_1. If $\gamma : [0,1] \to \mathbb{G}$ is a horizontal path we can, following (2.16), define

$$\text{Length}_{\mathbb{G},CC}(\gamma) = \int_0^1 \left(\sum_{j=1}^{m_1} \langle \gamma'(t), X_{1,j}|_{\gamma(t)} \rangle_{\mathbb{G}}^2 \right)^{1/2} dt. \tag{2.19}$$

Then the Carnot–Carathéodory distance on $\mathbb{G}$ is defined to be

$$d(x,y) = \inf \text{Length}_{\mathbb{G},CC}(\gamma),$$

where the infimum is taken over all horizontal paths connecting x to y. Clearly, d is left invariant, moreover, the maps δ_s are indeed a family of dilations with respect to this metric:

$$d(\delta_s x, \delta_s y) = s d(x,y). \tag{2.20}$$

These properties of the CC metric imply analogous properties for the resulting Hausdorff measures $\mathcal{H}^\alpha$, $\alpha > 0$. Recall that the *α-dimensional Hausdorff measure* $\mathcal{H}^\alpha$ on a metric space (X,d) is the outer measure defined as

$$\mathcal{H}^\alpha(S) = \lim_{\delta \to 0} \inf_{\mathcal{B}} \sum_i (\text{diam}\, B_i)^\alpha,$$

where the infimum is taken over all coverings $\mathcal{B}$ of the set S by balls B_i with diameter $\operatorname{diam} B_i < \delta$. The standard implication

$$\mathcal{H}^\alpha(S) < \infty \quad \Rightarrow \quad \mathcal{H}^{\alpha'}(S) = 0 \text{ for all } \alpha' > \alpha$$

ensures the existence of a unique value $\alpha_0 = \alpha_0(S) \in [0, \infty]$ with the property that $\mathcal{H}^\alpha(S) = 0$ for $\alpha > \alpha_0$ and $\mathcal{H}^\alpha(S) = +\infty$ for $0 \leq \alpha < \alpha_0$. The value α_0 is the *Hausdorff dimension* of S.

From the left invariance and scaling properties of the CC metric, one easily deduces corresponding properties for the Hausdorff measures in $(\mathbb{G}, d)$:

$$\mathcal{H}^\alpha(L_y E) = \mathcal{H}^\alpha(E),$$
$$\mathcal{H}^\alpha(\delta_s E) = s^\alpha \mathcal{H}^\alpha(E),$$

for all $s, \alpha > 0$, $y \in \mathbb{G}$, and $E \subset \mathbb{G}$. In particular, for each α there exists $c(\alpha) \in [0, \infty]$ so that

$$\mathcal{H}^\alpha(B(x,r)) = c(\alpha) r^\alpha$$

for all $x \in G$ and $r > 0$, where $B(x,r)$ denotes the metric ball with center x and radius r in $(\mathbb{G}, d)$. When $0 < \alpha < Q$ we have $c(\alpha) = +\infty$, while for $\alpha > Q$ we have $c(\alpha) = 0$. In case $\alpha = Q$,

$$c(Q) = \mathcal{H}^Q(B(o,1)) \in (0, +\infty).$$

Thus the Hausdorff dimension of $(\mathbb{G}, d)$ is Q; indeed $(\mathbb{G}, d)$ is an *Ahlfors Q-regular space* and $\mathcal{H}^Q$ agrees (up to a constant multiplicative factor) with the Haar measure on $\mathbb{G}$. In particular, for any non-abelian Carnot group $\mathbb{G}$ the Hausdorff dimension strictly exceeds the topological dimension; this gives $(\mathbb{G}, d)$ fractal character (in the sense of the term *fractal* advocated by Mandelbrot).

The gauge norm (2.11) has several natural extensions to general Carnot groups. Here we recall one of the more computationally friendly ones:

$$||x||_{\mathbb{G}}^{2r!} = \sum_{i=1}^{r} \sum_{j=1}^{m_i} |x_{ij}|^{\frac{2r!}{i}}, \qquad x = (x_{11}, \ldots, x_{rm_r}) \in \mathbb{G}. \tag{2.21}$$

In contrast with the Heisenberg situation, $|| \cdot ||_{\mathbb{G}}$ is typically only a quasinorm rather than a true norm: the inequality $||xy||_{\mathbb{G}} \leq ||x||_{\mathbb{G}} ||y||_{\mathbb{G}}$ must be replaced by $||xy||_{\mathbb{G}} \leq C ||x||_{\mathbb{G}} ||y||_{\mathbb{G}}$ for some (absolute) constant $C < \infty$. The latter fact easily follows from the Baker–Campbell–Hausdorff formula. As was the case in the Heisenberg group, the gauge quasimetric $d_{\mathbb{G}}(x,y) = ||y^{-1}x||_{\mathbb{G}}$ and the Carnot–Carathéodory metric d are comparable.

2.3 Geodesics and bubble sets

In this section we describe the length minimizing curves joining pairs of points in the Heisenberg group and define the so-called *bubble sets* which appear in Pansu's conjecture on the isoperimetric profile of $\mathbb{H}$. Without loss of generality, and thanks to left-invariance of the CC metric, we may assume that the two points are the origin $o = (0,0,0)$ and $x = (x_1, x_2, x_3)$. Let $\gamma : [0,1] \to \mathbb{H}$ denote a Legendrian curve joining o and x. Let S be the region in the (x_1, x_2)-plane bounded by $\pi(\gamma)$ and by the segment joining $\pi(x)$ to the origin, and let $\tilde{\gamma}$ be the closed curve obtained by closing $\pi(\gamma)$ with the segment. By Stokes' theorem we have

$$\begin{aligned} x_3 &= \int_0^1 \gamma_3'(t)\,dt = \frac{1}{2}\int_0^1 (\gamma_1\gamma_2' - \gamma_2\gamma_1')(t)\,dt \\ &= \frac{1}{2}\int_{\tilde{\gamma}} x_1\,dx_2 - x_2\,dx_1 = \int_S dx_1 \wedge dx_2 = \mathrm{Area}(S). \end{aligned} \tag{2.22}$$

In view of Remark 2.3 and (2.22) we can rephrase the problem of finding the Legendrian curve from o to x with minimal length with the following problem: *Find the plane curve from the origin to* (x_1, x_2) *with minimum length, subject to the constraint that the region S delimited by the curve and the segment joining* $(0,0)$ *to* (x_1, x_2) *has fixed area.* This is one formulation of *Dido's problem*, closely related to the isoperimetric problem, and is solved by choosing the plane curve to be an arc of a circle.

In conclusion, a length minimizing curve between o and x is the lift of a circular arc joining the origin in $\mathbb{C}$ with (x_1, x_2), whose convex hull has area x_3. The family of such curves emanating from o is parameterized by $e^{\mathbf{i}\,\phi} \in \mathbb{S}^1$ and $c \in \mathbb{R}$ and is given explicitly in the form

$$\gamma_{c,\phi}(s) = \left(e^{\mathbf{i}\,\phi}\frac{1 - e^{-\mathbf{i}\,cs}}{c}, \frac{cs - \sin(cs)}{2c^2} \right); \tag{2.23}$$

it is length minimizing over any interval of length $2\pi/|c|$. In particular, if $c = 0$ then $\gamma_{c,\phi}$ is a straight line through o in the xy-plane. We call c the *curvature* of the geodesic arc $\gamma_{c,\phi}$.

If $x = (0,0,x_3)$ lies on the x_3-axis, then the projection of the geodesic in $\mathbb{H}$ joining o to x is a circle of area x_3 passing through the origin. Clearly there are infinitely many such circles, so geodesics are not unique in this case. Choosing $t = cs$, $R = 1/c$ and $\phi = 0$ in (2.23) gives the following representation:

$$\gamma(t) = (Re^{\mathbf{i}\,\phi}(1 - e^{-\mathbf{i}\,t}), \frac{1}{2}R^2(t - \sin t)), \tag{2.24}$$

with $0 \le t \le 2\pi$ and $R^2 = x_3/\pi$.

Rotating such a geodesic about the x_3-axis produces a surface of revolution Σ whose profile curve is given parametrically as

$$t \mapsto \left(2R\sin(t/2), \frac{1}{2}R^2(t - \sin t) \right).$$

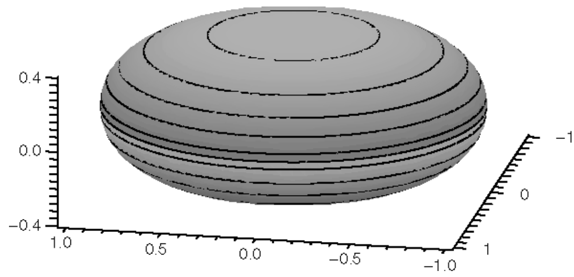


Figure 2.3: The conjectured isoperimetric set in $\mathbb{H}^1$.

Alternatively, Σ is the boundary of the open set Ω consisting of all points $(z, x_3) \in \mathbb{H}$ such that

$$R^2\left(\frac{\pi}{2} - \arccos(\frac{|z|}{2R})\right) - \frac{|z|}{2}\sqrt{R^2 - \frac{|z|^2}{4}} < |x_3| \\ < R^2\left(\frac{\pi}{2} + \arccos(\frac{|z|}{2R}\right) + \frac{|z|}{2}\sqrt{R^2 - \frac{|z|^2}{4}}. \quad (2.25)$$

To simplify the notation we dilate this ball by a factor of 2 and translate vertically by $-\pi R^2/2$. The resulting domains

$$\mathcal{B}(o, R) := \{(z, x_3) \in \mathbb{H} : |x_3| < f_R(|z|)\}, \quad (2.26)$$

$$f_R(r) = \frac{1}{4}\left(R^2 \arccos(\frac{r}{R}) + r\sqrt{R^2 - r^2}\right),$$

are the conjectured extremals for the sub-Riemannian isoperimetric problem in $\mathbb{H}$ and are often called *bubble sets*. See Figure 2.3 and the introduction to Chapter 8. For a computation of the horizontal mean curvature of $\partial\mathcal{B}(o, R)$, see Section 4.3. We emphasize that the boundary of the set $\mathcal{B}(o, R)$ is C^2 but not C^3.

2.4 Riemannian approximants to $\mathbb{H}$

The Heisenberg group equipped with the CC metric may be realized as the Gromov–Hausdorff limit of a sequence of Riemannian manifolds $(\mathbb{R}^3, g_L)$, as $L \to \infty$. This *Riemannian approximation scheme* plays a central role in the development of sub-Riemannian submanifold geometry and geometric measure theory in this survey.

2.4.1 The g_L metrics

To describe the Riemannian approximants to $\mathbb{H}$, let $L > 0$ and define a metric g_L on $\mathbb{R}^3$ so that the left invariant basis $X_1, X_2, X_3/\sqrt{L}$ of $\mathfrak{h}$ is orthonormal.[7] This family of metrics is essentially obtained as an anisotropic blow-up of the Riemannian metric g_1 defined in Section 2.2.2. We may represent g_L explicitly in exponential coordinates $x = (x_1, x_2, x_3)$ via the positive definite matrix

$$g_L(x) = \begin{pmatrix} 1 + \frac{1}{4}x_2^2 L & -\frac{1}{4}x_1 x_2 L & -\frac{1}{2}x_2 L \\ -\frac{1}{4}x_1 x_2 L & 1 + \frac{1}{4}x_1^2 L & \frac{1}{2}x_1 L \\ -\frac{1}{2}x_2 L & \frac{1}{2}x_1 L & L \end{pmatrix}. \tag{2.27}$$

Note that

$$g_L(x) = C^T I_L C,$$

where

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2}x_2 & \frac{1}{2}x_1 & 1 \end{pmatrix}$$

is the matrix defining left translation (see (2.4)) and

$$I_L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & L \end{pmatrix}.$$

Observe that the Riemannian volume element in $(\mathbb{H}, g_L)$ is

$$\sqrt{\det g_L}\, dx_1 \wedge dx_2 \wedge dx_3 = \sqrt{L}\, dx_1 \wedge dx_2 \wedge dx_3.$$

When calculating with the Riemannian metric g_L, we will sometimes use $\langle \cdot, \cdot \rangle_L$ to denote the inner product on vectors.[8] In other words, for $\vec{a} = \sum_{i=1}^3 a_i X_i$ and $\vec{b} = \sum_{i=1}^3 b_i X_i$ in $T\mathbb{H}$,

$$\langle \vec{a}, \vec{b}\, \rangle_L = a_1 b_1 + a_2 b_2 + L a_3 b_3. \tag{2.28}$$

As usual, the length of a vector is given as

$$|\vec{a}|_L = \langle \vec{a}, \vec{a} \rangle_L^{1/2}.$$

[7]These metrics also occur as restrictions of the Bergman metric on horospheres in the Siegel domain

$$D = \{(\xi_1, \xi_2) \in \mathbb{C}^2 : \mathrm{Im}\ \xi_2 - |\xi_1|^2 > 0\},$$

to orbits of $\mathbb{H}$. See Section 3.3.

[8]Recall that we will always reserve the notation $\langle \cdot, \cdot \rangle$ (with no subscript) for the standard Euclidean inner product in any dimension.

We note that we can recover the sub-Riemannian inner product on $\mathbb{H}$ by restricting $\langle\cdot,\cdot\rangle_L$ to the horizontal directions. Moreover, in the limit as $L \to \infty$, the only vectors of finite length are those which lie in the horizontal subbundle. We can capitalize on this observation by looking at the lengths of curves in the Riemannian approximants. Suppose $\gamma : [0,1] \to \mathbb{H}$ is a C^1 curve and that $\gamma' = a_1X_1 + a_2X_2 + a_3X_3$. Then

$$\begin{aligned} \mathrm{Length}_L(\gamma) &= \int_0^1 \langle\gamma'(t),\gamma'(t)\rangle_L^{1/2}\,dt \\ &= \int_0^1 (a_1(t)^2 + a_2(t)^2 + La_3(t)^2)^{1/2}\,dt, \end{aligned} \tag{2.29}$$

where we have written $\mathrm{Length}_L = \mathrm{Length}_{d_L}$ for simplicity. Note that γ has finite length in the limit as $L \to \infty$ if and only if $a_3 = 0$, i.e., γ is a horizontal curve. We define d_L to be the standard path metric associated to g_L.

2.4.2 Levi-Civita connection and curvature in the Riemannian approximants

In this section, we compute the sectional, Ricci and scalar curvatures of the Heisenberg group with respect to g_L. To this end, we use the Levi-Civita connection ∇ on $(\mathbb{H}, g_L)$. Given vector fields U, V, W on a Riemannian manifold (M, g), the *Kozul identity* states

$$\begin{aligned} \langle\nabla_U V, W\rangle_L = \frac{1}{2}\{&U\langle V,W\rangle_L + V\langle W,U\rangle_L - W\langle U,V\rangle_L \\ &- \langle W,[V,U]\rangle_L - \langle[V,W],U\rangle_L - \langle V,[U,W]\rangle_L\} \end{aligned} \tag{2.30}$$

where $\langle\cdot,\cdot\rangle_L$ is the inner product associated to g_L. To make the computation more clear, we introduce the functions

$$\alpha_{ijk} = \langle\tilde{X}_i, [\tilde{X}_j, \tilde{X}_k]\rangle_L$$

where $\tilde{X}_i = X_i$ for $i = 1, 2$ and $\tilde{X}_3 = L^{-1/2}X_3$. Note that

$$\langle\nabla_{\tilde{X}_i}\tilde{X}_j, \tilde{X}_k\rangle_L = -\frac{1}{2}\left(\alpha_{kji} + \alpha_{ijk} + \alpha_{jik}\right).$$

As the only nontrivial Lie bracket is $[X_1, X_2] = X_3 = \sqrt{L}\tilde{X}_3$, we have $\alpha_{312} = \sqrt{L}$, $\alpha_{321} = -\sqrt{L}$, and $\alpha_{ijk} = 0$ for all other triples (i, j, k). Consequently,

$$\nabla_{\tilde{X}_i}\tilde{X}_j = \begin{cases} \frac{1}{2}X_3 & \text{if } (i,j) = (1,2), \\ -\frac{1}{2}X_3 & \text{if } (i,j) = (2,1), \\ -\frac{\sqrt{L}}{2}X_2 & \text{if } (i,j) \in \{(1,3),(3,1)\}, \\ \frac{\sqrt{L}}{2}X_1 & \text{if } (i,j) \in \{(2,3),(3,2)\}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

Given a Riemannian manifold (M, g) with Riemannian connection ∇, we recall that the curvature tensor for M is defined by

$$R(U,V)W = \nabla_U \nabla_V W - \nabla_V \nabla_U W - \nabla_{[U,V]} W.$$

We remark that, in the literature, the sign of the curvature tensor is sometimes reversed. We compute the sectional curvatures of the two-planes spanned by the basis vectors $\tilde{X}_i$ and $\tilde{X}_j$: $K_{ij} = \langle R(\tilde{X}_i, \tilde{X}_j)\tilde{X}_i, \tilde{X}_j\rangle_L$:

$$\begin{aligned} K_{12} &= \langle R(X_1,X_2)X_1, X_2\rangle_L = \langle \nabla_{X_1}\nabla_{X_2}X_1 - \nabla_{X_2}\nabla_{X_1}X_1 - \nabla_{X_3}X_1, X_2\rangle_L \\ &= \left\langle \nabla_{X_1}\left(-\frac{1}{2}X_3\right) - \nabla_{X_2}(0) + \frac{L}{2}X_2, X_2 \right\rangle_L \\ &= \frac{L}{4} + \frac{L}{2} = \frac{3L}{4}, \end{aligned} \tag{2.32}$$

$$\begin{aligned} K_{13} &= \langle R(X_1,\tilde{X}_3)X_1, \tilde{X}_3\rangle_L = \left\langle \nabla_{X_1}\nabla_{\tilde{X}_3}X_1 - \nabla_{\tilde{X}_3}\nabla_{X_1}X_1, \tilde{X}_3 \right\rangle_L \\ &= \left\langle \nabla_{X_1}\left(-\frac{\sqrt{L}}{2}\right)X_2 - \nabla_{\tilde{X}_3}(0), \tilde{X}_3 \right\rangle_L \\ &= \left\langle -\frac{\sqrt{L}}{4}X_3, \tilde{X}_3 \right\rangle_L = -\frac{L}{4} \end{aligned} \tag{2.33}$$

and

$$\begin{aligned} K_{23} &= \langle R(X_2,\tilde{X}_3)X_2, \tilde{X}_3\rangle_L = \left\langle \nabla_{X_2}\nabla_{\tilde{X}_3}X_2 - \nabla_{\tilde{X}_3}\nabla_{X_2}X_2, \tilde{X}_3 \right\rangle_L \\ &= \left\langle \nabla_{X_2}\frac{\sqrt{L}}{2}X_1 - \nabla_{\tilde{X}_3}(0), \tilde{X}_3 \right\rangle_L \\ &= \left\langle -\frac{\sqrt{L}}{4}X_3, \tilde{X}_3 \right\rangle_L = -\frac{L}{4}. \end{aligned} \tag{2.34}$$

In fact, the full Riemannian curvature tensor $R_{ijkl} = \langle R(\tilde{X}_i, \tilde{X}_j)\tilde{X}_k, \tilde{X}_l\rangle$ is

$$R_{ijkl} = \begin{cases} \frac{3L}{4} & \text{if } (ijkl) = (1212) \text{ or } (2121), \\ -\frac{3L}{4} & \text{if } (ijkl) = (1221) \text{ or } (2112), \\ -\frac{L}{4} & \text{if } (ijkl) = (1313), (3131), (2323) \text{ or } (3232), \\ \frac{L}{4} & \text{if } (ijkl) = (1331), (3113), (2332) \text{ or } (3223), \\ 0 & \text{otherwise.} \end{cases}$$

Next, the Ricci curvatures $\mathrm{Ric}_i = K_{i1} + K_{i2} + K_{i3}$ are

$$\mathrm{Ric}_1 = \mathrm{Ric}_2 = \frac{L}{2} \quad \text{and} \quad \mathrm{Ric}_3 = -\frac{L}{2},$$

while the scalar curvature $\sigma = \mathrm{Ric}_1 + \mathrm{Ric}_2 + \mathrm{Ric}_3$ is

$$\sigma = \frac{L}{2}.$$

Observe that the sectional, Ricci and scalar curvatures all diverge as $L \to \infty$.

2.4.3 Gromov–Hausdorff convergence

In this section we define the Gromov–Hausdorff distance between two metric spaces and the corresponding notion of convergence. Recall that for two subsets E_1, E_2 of a metric space Z, the Hausdorff distance between E_1 and E_2 is given by

$$\mathrm{Haus}_Z(E_1, E_2) = \inf\{\epsilon > 0 | E_1 \subset (E_2)_\epsilon, E_2 \subset (E_1)_\epsilon\}$$

where $(E_i)_\epsilon = \{z \in Z : \mathrm{dist}(z, E_i) < \epsilon\}$ is the ϵ-neighborhood of E_i in (Z, d).

Definition 2.5. Let (X, d_X) and (Y, d_Y) be metric spaces. The *Gromov–Hausdorff distance* between X and Y is given by

$$d_{GH}(X, Y) = \inf_{f,g,Z} \mathrm{Haus}_Z(f(X), g(Y)),$$

where the infimum is taken over all metric spaces Z and isometric embeddings f and g of X and Y (respectively) into Z.

Using this metric, we have a notion of convergence:

Definition 2.6. A sequence of compact metric spaces (X_n) *Gromov–Hausdorff converges* to a compact metric space X if $d_{GH}(X_n, X) \to 0$ as $n \to \infty$.

The notion of convergence in Definition 2.6 is unnatural for noncompact spaces and limits. For example, we would like to assert that the dilated spheres $(\mathbb{S}^n, \lambda d_{\mathbb{S}^n})$ ($d_{\mathbb{S}^n}$ the geodesic distance on $\mathbb{S}^n$) converge to $\mathbb{R}^n$ as $\lambda \to \infty$. However, according to Definition 2.5, $d_{GH}((\mathbb{S}^n, \lambda d_{\mathbb{S}^n}), \mathbb{R}^n) = +\infty$ for all $\lambda > 0$. In the general case we work in the category of proper pointed metric length spaces. Recall that a metric space (X, d) is *proper* if all closed balls in X are compact, and *length* if the distance between any two points x and y is realized by the infimum of the lengths of rectifiable paths joining x to y. Note that every proper length space is in fact geodesic. By a *pointed* metric space (X, d, x) we mean a metric space (X, d) equipped with a fixed basepoint $x \in X$.

Definition 2.7. A sequence of proper pointed length spaces (X_n, d_n, x_n) *Gromov–Hausdorff converges* to a proper pointed length space (X, d, x) if the sequence of closed balls $\overline{B}_{X_n}(x_n, r)$ Gromov–Hausdorff converges (in the sense of Definition 2.6) to $\overline{B}_X(x, r)$, uniformly in r.

The following proposition abstracts the key geometric features of the Riemannian approximation scheme for the sub-Riemannian Heisenberg group which guarantees that the approximating manifolds converge in the Gromov–Hausdorff sense.

Proposition 2.8. *Let X be a set equipped with a family of metrics $(d_t)_{t\geq 0}$ generating a common topology. For K compact in X, let:*

$$\omega_K(\epsilon) := \sup_{x,y\in K, t\geq 0} d_t(x,y) - d_{t+\epsilon}(x,y).$$

Assume:

(i) *For each $t \geq 0$, (X, d_t) is a proper length space.*
(ii) *For fixed $x, y \in X$, the function $t \mapsto d_t(x,y)$ is non-increasing.*
(iii) *For each compact set K in X, $\omega_K(\epsilon) \to 0$ as $\epsilon \to 0$.*

Then (X, d_t) converges, in the sense of pointed Gromov–Hausdorff convergence, to (X, d_0).

Proof of Proposition 2.8. By hypothesis (ii) and the definition of ω_K we easily verify the following additional facts for each compact set K:

(iv) the map $\epsilon \mapsto \omega_K(\epsilon)$ is increasing in ϵ,
(v) ω_K is sublinear: $\omega_K(a+b) \leq \omega_K(a) + \omega_K(b)$ for all $a, b \geq 0$,
(vi) if we denote by $B_t(x_0, R)$ the closed metric ball with center x_0 and radius R in the metric space (X, d_t), $t \geq 0$, then

$$B_0(x_0, R) \subset B_t(x_0, R) \subset B_0(x_0, R + \omega_K(t))$$

for any x_0, $R > 0$ and $t \geq 0$ so that $B_t(x_0, R) \subset K$.

From (vi) and (i) we further conclude

(vii) For x_0, R and t as in (vi), to each $y \in B_t(x, 0, R)$ there corresponds a point $x \in B_0(x_0, R)$ with $d_t(x,y) \leq \omega_K(t)$.

We now establish the desired conclusion. Fix a basepoint $x_0 \in X$. It suffices to prove that the compact metric balls $B_t(x_0, R)$ converge (in the sense of Definition 2.6) to $B_0(x_0, R)$, for each $R > 0$. We restrict attention to $t \in [0,1]$, let $K = B_1(x_0, R)$, and equip the space

$$Z = K \times [0,1]$$

with the metric

$$d_Z((x,t),(x',t')) = d_{\max\{t,t'\}}(x,x') + \omega_K(|t-t'|) + |t-t'|.$$

We claim that d_Z is a metric. If $d_Z((x,t),(x',t')) = 0$ then $t = t'$ and $d_t(x,x') = 0$, hence also $x = x'$ (since d_t is a metric).[9] Next we verify the triangle inequality.

[9] We include the term $|t-t'|$ in the definition of d_Z for this argument, in order to conclude $t = t'$; note that we have no guarantee that $\epsilon > 0 \Rightarrow \omega_K(\epsilon) > 0$.

Let $(x,t),(x',t'),(x'',t'')\in Z$. If $\max\{t,t',t''\}\neq t''$, then

$$\begin{aligned} d_Z((x,t),(x',t')) &= d_{\max\{t,t',t''\}}(x,x') + \omega_K(|t-t'|) + |t-t'| \\ &\le d_{\max\{t,t''\}}(x,x'') + d_{\max\{t'',t'\}}(x'',x') + \omega_K(|t-t''|) \\ &\qquad + \omega_K(|t''-t'|) + |t-t''| + |t''-t'| \\ &= d_Z((x,t),(x'',t'')) + d_Z((x'',t''),(x',t')) \end{aligned}$$

where we used (ii) and (v) in the middle step. If $\max\{t,t',t''\}=t''$, then

$$\begin{aligned} d_Z((x,t),(x',t')) &= d_{\max\{t,t'\}}(x,x') + \omega_K(|t-t'|) + |t-t'| \\ &\le d_{t''}(x,x') + \omega_K(t''-\max\{t,t'\}) + \omega_K(|t-t'|) + |t-t'| \end{aligned}$$

by the definition of ω_K

$$\begin{aligned} &\le d_{t''}(x,x'') + d_{t''}(x'',x') + \omega_K(t''-t) + \omega_K(t''-t') + |t-t''| + |t''-t'| \\ &= d_Z((x,t),(x'',t'')) + d_Z((x'',t''),(x',t')) \end{aligned}$$

by (iv).

It is clear that the map $x\mapsto(x,t)$ is an isometric embedding of $B_t(x_0,R)$ into Z. To complete the proof, it suffices to verify that the Hausdorff distance (in Z) between $B_t(x_0,R)\times\{t\}$ and $B_0(x_0,R)\times\{0\}$ tends to zero as $t\to 0$. In fact, we claim that

$$\mathrm{Haus}_Z(B_t(x_0,R)\times\{t\}, B_0(x_0,R)\times\{0\}) \le 2\omega_K(t)+t; \tag{2.35}$$

the result then follows from assumption (iii).

To see that (2.35) is true: if $x\in B_0(x_0,R)$, then $x\in B_t(x_0,R)$ and

$$d_Z((x,t),(x,0)) = \omega_K(t)+t \le 2\omega_K(t)+t,$$

while if $y\in B_t(x_0,R)$ we choose x as in (vii) and conclude

$$d_Z((y,t),(x,0)) = d_t(x,y)+\omega_K(t)+t \le 2\omega_K(t)+t. \qquad \square$$

2.4.4 Carnot–Carathéodory geodesics and Gromov–Hausdorff convergence

The CC geodesics in the Heisenberg group can be recovered through the approximation scheme using the geodesics in the Riemannian manifolds $(\mathbb{R}^3,g_L)$. In this section we sketch two different ways of recovering this result. In the first approach we find a differential equation whose solutions are the CC geodesics; in the second approach we compute the g_L-geodesics explicitly and show that they converge to length minimizing curves in $(\mathbb{H},d)$.

Let $\gamma : [0,1] \to \mathbb{H}$ be a Lipschitz curve and let $\omega = dx_3 - \frac{1}{2}(x_1 dx_2 - x_2 dx_1)$ denote the contact form defined in (2.5). We consider the "penalized" energy of γ given by

$$E_L(\gamma) = \int_0^1 |\gamma_1'(t)|^2 + |\gamma_2'(t)|^2 + L\left|\omega(\gamma'(t))\big|_{\gamma(t)}\right|^2 dt, \tag{2.36}$$

while the g_L-length of γ is

$$L_L(\gamma) = \int_0^1 \sqrt{|\gamma_1'(t)|^2 + |\gamma_2'(t)|^2 + L\left|\omega(\gamma'(t))\big|_{\gamma(t)}\right|^2}\, dt. \tag{2.37}$$

In our computation we will find minimizing arcs as solutions of the Euler–Lagrange equations of the energy rather than of the length.

Method I. We substitute for γ a one-parameter family of compactly supported perturbations $\{\gamma^\lambda\}$, where $\gamma_i^\lambda(s) = \gamma_i(s) + \lambda f_i(s)$, $f_i \in C_0^\infty([0,1])$, $i = 1,2,3$, and compute the derivative of $E_L(\gamma^\lambda)$ with respect to λ at $\lambda = 0$ to obtain the Euler–Lagrange equations for the critical points of the penalized energy:

$$\gamma_1'' = -L\omega\gamma_2', \qquad \gamma_2'' = L\omega\gamma_1', \qquad (\omega(\gamma')|_\gamma)' = 0. \tag{2.38}$$

The right-hand side of the first two equations contains L and hence could potentially blow up as $L \to \infty$. However, note that $\omega(\gamma')|_\gamma$ equals a constant depending only on the initial data $\gamma(0)$ and $\gamma'(0)$ by the third equation. If we choose $\gamma(0) = o$ and $\gamma'(0) = (h_1, h_2, a_L/L)$, with $h_1, h_2, a_L \in \mathbb{R}$, then $\omega(\gamma'(t))|_{\gamma(t)} = a_L/L$ for all t and the (2.38) yield

$$(\gamma_1 + \mathbf{i}\,\gamma_2)'' = -\,\mathbf{i}\,a_L(\gamma_1 + \mathbf{i}\,\gamma_2)', \tag{2.39}$$

with $(\gamma_1, \gamma_2)(0) = (0,0)$ and $(\gamma_1', \gamma_2')(0) = (h_1, h_2)$. Comparing (2.23) with (2.39) indicates that solutions to (2.39) corresponding to $h_1^2 + h_2^2 \neq 0$ (which are arcs of circles) are projections of length minimizing arcs emanating from the origin with initial velocity $h_1 X_1 + h_2 X_2$, parameterized by $a_L \in \mathbb{R}$. In particular, g_L-geodesic arcs with horizontal initial velocity ($a_L = 0$) are also length minimizing arcs in $(\mathbb{H}, d)$, the horizontal segments. In general, if $a_L \neq 0$, the solution of (2.38) may be not horizontal. However, choosing a sequence $a_L \to a_\infty \in \mathbb{R}$ as $L \to \infty$, one can easily prove, using energy estimates for the ODE (2.39), that the corresponding solutions γ_L of (2.38) converge uniformly to a length minimizing arc in $(\mathbb{H}, d)$ as described in (2.23).

Method II. We can approach the problem differently, and solve (2.38) directly. Integrating (2.38) we obtain Euler-Lagrange equations for g_L-geodesics:

$$\gamma_1' - a_L\gamma_2 = b_L^1, \quad \gamma_2' + a_L\gamma_1 = b_L^2, \quad \text{and} \quad \gamma_3' - \frac{1}{2}\Big(\gamma_1\gamma_2' - \gamma_2\gamma_1'\Big) = \frac{a_L}{L}, \tag{2.40}$$

with $a_L \in \mathbb{R}$ and $b_L = (b_L^1, b_L^2) \in \mathbb{R}^2$ arbitrary. We look for special solutions $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ of (2.40) such that $\gamma(0) = o$ and $\gamma(1) = x$, a fixed point in $\mathbb{H}$. To simplify our notation it is convenient to define $m(s) = s - \sin s$.

In case $a_L = 0$, all solutions of (2.40) have the form

$$\begin{aligned} \gamma_1(s) + \mathbf{i}\,\gamma_2(s) &= sb, \\ \gamma_3(s) &= 0 \end{aligned} \tag{2.41}$$

for some $b \in \mathbb{C}$.

In case $a_L \neq 0$, all solutions of (2.40) have the form

$$\begin{aligned} \gamma_1(s) + \mathbf{i}\,\gamma_2(s) &= u_L\Big(e^{-\mathbf{i}\,a_L s} - 1\Big), \\ \gamma_3(s) &= \frac{a_L}{L}s + \frac{1}{2}|u_L|^2 m(sa_L), \end{aligned} \tag{2.42}$$

for some $u_L \in \mathbb{C}$.

Next, we impose the endpoint constraint $\gamma(1) = x = (x_1, x_2, x_3)$ to obtain the geodesic arcs:

1. If $x_3 = 0$ then there exists a unique geodesic arc, given by (2.41) with the choice $b = (x_1, x_2)$. The length is $L_L(\gamma) = |b|$.
2. If $x_3 \neq 0$ and $R^2 = x_1^2 + x_2^2 \neq 0$, then one has a finite number of geodesics all given by (2.42) with a_L a solution to the equation

$$\frac{R^2}{x_3 - \frac{a_L}{L}} = \frac{4m'(a_L)}{m(a_L)}. \tag{2.43}$$

   Setting $s = 1$ in (2.42) we see that u_L is completely determined by the choice of a_L. In particular

$$|u_L|^2 = 2\frac{x_3 - \frac{a_L}{L}}{m(a_L)}. \tag{2.44}$$

   The length is $\mathrm{Length}_L(\gamma) = |a_L|\sqrt{|u_L|^2 + 1/L}$.
3. If $x_3 \neq 0$ and $R^2 = x_1^2 + x_2^2 = 0$ there will be a geodesic arc of the form

$$\gamma_1(s) = \gamma_2(s) = 0, \text{ and } \gamma_3(s) = sx_3. \tag{2.45}$$

   The length is $\mathrm{Length}_L(\gamma) = \sqrt{L}|x_3|$.

Remark 2.9. If L or x_3 are sufficiently large there will be shorter arcs joining o to x, namely, infinitely many more geodesics given by (2.42) with $a_L = 2k\pi$, $k = \pm1, \pm2, \ldots$, $\mathrm{sign}(a_L) = \mathrm{sign}\, x_3$, $|a_L| < |x_3|L$ and for arbitrary choice of u_L. The length of these arcs is given by

$$\mathrm{Length}_L(\gamma) = 2\sqrt{k\pi(x_3 - \frac{2k\pi}{L}) + \frac{(k\pi)^2}{L}} = 2\sqrt{k\pi(x_3 - \frac{k\pi}{L})}.$$

As a result of the above computations, we have the following

Proposition 2.10. *Given $x \in \mathbb{H}$, any length minimizing horizontal curve γ joining x to the origin $o \in \mathbb{H}$ is the uniform limit as $L \to \infty$ of geodesic arcs from o to x in the Riemannian spaces $(\mathbb{R}^3, g_L)$. The limit arcs are of the form* (2.23) *and are given by* (2.41) *and* (2.42) *in the limit $L \to \infty$. Moreover, the convergence is uniform both in x and in the parameter L, in the sense that*

$$\lim_{\epsilon \to 0} \sup_{0<L\leq\infty, x\in K} d_L(o,x) - d_{L-\epsilon}(o,x) = 0$$

for any compact $K \subset \mathbb{R}^3$.

We now state the main result of this section.

Theorem 2.11. *The sequence of metric spaces $(\mathbb{R}^3, d_L)$ converges to $(\mathbb{H}, d)$ in the pointed Gromov–Hausdorff sense as $L \to \infty$.*

Theorem 2.11 is an immediate corollary of Proposition 2.8, when we choose $X = \mathbb{R}^3$, d_0 the Carnot–Carathéodory metric d on X for the usual Heisenberg structure, and d_t the distance function associated to the Riemannian metric g_L, where $L = 1/t$.

2.4.5 Riemannian approximants to $\mathbb{H}^n$ and Carnot groups

Approximating sequences of Riemannian metrics can be defined also in the higher-dimensional Heisenberg group and in general Carnot groups.

We recall from Section 2.1.2 the left invariant basis $X_1, \dots, X_{2n}, X_{2n+1}$ for the Lie algebra of $\mathbb{H}^n$, where the first $2n$ vector fields span the horizontal bundle and the final vector field generates the center. For any $L > 0$ we define Riemannian metrics g_L in $\mathbb{R}^{2n+1}$ so that the set $\{\tilde{X}_1, \dots, \tilde{X}_{2n+1}\}$ is orthonormal, where we have let $\tilde{X}_i = X_i$ for $i = 1, \dots, 2n$ and $\tilde{X}_{2n+1} = L^{-1/2} X_{2n+1}$. The norm of a tangent vector $\vec{v} = \sum_{i=1}^{2n+1} v_i X_i$ is $|\vec{v}| = \langle \vec{v}, \vec{v} \rangle^{1/2}$, where $\langle \vec{v}, \vec{v} \rangle_L = \sum_{i=1}^{2n} v_i^2 + L v_{2n+1}^2$. Thus, the only curves with finite velocity in the limit as $L \to \infty$ are the horizontal paths.

A simple extension of the arguments in the previous section yields the analogue of Theorems 2.10 and 2.11 in the $\mathbb{H}^n$, $n \geq 1$ setting.

Next we discuss the Riemannian approximation scheme for general Carnot groups. Let $\mathbb{G}$ be a Carnot group of dimension N with Lie algebra $\mathfrak{g} = V_1 \oplus \cdots \oplus V_r$, equipped with a background left invariant Riemannian metric which makes the layers V_i orthogonal. Let $\{X_1, \dots, X_m\}$ denote an orthonormal basis of the horizontal layer $H = V_1$ and let $\{Y_1, \dots, Y_n\}$ denote an orthonormal basis of the vertical layer $V = V_2 \oplus \cdots \oplus V_r$. Finally, let $d(i)$ be the index of the layer to which Y_i belongs (so that $d(i) \geq 2$). Set $\tilde{Y}_i = L^{-(d(i)-1)/2} Y_i$.

We now define a family of Riemannian metrics $(g_L)_{L>0}$ in $\mathbb{R}^N$ so that the family $\{X_1, \dots, X_m, \tilde{Y}_1, \dots, \tilde{Y}_n\}$ is orthonormal. Note that $\lim_{L\to\infty} \tilde{Y}_i = 0$ and again, the only curves with finite velocity in the limit are the horizontal ones.

The convergence of geodesic arcs in this more general (higher step) setting is quite delicate and presents an obstacle which does not appear in the step two case: the Riemannian geodesics in the approximants may converge to singular geodesics. This line of investigation goes beyond the scope of this survey and we refer the reader to the monograph [203] for more details. Nevertheless, the analog of Theorem 2.11 continues to hold in this setting. Denoting by d_L the global distance function associated to the Riemannian metric g_L, we have

Theorem 2.12. *The sequence of metric spaces* $(\mathbb{R}^N, d_L)$ *converges to* $(\mathbb{G}, d)$ *in the pointed Gromov–Hausdorff sense as* $L \to \infty$*.*

2.5 Notes

Notes for Section 2.1. Useful surveys of aspects of analysis and geometry in the Heisenberg group or on more general Carnot groups include Semmes [241] and Heinonen [135]. For calculus on Heisenberg manifolds, we recommend Beals and Greiner [30] and Gaveau [120]. One of the standard references for analysis on the Heisenberg group is Chapters XII and XIII of Stein's book on harmonic analysis [243]. Folland [100] has a detailed introduction to the Heisenberg group, its representations and applications, and among other things discusses polarized coordinates and the matrix model for $\mathbb{H}$.

Notes for Subsection 2.1.3. The notion of sub-Riemannian geometry rests on the accessibility condition for horizontal paths. By the fundamental theorem of Chow and Rashevsky, local accessibility is equivalent to the bracket generating condition for a frame of smooth vector fields $X_1, \ldots, X_m$ which generate the tangent bundle of an n-dimensional manifold M:

$$\mathrm{Rank}(\mathrm{Lie}[X_1, \ldots, X_m])(x) = n \tag{2.46}$$

for every $x \in M$. The analytic form of this condition first appeared in the literature in 1967 in the celebrated paper of Hörmander [149], who proved that it is a sufficient condition for hypoellipticity of second order differential operators of the form $\mathcal{L} = \sum_{i=1}^m X_j^2$. Operators of this form are known as *sum of squares* or *sub-Laplacians*.

Carnot groups arise naturally as ideal boundaries of noncompact rank 1 symmetric spaces. For instance, $\mathbb{H}^n$ is isomorphic to the nilpotent part of the Iwasawa decomposition of $U(1,n)$, the isometry group of the complex hyperbolic space of dimension n. The key role played by Carnot groups[10] become evident in the early 1970s with a number of important papers following a circle of ideas outlined by E.M. Stein in his address at the 1970 International Congress of Mathematicians in

[10] The name *Carnot group* is relatively recent, as it emerged in the late 1980s from the papers of Gromov, Pansu and others. In the work of Stein and his collaborators such groups are known as *nilpotent homogeneous* or *stratified Lie groups*.

Nice. The homogeneous structure of such groups allows the development of harmonic analysis [99] which, in turn, plays a central role in the regularity theory of general Hörmander type operators. In their celebrated article [102] G. Folland and E.M. Stein use the Heisenberg group as an osculating space for strictly pseudoconvex CR manifolds in $\mathbb{C}^n$ to study properties of singular integral operators and solve the $\bar{\partial}_b$ operator. In the following chapter, we describe in some detail the role of the Heisenberg group in CR geometry and its natural occurrence in connection with the Gromov hyperbolic geometry of complex hyperbolic space.

The program of using harmonic analysis in stratified Lie groups as a model for harmonic analysis in more general sub-Riemannian spaces was developed in the pathbreaking paper [233], where Rothschild and Stein extended the Folland–Stein approach to general sub-Riemannian spaces associated to smooth, bracket generating, sets of vector fields. A central result in the work of Rothschild–Stein is an approximation scheme that allows one to

- lift a set of bracket generating vector fields to a higher-dimensional space so that the *lifted* vectors are free (i.e., the only relations among the vectors and their commutators up to the step needed to span the whole tangent space, are those arising from the bracket structure and the Jacobi identity), and
- approximate the sub-Riemannian structure of the lifted vector fields with an osculating Carnot group structure, with very precise estimates on the nonlinear remainder terms.

Subanalyticity of real analytic Carnot–Carathéodory metrics on sub-Riemannian manifolds was recently established in a significant and extremely intricate analysis by Agrachev and Gauthier [6].

Notes for Section 2.2. For more details on the lifting procedure and Legendrian paths as discussed in Remark 2.3, see [57], [22] and other references therein.

The metric $d_{\mathbb{H}}$ defined by the gauge (2.11) is nowadays associated with the name of Adam Korányi, who used it extensively in connection with harmonic analysis and potential theory in the Heisenberg group and more general Carnot groups of Heisenberg type [166]. The properties of the Korányi gauge and inversion are mostly due to Korányi and Reimann [168], [169], in particular, the elegant proof of (2.15) can be found in Section 1 of [169]. See also [75].

As noted in the chapter, while it is easy to write down a variety of homogeneous gauges on general Carnot groups such as (2.21) which agree with the Korányi gauge $||\cdot||_{\mathbb{H}}$ in the Heisenberg setting, the fact that $d_{\mathbb{H}}$ defines a metric is a specific feature of that setting. No simple expressions of Korányi type for homogeneous metrics in general Carnot groups are known.

We note that in the literature one can find several different gauges of Korányi type, each designed for a particular application. See, for example, Franchi–Serapioni–Serra-Cassano [106] or Bieske [33]. See the notes to Chapter 9 for more information.

Useful references for Hausdorff measure and dimension in metric spaces include Mattila [196] and Falconer [94]. The study of sub-Riemannian spaces *qua* metric spaces, with a focus on the intrinsic metric geometry of subsets, was advocated by Gromov in [130].

Notes for Section 2.3. The description of the CC geodesics in the Heisenberg group can be found in numerous references, see for example [32], p. 28. The bubble sets were described by Pansu in his 1982 paper [217].

Notes for Section 2.4. The Riemannian approximation scheme appeared first in the paper of Korányi [167], who used it to derive explicit expressions for the CC geodesics. His method is sketched in 2.4.4. Later, the same approach was used by several authors, see for instance [10] and [21]. Roughly speaking, in this scheme lies the main idea of our general approach: to define horizontal geometric objects as limits of horizontal restrictions of classical Riemannian analogs. The situation for general Carnot groups is rather more complicated than the step two case; the proof via convergence of geodesics which we gave for Theorem 2.11 encounters obstacles in the higher step setting due to the possibility of abnormal geodesics. Theorem 2.12 was proved by Pansu in [218]. In even greater generality, Riemannian approximations to sub-Riemannian manifolds were studied by Roberto Monti in his Ph.D. dissertation at the Università di Trento (unpublished).

Gromov's notion of convergence of metric spaces was introduced in his groundbreaking paper on groups of polynomial growth [129], see also Chapter 3 of [131]. Proposition 2.7 can be found in [131, Chapter 3] as Propositions 3.7 and 3.13, respectively. A very readable account of the theory of Gromov–Hausdorff convergence of metric spaces can be found in Chapters 7 and 8 of [47].

The basic ingredients of Riemannian geometry which we use in Subsection 2.4.2 can be found in the standard texts. For the detailed definition and proof of the existence of the Levi-Civita connection, we refer the reader to [113, Theorem 2.51]. The Kozul identity can be found on p. 55 in [88].

Additional notes. Among the many advantages of the special structure of the Heisenberg groups $\mathbb{H}^n$ are the facts that the center is of dimension 1 and that the explicit solution of the sub-Laplacian operator $\mathcal{L} = \sum_{i=1}^{2n} X_i^2$ is explicitly known. In [98], Folland proved that for a specific choice of a constant C_n, the function $\Gamma(x,y) = C_n ||y^{-1}x||_{\mathbb{H}}^{2-(2n+2)}$ satisfies the equation $\mathcal{L}\Gamma(x,y) = \delta_x(y)$ in $\mathbb{H}^n$. (We will prove this in the setting of the first Heisenberg group in Section 5.2.) In 1980, Kaplan [160] introduced a new class of groups, now called H-type groups, in which a generalization of Folland's formula holds. In any step two Carnot group define the linear map $J : V_2 \to \mathrm{End}(V_1)$ from the second layer of the Lie algebra stratification to the endomorphisms of the first layer via the identity $\langle J(Y)X, X'\rangle = \langle [X,X'], Y\rangle$ for all $Y \in V_2$ and $X, X' \in V_1$. A step two Carnot group $\mathbb{G}$ is called an *H-type group* (or *group of Heisenberg type* or *Kaplan group*) if the map J is orthogonal, i.e.,

$$\langle J(Y)X, J(Y)X'\rangle = \langle X, X'\rangle\, |Y|^2.$$

H-type groups have a rich analytic and algebraic structure. The ideal boundaries of noncompact, constant negatively curved, symmetric spaces of rank 1 are one-point compactifications of H-type groups [73]. If we write $g = \exp(x(g) + y(g))$ with $x(g) \in V_1$ and $y(g) \in V_2$, then the fundamental solution for the sub-Laplacian in an H-type group $\mathbb{G}$ is given in terms of a gauge metric

$$||g||_{\mathbb{G}}^4 = |x(g)|^4 + 16|y(g)|^2,$$

and has the form

$$\Gamma_{\mathbb{G}}(g, g') = C_{\mathbb{G}}||(g')^{-1}g||_{\mathbb{G}}^{2-Q},$$

where $C_{\mathbb{G}} > 0$ is a constant depending only on the group $\mathbb{G}$ and $Q = \dim V_1 + 2\dim V_2$ is the homogenous dimension of $\mathbb{G}$. See also [54], [138], [26] for further results connected with linear and nonlinear potential theory and H-type groups. There is a rich theory of conformal geometry, including analogs of the Korányi inversion $j_{\mathbb{H}}$ on these groups; for further information, see [118].

Chapter 3

Applications of Heisenberg Geometry

A very intuitive way to think of the sub-Riemannian Heisenberg group is as a medium in which motion is only possible along a given set of directions, changing from point to point. If the constraints are too tight, then it may be impossible to join any two points with an admissible trajectory, hence one needs to find conditions on the constraints implying "horizontal accessibility".

Constrained motion as defined above is studied in depth in control theory and has numerous applications in engineering (motion of robot arms and wheeled motion) and biology (models of perceptual completion). It also arises naturally in other branches of pure mathematics. In this chapter we briefly describe occurrences of Heisenberg geometry in other areas of pure mathematics (CR geometry, Gromov hyperbolic geometry of complex hyperbolic space, and jet spaces), as well as in the engineering and neurobiological applications mentioned above.

3.1 Jet spaces

The Heisenberg group, as well as a large class of other Carnot groups, can be represented as jet spaces. The concept of jet space gives geometric structure to the classical framework of Taylor polynomials.

To define the first jet space $J^1(\mathbb{R},\mathbb{R})$ we begin by introducing an equivalence relation in $C^1(\mathbb{R})$: two functions f and g are equivalent at a point $t \in \mathbb{R}$ if $f(t) = g(t)$ and $f'(t) = g'(t)$. To stress the role of the basepoint t we will write $f \sim_t g$. We then define $J^1(\mathbb{R},\mathbb{R})$ as the disjoint union of the quotient spaces $C^1(\mathbb{R})/\sim_t$:

$$J^1(\mathbb{R},\mathbb{R}) = \coprod_{t\in\mathbb{R}} C^1(\mathbb{R},\mathbb{R})/\sim_t\ . \tag{3.1}$$

Let us denote elements of $J^1(\mathbb{R},\mathbb{R})$ by $j_t(f)$. The space $J^1(\mathbb{R},\mathbb{R})$ can be naturally identified with $\mathbb{R}^3$ via the global coordinates

$$j_t(f) \longleftrightarrow x = (x_1, x_2, x_3) = (f'(t), t, f). \tag{3.2}$$

Jet spaces are naturally equipped with sub-Riemannian structure. We illustrate this in the simplest case of the contact structure on $J^1(\mathbb{R},\mathbb{R})$. We want to define a 1-form θ which vanishes on all 1-jets $t \to j_t(f)$ of C^1 functions. The basic relation is

$$df = f'(t)\,dt,$$

which, in local coordinates, reads

$$(dx_3 - x_1\,dx_2)j_t(f) = 0.$$

The latter suggests the choice $\theta = dx_3 - x_1\,dx_2$; one immediately verifies that θ is contact: $\theta \wedge d\theta = dx_1 \wedge dx_2 \wedge dx_3$. The horizontal tangent bundle defined by θ is

$$\mathcal{H}^1_x = \{\vec{v} \in T_x J^1(\mathbb{R},\mathbb{R}) \mid \theta(\vec{v}) = 0\} = \operatorname{span}(\mathcal{X}_1, \mathcal{X}_2), \tag{3.3}$$

where $\mathcal{X}_1 = \partial_{x_1}$ and $\mathcal{X}_2 = \partial_{x_2} + x_1\partial_{x_3}$. Note that

$$TJ^1(\mathbb{R},\mathbb{R}) = \mathcal{H}^1 \oplus [\mathcal{H}^1, \mathcal{H}^1]. \tag{3.4}$$

Next, we define a group law on $J^1(\mathbb{R},\mathbb{R})$ such that $\mathcal{H}^1$ and θ are left invariant:

$$(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3 + x_1 y_2). \tag{3.5}$$

Observe that $J^1(\mathbb{R},\mathbb{R})$ with this group law is isomorphic with the Heisenberg group in its matrix model (2.1).

3.2 Applied models

The sub-Riemannian geometry of the Heisenberg group in its incarnation as the first jet space arises in a variety of physical and biological systems. Such geometries arise naturally in the study of optimal control and path planning where one is dealing with vehicles with limited degrees of freedom, such as wheeled vehicles. Less obviously (and perhaps more surprisingly), this representation is also relevant to the study of the geometry of the first layer of the mammalian visual cortex $V1$. In this section, we describe the role of contact geometries modeled by the Heisenberg group in a simple mechanical model of nonholonomic motion, and in the aforementioned model of the functional structure of the first layer of the visual cortex.

In the process of describing these examples, we will encounter a new sub-Riemannian space, the roto-translation group $\mathcal{RT}$ which, while distinct from the

Heisenberg group, is locally approximated by it. We first introduce $\mathcal{RT}$ and then discuss the approximation by $\mathbb{H}$.

The *roto-translation group*, $\mathcal{RT}$, is the group of Euclidean rotations and translations of the plane equipped with a particular sub-Riemannian metric. More precisely, $\mathcal{RT}$ is a three-dimensional topological manifold diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$ with coordinates (x, y, θ). We identify the vector fields

$$X_1 = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y},$$

$$X_2 = \frac{\partial}{\partial \theta}$$

and

$$X_3 = \sin\theta \frac{\partial}{\partial x} - \cos\theta \frac{\partial}{\partial y}.$$

Observe that, similarly to the Heisenberg group, $[X_1, X_2] = X_3$. However, this is not the only nontrivial bracket, indeed $[X_3, X_2] = -X_1$. Note also that $\omega = \sin\theta\, dx - \cos\theta\, dy$ is a contact form on $\mathcal{RT}$ whose kernel is spanned by X_1 and X_2.

We equip $\mathcal{RT}$ with a sub-Riemannian metric using the same construction as Section 2.2.3. First, we introduce an inner product, $\langle\cdot,\cdot\rangle_{\mathcal{RT}}$ on the subbundle of the tangent bundle generated by $\{X_1, X_2\}$. Then, if $\gamma : I \to \mathcal{RT}$ is a path in $\mathcal{RT}$ so that γ' is always in the span of $\{X_1, X_2\}$, we define

$$\mathrm{Length}_{\mathcal{RT}}(\gamma) = \int_I \langle\gamma'(t), \gamma'(t)\rangle_{\mathcal{RT}}^{1/2}\, dt.$$

The induced sub-Riemannian distance is

$$d_{\mathcal{RT}}(p, q) = \inf\{\mathrm{Length}_{\mathcal{RT}}(\gamma) \,:\, \gamma(0) = p, \gamma(1) = q, \gamma' \in \mathrm{span}\{X_1, X_2\}\}.$$

While $\mathcal{RT}$ is certainly distinct from $\mathbb{H}$, we emphasize that, locally, the two are equivalent. One way to see this is to invoke Darboux's theorem: any contact form on a three-dimensional manifold is locally diffeomorphic to the standard contact form ω in $\mathbb{R}^3$ (defined in (2.5)). Another way to see this is to use a weighted Taylor expansion adapted to the bracket structure to describe the vector fields locally. We follow the second approach. First, some definitions:

Definition 3.1. In $\mathcal{RT}$, let $\mathcal{L}^1$ be the set of linear combinations of $\{X_1, X_2\}$ with smooth coefficients and let $\mathcal{L}^2 = \mathcal{L}^1 + [\mathcal{L}^1, \mathcal{L}^1]$. Let $L^i(p)$ be the subspace of $T_p\mathcal{RT}$ given by the collection of vectors $X(p)$, $X \in \mathcal{L}^i$.

By the above remarks, $\mathcal{L}^2 = T(\mathcal{RT})$ and $L^2(p) = T_p\mathcal{RT}$.

Definition 3.2. If $\{x_1, x_2, x_3\}$ are local coordinates at p so that $\{dx_1, dx_2, dx_3\}$ form a basis of $T_p^*\mathcal{RT}$ adapted to the flag $L^1(p) \subset L^2(p)$, we define the *weight* of x_i at p by

$$\mathrm{weight}(x_i) = \min\{i : \partial_{x_1}(p) \in L^i(p)\}.$$

Moreover, we let

$$\text{weight}(\partial_{x_i}) = -\text{weight}(x_i).$$

This notion of weight captures the bracket structure of $\mathcal{RT}$ and formalizes the intuitive notion that X_3 is a second order derivative as it arises as a bracket of two horizontal vector fields. We note that at $(0,0,0)$, the standard coordinates (x, y, θ) have weights 1, 2 and 1 respectively. Using this weighting, we expand the vector fields, X_1, X_2 and X_3 at $(0,0,0)$:

$$X_1 = \underbrace{\partial_x + \theta\partial_y}_{\text{weight } -1} + \underbrace{\left(\frac{\theta^2}{2}\partial_x - \frac{\theta^3}{3!}\partial_y\right)}_{\text{weight } 1} + \cdots,$$

$$X_2 = \underbrace{\partial_\theta}_{\text{weight } -1}$$

and

$$X_3 = \underbrace{\partial_y}_{\text{weight } -2} - \underbrace{\left(\theta\partial_x + \frac{\theta^2}{2}\partial_y\right)}_{\text{weight } 0} + \cdots.$$

Considering only the weight -1 terms in the expansions of X_1, X_2 and the weight -2 term in the expansion of X_3, we recover the presentation of the Heisenberg group in its matrix model (2.1). It is in this sense that $\mathcal{RT}$ is locally modeled by $\mathbb{H}$ in a neighborhood of $(0,0,0)$. A similar computation, left to the reader, provides the same result in a neighborhood of any point of $\mathcal{RT}$.

We remark that the roto-translation group goes by other names in the literature, e.g., "group of planar Euclidean rigid motions".

3.2.1 Nonholonomic path planning

To illustrate the use of the roto-translation group in the context of nonholonomic path planning and optimal control, we consider the simplest example of a wheeled vehicle: the unicycle. To model the motion and control of a unicycle in a planar region, we introduce standard (x, y) coordinates in the plane, together with an angular variable θ describing the deviation of the wheel from the x-axis. See Figure 3.1.

The *state space* $\mathcal{S} = \mathbb{R}^2 \times \mathbb{S}^1$ describes all possible configurations of the unicycle. If the operator pedals the unicycle forward from a point $(x, y, \theta) \in \mathcal{S}$ without changing the angle of the wheel, the unicycle follows the parametric path

$$(x + t\cos\theta, y + t\sin\theta, \theta).$$

Taking one derivative in t yields one of the allowable directions of instantaneous motion:

$$X_1 = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}.$$

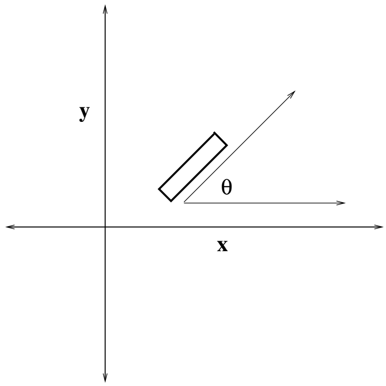


Figure 3.1: Coordinates describing the unicycle.

As the operator can change the angle of the wheel at will, another direction of instantaneous motion is simply

$$X_2 = \frac{\partial}{\partial \theta}.$$

To complete to a basis for $T\mathcal{S}$ we add the vector

$$X_3 = \sin\theta \frac{\partial}{\partial x} - \cos\theta \frac{\partial}{\partial y}.$$

Thus, we recover the roto-translation group as a model space for the motion of a unicycle.

We note that the physical model provides an interpretation of the vector field X_3 since a unicycle cannot move instantaneously in this direction as it is perpendicular to the axis of the wheel. However, using a combination of angle rotation and forward motion, the operator can access any position in the plane. This is reflected mathematically in the bracket relation $[X_1, X_2] = X_3$.

In this context, sub-Riemannian metric geometry has a close connection with path planning. To plan an optimally efficient path between two points p, q in the state space, we must minimize the length of the path connecting p to q while moving only in allowable directions. In other words, optimal path planning is equivalent to the geodesic problem with respect to the metric $d_{\mathcal{RT}}$.

3.2.2 Geometry of the visual cortex

Neuralbiological research over the past few decades has greatly clarified the functional mechanisms of the first layer (V1) of the visual cortex. Early research showed that V1 contains a variety of types of cells, including the so-called "simple cells".

They found that simple cells are sensitive to orientation specific brightness gradients and are associated to a specific retinotopic field (i.e., a region of the retina). Simple cells are arranged into columns of cells with shared orientation preference and these columns are further arranged into the so-called "hypercolumns" Each hypercolumn is a stack of columns of simple cells that are all associated to specific spatial points on the retina but with with orientation preferences, or "tuning", ranging over all possible angles. Figure 3.2 shows a schematic representation of the hypercolumn structure. In this figure, the circles represent a column of simple cells and the arrow in each circle represents the orientation tuning of that column. The "horizontal" direction is one possible direction of motion between retinotopic fields.

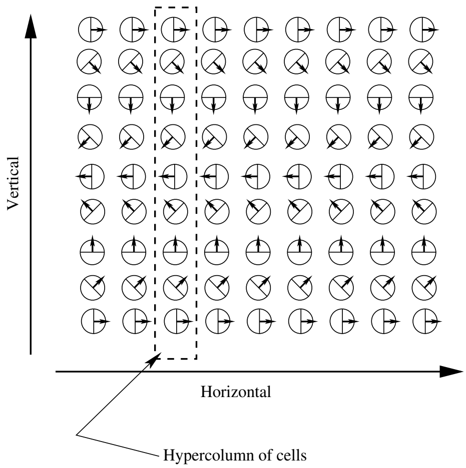


Figure 3.2: The hypercolumn structure of V1.

Early assumptions – supported by some research – that cortical connectivity runs mostly vertically along the hypercolumns and is severely restricted in horizontal directions (between hypercolumns), were contradicted by later evidence showing "long range horizontal" connectivity in the cortex. This experimental evidence demonstrated the properties of a specific geometric structure in the first layer of the visual cortex associated to intracortical communication. We may mathematically model this structure of the cortex using a sub-Riemannian structure. The retinotopic fields are modeled by two spatial directions which we will denote by $(x, y) \in \mathbb{R}^2$. Ignoring the column structure in favor of the hypercolumn structure, we may describe each hypercolumn using a copy of $\mathbb{S}^1$. We model this situation using the roto-translation group, $\mathcal{RT}$, as $\mathbb{R}^2 \times \mathbb{S}^1$ where each point (x, y, θ) represents a column of cells associated to a point of retinal data $(x, y) \in \mathbb{R}^2$, all of which are attuned to the orientation given by the angle $\theta \in \mathbb{S}^1$.

The experimental evidence referred to above shows that horizontal connections are made between points (x_1, y_1, θ_1) and (x_2, y_2, θ_2) if $\theta_1 = \theta_2$, and that there is a stronger probability of connection if the vector $(x_1, y_1) - (x_2, y_2)$ is parallel to the direction given by $\theta_1 = \theta_2$. These restrictions are described by the sub-Riemannian structure $\mathcal{RT}$ where the horizontal directions are spanned by the vector fields

$$X_1 = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y} \qquad \text{and} \qquad X_2 = \frac{\partial}{\partial \theta}.$$

3.3 CR structures

A CR 3-manifold is a real, oriented smooth manifold M of dimension 3, together with a subbundle $T_{1,0}$ of the complex tangent bundle $T^{\mathbb{C}}M$ satisfying:

(i) $\dim_{\mathbb{C}} T_{1,0} = 1$,

(ii) $T_{1,0} \cap \overline{T}_{1,0} = \{0\}$, and

(iii) $T_{1,0}$ is integrable, i.e., if Z_1, Z_2 are smooth sections of $T_{1,0}$, then so is $[Z_1, Z_2]$.

The Heisenberg group $\mathbb{H}$ carries a CR structure. To see this we let

$$Z = \frac{1}{2}\Big(X_1 - \mathbf{i}\, X_2\Big) = \frac{\partial}{\partial z} - \mathbf{i}\, \frac{\overline{z}}{4} \frac{\partial}{\partial x_3},$$

$$\overline{Z} = \frac{1}{2}\Big(X_1 + \mathbf{i}\, X_2\Big) = \frac{\partial}{\partial \overline{z}} + \mathbf{i}\, \frac{z}{4} \frac{\partial}{\partial x_3}.$$

Here we have used the standard notation $\partial/\partial\overline{z} = \frac{1}{2}(\partial/\partial x_1 + \mathbf{i}\, \partial/\partial x_2)$ and $\partial/\partial z = \frac{1}{2}(\partial/\partial x_1 - \mathbf{i}\, \partial/\partial x_2)$. Set $T_{1,0} = \mathrm{span}(Z)$. Then (i)–(iii) above are immediate. We also observe the identity $[Z, \overline{Z}] = \frac{1}{2}\, \mathbf{i}\, \partial/\partial x_3$.

In order to obtain a more geometric insight we recall the notion of an embedded CR manifold. Let $\Omega = \{z = (z_1, z_2) \in \mathbb{C}^2 : \phi(z) < 0\}$, $\phi \in C^2(\mathbb{C}^2)$, $\nabla\phi \neq 0$, define a smooth subset of $\mathbb{C}^2$. The tangent space to $\partial\Omega$ at $p \in \partial\Omega$ is given by

$$T_p\partial\Omega = \{Z \in \mathbb{C}^2 : \mathrm{Re}\langle\langle\overline{\partial}\phi(p), Z\rangle\rangle = 0\},$$

where

$$\overline{\partial}\phi = \left(\frac{\partial}{\partial \bar{z}_1}\phi, \frac{\partial}{\partial \bar{z}_2}\phi\right),$$

and for $Z, W \in \mathbb{C}^2$, $\langle\langle Z, W\rangle\rangle = Z_1\bar{W}_1 + Z_2\bar{W}_2$, denotes the complex scalar product. The maximal complex, or *horizontal*, plane at p is given by

$$H_p\partial\Omega = \{Z \in \mathbb{C}^2 : \langle\langle\overline{\partial}\phi(p), Z\rangle\rangle = 0\}.$$

Combining the conditions defining tangential complex lines $(\mathrm{Re}\langle\langle\overline{\partial}\phi(p), Z\rangle\rangle = 0)$ and horizontal complex lines $(\langle\langle\overline{\partial}\phi(p), Z\rangle\rangle = 0)$ we see that the horizontal lines tangential to $\partial\Omega$ are given by $\mathrm{Im}\langle\langle\overline{\partial}\phi(p), Z\rangle\rangle\langle\langle(\partial - \overline{\partial})\phi(p), Z\rangle\rangle = 0$. Thus, the horizontal distribution on $\partial\Omega$ is given by the tangential vector fields which are in the kernel of the form

$$\sigma = \frac{\partial\phi}{\partial\bar{z}_1}d\bar{z}_1 + \frac{\partial\phi}{\partial\bar{z}_2}d\bar{z}_2 - \frac{\partial\phi}{\partial z_1}dz_1 - \frac{\partial\phi}{\partial z_2}dz_2.$$

If Ω is strictly pseudoconvex, i.e., the *Levi form*

$$Z \mapsto L(p, Z) = \sum_{i,j=1}^{2} \frac{\partial^2\phi(p)}{\partial z_i \partial\bar{z}_j} Z_i \overline{Z}_j$$

is positive definite on $H_p\partial\Omega$ for all $p \in \partial\Omega$, then it is easy to check that $H_p\partial\Omega$ is a *contact distribution* on $\partial\Omega$. In this case a defining contact form is given by σ.

If $\phi(w) = |w|^2 - 1$ then Ω is the unit ball and $\partial\Omega = \mathbb{S}^3$. The Levi form is a constant multiple of the identity (and hence positive definite), and the horizontal distribution is given by the kernel of the contact form

$$\sigma = \overline{w}_1 dw_1 - w_1 d\overline{w}_1 + \overline{w}_2 dw_2 - w_2 d\overline{w}_2. \tag{3.6}$$

The vector fields

$$W_1 = \mathbf{i}\,\frac{(1+w_2)^2}{1+\overline{w}_2}(\overline{w}_2\partial_{w_1} - \overline{w}_1\partial_{w_2})$$

and

$$W_2 = -\,\mathbf{i}\,\frac{(1+\overline{w}_2)^2}{1+w_2}(w_2\partial_{\overline{w}_1} - w_1\partial_{\overline{w}_2})$$

are a basis for $H_p\mathbb{S}^3$. We want to show that the resulting CR structure on $\mathbb{S}^3$ can be viewed as the one-point compactification of the Heisenberg group $\mathbb{H}$, and that under this identification, $H\mathbb{S}^3$ corresponds to the horizontal distribution $H\mathbb{H}$ in $\mathbb{H}$, and σ corresponds to the contact form $\omega = dx_3 - \frac{1}{2}(x_1 dx_2 - x_2 dx_1)$ in $\mathbb{H}$. In order to write the exact correspondence between $\mathbb{S}^3$ and $\mathbb{H}$, we require a special stereographic projection π based on the Cayley transform, which we define below (see 3.11).

First, we recall the definition of the Siegel domain from Section 2.4.1:

$$D = \{(\xi_1, \xi_2) \in \mathbb{C}^2 : \mathrm{Im}\ \xi_2 - |\xi_1|^2 > 0\}. \tag{3.7}$$

A defining function for D is $\phi(\xi) = \xi_1\overline{\xi_1} + \frac{1}{2}\,\mathbf{i}\,(\xi_2 - \overline{\xi_2})$. The horizontal structure $H\partial D$ is given by the kernel of the contact form

$$\tau = -\,\mathbf{i}\,\overline{\xi}_1 d\xi_1 + \,\mathbf{i}\,\xi_1 d\overline{\xi}_1 + \frac{1}{2}(d\xi_2 + d\overline{\xi}_2), \tag{3.8}$$

while the vector fields

$$\Xi_1 = \partial_{\xi_1} + 2\,\mathbf{i}\,\overline{\xi}_1\partial_{\xi_2} \qquad \text{and} \qquad \Xi_2 = \partial_{\overline{\xi}_1} - 2\,\mathbf{i}\,\xi_1\partial_{\xi_2}$$

are a basis for $H\partial D$.

Observe that the Heisenberg group acts[1] on $\mathbb{C}^2$ from the right by holomorphic affine transformations:

$$(\xi_1, \xi_2) \cdot (z, x_3) = \Big(\xi_1 + z,\ \xi_2 + 4x_3 + \mathbf{i}\,|z|^2 + 2\,\mathbf{i}\,\xi_1 \bar{z}\Big),$$

for $(z, x_3) \in \mathbb{H}$ and $(\xi_1, \xi_2) \in \mathbb{C}^2$. This action preserves D and ∂D, since

$$\begin{aligned} \operatorname{Im}(\xi_2 + 4x_3 + \mathbf{i}\,|z|^2 + 2\,\mathbf{i}\,\xi_1\bar{z}) - |\xi_1 + z|^2 &= \operatorname{Im}\xi_2 + |z|^2 + 2\,\mathbf{i}\,\xi_1\bar{z}) - |\xi_1 + z|^2 \\ &= \operatorname{Im}\xi_2 - |\xi_1|^2. \end{aligned}$$

Since this action is simply transitive[2] on ∂D, we may identify $\mathbb{H}$ with ∂D by the correspondence

$$(z, x_3) \mapsto (0, 0) \cdot (z, x_3) = (z, 4x_3 + \mathbf{i}\,|z|^2).$$

Under this identification the CR structure of $\mathbb{H}$ defined above coincides with the CR structure induced by the Euclidean metric in $\mathbb{C}^2$, i.e.,

$$H_{(4x_3 + \mathbf{i}\,|z|^2, z)}\partial\Omega = T_{1,0}(z, t). \tag{3.9}$$

In order to prove this we observe that the holomorphic subspaces at the origin coincide as $\partial\Omega$ is tangent to the hyperplane $\xi_0 = 0$ there. Next, we remark that the CR structure on $\mathbb{H}$ is left invariant, and the action of $\mathbb{H}$ on $\partial\Omega$ is holomorphic, hence preserves the CR structure on $\partial\Omega$. In conclusion, by invariance under left translation (3.9) holds everywhere.

The Cayley transform

$$C(w_1, w_2) = \left(\frac{\mathbf{i}\,w_1}{1 + w_2}, \mathbf{i}\,\frac{1 - w_2}{1 + w_2}\right) \tag{3.10}$$

maps the unit ball $B(0, 1) \subset \mathbb{C}^2$ biholomorphically onto the Siegel domain D. With the help of the Cayley transform we can define a CR generalization of stereographic projection

$$\pi : \partial B(0, 1) \setminus \{(0, -1)\} \to \mathbb{R}^3 \tag{3.11}$$

as the composition of C and the projection $(\xi_1, \xi_2) \mapsto (\xi_1, \frac{1}{4}\operatorname{Re}(\xi_2))$. The stereographic projection π can be extended to a map from $\partial B(0, 1)$ to the one-point compactification of $\mathbb{R}^3$, and the inverse map is given by

$$w_1 = \frac{-2\,\mathbf{i}\,z}{1 + |z|^2 - 4\,\mathbf{i}\,x_3} \quad \text{and} \quad w_2 = \frac{1 - |z|^2 + 4\,\mathbf{i}\,x_3}{1 + |z|^2 - 4\,\mathbf{i}\,x_3}, \tag{3.12}$$

[1]This action can be seen as a higher-dimensional analogue of the action of $\mathbb{R}$ on the upper half-plane via translations.

[2]Simply transitive means that for every two points $P, Q \in \partial D$ there exists only one $(z, x_3) \in \mathbb{H}$ that maps P to Q.

with $x_3 = \mathrm{Re}(\xi_2)$ and $z = \xi_1$. The differential of the Cayley map transforms the frame W_1, W_2 for $H\mathbb{S}^3$ into the frame Ξ_1, Ξ_2 generating $H\partial D$, while the pull-back of the contact form τ in (3.8) is the contact form σ in (3.6) on the sphere $\mathbb{S}^3$. The expression for the Heisenberg inversion $j_{\mathbb{H}}$ (see (2.14)) takes a very simple form when transported via π to $\partial B(0,1)$, to wit,

$$\pi^{-1} \circ j_{\mathbb{H}} \circ \pi(w_1, w_2) = (-w_1, -w_2).$$

3.4 Boundary of complex hyperbolic space

The sub-Riemannian Heisenberg group arises naturally as the *boundary at infinity* of the complex hyperbolic space of dimension 2. This fact is a special instance of a more general construction of sub-Riemannian spaces as boundaries at infinity of the Gromov hyperbolic spaces obtained from smooth strictly pseudoconvex domains in $\mathbb{C}^n$ equipped with the Bergman or the Kobayashi metric.

3.4.1 Gromov hyperbolic spaces

Let (X, d) be a metric space. For $x, y, p \in X$ define the Gromov product $(x|y)_p = \frac{1}{2}(d(x,p) + d(y,p) - d(x,y))$. The space (X, d) is δ-*Gromov hyperbolic*, $\delta \geq 0$, relative to the fixed basepoint $p \in X$, if

$$(x|z)_p \geq \min\{(x|y)_p, (y|z)_p\} - \delta \tag{3.13}$$

for all $x, y, z \in X$. If a different basepoint is chosen, the space continues to be hyperbolic, with δ replaced by 2δ.

If (X, d) is a length space, a more familiar (and equivalent) definition can be given in terms of geodesic triangles. For all $x, y, z \in X$ consider the geodesic triangle with vertices x, y, z and denote by $[x,y], [x,z], [y,z]$ its sides. Then (X, d) is δ-hyperbolic if

$$d(u, [x,y] \cup [x,z]) \leq \delta$$

for all $u \in [y,z]$. Examples of Gromov hyperbolic spaces include Cartan–Hadamard manifolds, metric trees, and Cayley graphs of free groups equipped with the word metric associated to a system of generators. An uninteresting class of examples are bounded metric spaces, which are trivially δ-hyperbolic with δ equal to the diameter of the space.

3.4.2 Gromov boundary and visual metric

A *ray* is an isometric image of $[0, \infty)$ in X. Let $\mathcal{R}$ be the set of rays from a fixed basepoint p. We define an equivalence relation in $\mathcal{R}$ as follows: $\xi \approx \eta$ if $\mathrm{Haus}_X(\xi, \eta) < \infty$. The *boundary at infinity*[3] $\partial_\infty X$ of a Gromov hyperbolic space

[3] Also referred to as the *Gromov boundary* or *horizon*.

X is the set $\mathcal{R}/\approx$. We say that $r \in \mathcal{R}$ *converges at infinity* to $[r] \in \partial_\infty X$ and extend the Gromov product to $\partial_\infty X$ by letting $([\xi]|[\eta])_p = \sup \liminf (x_i|y_j)$ where the supremum is taken over all representatives ξ, η of $[\xi], [\eta]$ respectively, and $(x_i), (y_j)$ are sequences of points on the rays ξ, η respectively which tend to infinity (in the sense that $\lim_{i,j\to\infty}(x_i|y_j) = \infty$). For the sake of clarity we will consistently abuse notation by writing $(\xi|\eta)_p$ for $([\xi]|[\eta])_p$. It is easy to see that (3.13) continues to hold for this extended Gromov product, with δ replaced by 2δ.

The boundary at infinity inherits a family of metrics. First, define for all $\epsilon > 0$ the distance function

$$\rho_\epsilon(\xi, \eta) = \exp(-\epsilon(\xi|\eta)).$$

This function is not necessarily a metric as it only satisfies a weak version of the triangle inequality:

$$\rho_\epsilon(\xi, \eta) \leq e^{2\delta\epsilon} \max\{\rho_\epsilon(\xi, \zeta), \rho_\epsilon(\zeta, \eta)\}.$$

Such a distance function is typically known as a *quasimetric*. There is a standard method to construct a metric from a quasimetric which we now recall. A *chain* between two points $\xi, \eta \in \partial_\infty X$ is a finite sequence of points $z_1, \ldots, z_N$ in $\partial_\infty X$ starting at ξ and ending at η. The *length* of such a chain is $\sum_i \rho_\epsilon(z_i, z_{i-1})$. We define $d_\epsilon(\xi, \eta)$ to be the infimum of the lengths of all chains joining ξ and η. One can show that, for sufficiently small ϵ, the resulting function d_ϵ is bi-Lipschitz equivalent with a metric on $\partial_\infty X$. Any metric on $\partial_\infty X$ which arises in this fashion is called a *visual metric*.

We will also need a "parabolic" analog of the visual metric defined above. We define

$$(\xi, \eta)_{\chi,q} := \lim_{p\to\chi} \Big((\xi|\eta)_p - d(p, q) \Big) \tag{3.14}$$

for $\chi \in \partial_\infty X$, $\xi, \eta \in \partial_\infty X \setminus \{\chi\}$, and $q \in X$. A corresponding quasimetric $\rho_{\chi,q}(\xi, \eta) = \exp\Big((\xi, \eta)_{\chi,q}\Big)$ can be defined. If ρ_1 happens to be a distance function itself, then $\rho_{\chi,q}$ will be a distance function as well.

Example 3.3. As a simple example, we compute the visual metric on the boundary at infinity of the real hyperbolic space of dimension 2, $H^2_{\mathbb{R}}$, in the Poincaré disc model $(\mathcal{D}, d_h)$, where d_h is the Poincaré distance

$$d_h(z, w) = \log \left(\frac{1 + \left|\frac{z-w}{1-\overline{z}w}\right|}{1 - \left|\frac{z-w}{1-\overline{z}w}\right|} \right). \tag{3.15}$$

Clearly, $\partial_\infty H^2_{\mathbb{R}} = \mathbb{S}^1$. This follows from the fact that geodesic rays from the origin are segments joining the origin to points in $\mathbb{S}^1$. We use the elementary identity

$$\frac{1 - \left|\frac{z-w}{1-\overline{z}w}\right|}{(1-|w|)(1-|z|)} = \frac{1}{|1-\overline{z}w|^2} \left(\frac{(1+|w|)(1+|z|)}{1 + \left|\frac{z-w}{1-\overline{z}w}\right|} \right),$$

to compute the Gromov product in $H^2_{\mathbb{R}}$ as follows:

$$
\begin{aligned}
-(z|w)_0 &= -\frac{1}{2}\log\left(\frac{\frac{1+|z|}{1-|z|}\cdot\frac{1+|w|}{1-|w|}}{\frac{1+\left|\frac{z-w}{1-\overline{z}w}\right|}{1-\left|\frac{z-w}{1-\overline{z}w}\right|}}\right) \\
&= -\frac{1}{2}\log\left(\frac{(1+|w|)(1+|z|)}{1+\left|\frac{z-w}{1-\overline{z}w}\right|}\,\frac{1-\left|\frac{z-w}{1-\overline{z}w}\right|}{(1-|w|)(1-|z|)}\right) \qquad (3.16)\\
&= \log\left(\frac{1+\left|\frac{z-w}{1-\overline{z}w}\right|}{(1+|w|)(1+|z|)}\cdot|1-\overline{z}w|\right).
\end{aligned}
$$

Observing that $|1-\overline{z}w|\to|\xi-\eta|$ as $z\to\xi\in\mathbb{S}^1$ and $w\to\eta\in\mathbb{S}^1$, we immediately deduce an explicit expression for the metric on the boundary at infinity of the hyperbolic disc:[4]

$$\rho_1(\xi,\eta)=\lim_{z\to\xi\in\mathbb{S}^1,w\to\eta\in\mathbb{S}^1} e^{-(z,w)_0}=\frac{1}{2}|\xi-\eta|.$$

A similar computation shows that the boundary at infinity of the real hyperbolic space $H^n_{\mathbb{R}}$ is the round sphere $\mathbb{S}^{n-1}$, and that the metric on the boundary is precisely a multiple of the Euclidean metric on $\mathbb{S}^{n-1}$.

3.4.3 Complex hyperbolic space $H^2_{\mathbb{C}}$ and its boundary at infinity

The two-dimensional complex hyperbolic space $H^2_{\mathbb{C}}$ can be considered through a variety of models. We present a few such models, describe the identification of the Gromov boundary with a one-point compactification of the Heisenberg group, and relate the associated visual metric with the Korányi gauge.

$\boxed{\text{I}}$ **The projective model** Consider the $(2,1)$ form

$$Q(x,y)=\overline{x_1}y_1+\overline{x_2}y_2-\overline{x_3}y_3$$

on $\mathbb{C}^3$. Define the equivalence relation $x\approx y$ if and only if $x=\lambda y$ for some $\lambda\in\mathbb{C}\setminus\{0\}$. *Complex projective space* is the quotient $\mathbb{C}P^2=\mathbb{C}^3/\approx$, and *complex hyperbolic space* is the subset

$$H^2_{\mathbb{C}}=\{[x]\in\mathbb{C}P^2 : Q(x,x)<0\}.$$

[4]Note that in this case $\rho_1=\exp(-(\cdot|\cdot)_0)$ is already a metric, so the additional step described above is not necessary. This holds also for the complex hyperbolic spaces – see Section 3.4.5 – and is a general feature of the class of Gromov hyperbolic spaces satisfying the synthetic curvature condition CAT(κ) for some $\kappa<0$. See [41] and [43, p. 435] for more details.

We equip $H^2_{\mathbb{C}}$ with the metric

$$d([x],[y]) = \operatorname{arc\,cosh}\left(\frac{Q(x,y)Q(y,x)}{Q(x,x)Q(y,y)}\right)^{1/2}, \tag{3.17}$$

where $[x],[y]$ are equivalence classes in $H^2_{\mathbb{C}}$.

The metric space $(H^2_{\mathbb{C}}, d)$ is Gromov hyperbolic and its boundary at infinity coincides with the set $\{[x] \in \mathbb{C}P^2 \mid Q(x,x) = 0\}$. Topologically, this is the three-dimensional sphere $\mathbb{S}^3$.

$\boxed{\text{II}}$ **The parabolic model** Instead of constructing $H^2_{\mathbb{C}}$ from the form Q as above, we now consider

$$Q'(x,y) = -\overline{x_1}y_3 + \overline{x_2}y_2 - x_1\overline{y_3}.$$

This form is obtained from Q by a linear change of coordinates and as such induces the same hyperbolic geometry on $\{x \in \mathbb{C}P^2 : Q'(x,x) < 0\}$, equipped with the metric analog of (3.17) with Q' substituting for Q. In nonhomogeneous coordinates[5] Q' reads $Q'(x,x) = |x_2|^2 - 2\operatorname{Re}(x_1)$, whence the parabolic model of $\mathbb{C}P^2$ is given by the domain

$$\left\{(x_1,x_2) \in \mathbb{C}^2 : \operatorname{Re}(x_1) > \frac{1}{2}|x_2|^2\right\},$$

which is clearly (biholomorphically) equivalent to the Siegel domain (3.7).

$\boxed{\text{III}}$ **The ball model** Consider the unit ball $B(0,1) \subset \mathbb{C}^2$ equipped with the distance function

$$d(x,y) = \operatorname{arc\,cosh}\left(\frac{1 - \langle x,y\rangle}{\sqrt{(1-|x|^2)(1-|y|^2)}}\right), \tag{3.18}$$

where $\langle x,y\rangle = \overline{x_1}y_1 + \overline{x_2}y_2$. The map $(x_1,x_2) \mapsto [(x_1,x_2,1)]$ from $(B(0,1),d)$ to $H^2_{\mathbb{C}}$ is an isometry which extends to an isometry of $\overline{B}(0,1)$ onto $H^2_{\mathbb{C}} \cup \partial_\infty H^2_{\mathbb{C}}$.

3.4.4 The Bergman metric

The ball model for $H^2_{\mathbb{C}}$ also arises when considering the Bergman metric in the unit ball of $\mathbb{C}^2$. The *Bergman space* associated to a domain $\Omega \subset \mathbb{C}^2$ is

$$A^2(\Omega) = \{f \text{ holomorphic in } \Omega : ||f||_{L^2(\Omega)} < \infty\}.$$

The mean value property guarantees that $\sup_K |f| \le C_K ||f||_{L^2(\Omega)}$ for all $K \subset\subset \Omega$, where $C_K > 0$ depends only on K. This estimate implies the continuity of the evaluation functionals $\Phi_z : A^2(\Omega) \to \mathbb{C}$, $z \in \Omega$, defined by $\Phi_z(u) = u(z)$. In view

[5]We remind the reader that the nonhomogeneous coordinates of $[(x_1,x_2,1)] \in \mathbb{C}P^2$ are (x_1,x_2).

of the Riesz Representation Theorem, there exists a function $K(z,\xi)$ (the *Bergman kernel*[6]) so that $f(z) = \int_\Omega K(z,\xi) f(\xi)\, d\,\mathrm{Vol}(\xi)$ for all $f \in A^2(\Omega)$. The *Bergman metric* (b_{ij}) is the Riemannian metric on Ω given by the quadratic form

$$b_{ij}(z) = \partial_{z_i}\partial_{\overline{z_j}} \log K(z,z) \tag{3.19}$$

for $z \in \Omega$. An important feature of this metric is that it turns biholomorphisms into isometries: if $f : \Omega \to \Omega'$ is biholomorphic, then

$$\sum_{ij} b_{ij}(z) v_i v_j = \sum_{ij} b_{ij}(f(z))(f_* v)_i (f_* v)_j$$

for all $v \in \mathbb{C}^2$ and $z \in \Omega$.

The Bergman kernel $K(z,\xi) = \frac{2}{\pi^2}(1-\langle z,\xi\rangle)^{-3}$ and Bergman metric $b_{ij}(z) = 3(1-|z|^2)^{-2}\left((1-|z|^2)\delta_{ij} + \overline{z_i} z_j\right)$ for $\Omega = B(0,1)$ can be explicitly computed by using the symmetries of the ball. The corresponding distance function[7] is given by

$$d_B(z,w) = \frac{\sqrt{3}}{2}\log\left(\frac{|1-\langle z,w\rangle| + \sqrt{|w-z|^2 + |\langle z,w\rangle|^2 - |z|^2|w|^2}}{|1-\langle z,w\rangle| - \sqrt{|w-z|^2 + |\langle z,w\rangle|^2 - |z|^2|w|^2}}\right) \tag{3.20}$$

for $z, w \in B(0,1)$. The following computation shows that (3.20) and (3.18) agree modulo a multiplicative constant:

$$\begin{aligned} d(z,w) &= \log(\cosh d(z,w) + \sinh d(z,w)) \\ &= \log\left(\frac{|1-\langle z,w\rangle| + \sqrt{|1-\langle z,w\rangle|^2 - (1-|z|^2)(1-|w|^2)}}{\sqrt{(1-|z|^2)(1-|w|^2)}}\right) \\ &= \frac{1}{\sqrt{3}} d_B(z,w). \end{aligned}$$

Using the inverse of the Cayley transform

$$C^{-1}(\zeta_1,\zeta_2) = \left(\frac{-2\,\mathbf{i}\,\zeta_1}{1-\mathbf{i}\,\zeta_2}, \frac{1+\mathbf{i}\,\zeta_2}{1-\mathbf{i}\,\zeta_2}\right),$$

and the invariance of the Bergman distance with respect to biholomorphisms we can present the Bergman distance d_B^D explicitly in the Siegel domain D.

[6]Also known as the *reproducing kernel*.

[7]The distance function corresponding to the metric (b_{ij}) is obtained by first defining the b-length of a curve $\gamma : [0,1] \to \Omega$ as $L_b(\gamma) = \int_0^1 (b_{ij}\gamma_i'\gamma_j')^{1/2}\, ds$ and then setting $d_b(z,w) = \inf\{L_b(\gamma) | \gamma \text{ a curve joining } z \text{ to } w\}$.

For any $\zeta, \xi \in D$,

$$
\begin{aligned}
d_B^D(\zeta,\xi) &= d_B\left(\left[\frac{-2\,\mathbf{i}\,\zeta_1}{1-\,\mathbf{i}\,\zeta_2}, \frac{1+\,\mathbf{i}\,\zeta_2}{1-\,\mathbf{i}\,\zeta_2}\right], \left[\frac{-2\,\mathbf{i}\,\xi_1}{1-\,\mathbf{i}\,\xi_2}, \frac{1+\,\mathbf{i}\,\xi_2}{1-\,\mathbf{i}\,\xi_2}\right]\right) \\
&= \sqrt{3}\log\left(\frac{2\left|\frac{\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-2\overline{\zeta_1}\xi_1}{1+\mathbf{i}\,(\overline{\zeta_2}-\xi_2)+\overline{\zeta_2}\xi_2}\right| + \sqrt{\frac{4\left|\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-2\overline{\zeta_1}\xi_1\right|^2+16(\mathrm{Im}(\zeta_2)-|\zeta_1|^2)(\mathrm{Im}(\xi_2)-|\xi_1|^2)}{|\,\mathbf{i}+\zeta_2|^2|\,\mathbf{i}+\xi_2|^2}}}{4\sqrt{\frac{(\mathrm{Im}(\zeta_2)-|\zeta_1|^2)(\mathrm{Im}(\xi_2)-|\xi_1|^2)}{|\,\mathbf{i}+\zeta_2|^2|\,\mathbf{i}+\xi_2|^2}}}\right) \\
&= \frac{\sqrt{3}}{2}\log\left(\frac{|\,\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-\overline{\zeta_1}\xi_1| + \sqrt{|\,\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-2\overline{\zeta_1}\xi_1|^2+4h(\zeta)h(\xi)}}{|\,\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-\overline{\zeta_1}\xi_1| - \sqrt{|\,\mathbf{i}\,(\overline{\zeta_2}-\xi_2)-2\overline{\zeta_1}\xi_1|^2+4h(\zeta)h(\xi)}}\right), \quad (3.21)
\end{aligned}
$$

where we have let $h(\zeta) = \mathrm{Im}(\zeta_2) - |\zeta_1|^2$ (the *height* function in the Siegel domain).

3.4.5 Boundary at infinity of $H^2_{\mathbb{C}}$ and the Heisenberg group

In this section we compute the distance function

$$\rho_1(\xi,\eta) = \lim_{z\to\xi, w\to\eta} \exp(-(z|w)_0)$$

on the boundary at infinity $\partial_\infty H^2_{\mathbb{C}}$ of $H^2_{\mathbb{C}}$, using the ball model. Note that while the Bergman and hyperbolic metrics are the same modulo a multiplicative factor, the corresponding visual metrics on $\partial_\infty H^2_{\mathbb{C}}$ are only Hölder equivalent, exactly because of the multiplicative factor.

Our starting point is an evaluation of the Gromov product:

$$
\begin{aligned}
\exp(-(z|w)_0) &= \exp\left(\frac{1}{2}(d(z,w)-d(z,0)-d(w,0))\right) \\
&= \left(\frac{(|1-\langle z,w\rangle| + \sqrt{|1-\langle z,w\rangle|^2-(1-|z|^2)(1-|w|^2)})(1-|z|)(1-|w|)}{(|1-\langle z,w\rangle| - \sqrt{|1-\langle z,w\rangle|^2-(1-|z|^2)(1-|w|^2)})(1+|z|)(1+|w|)}\right)^{1/4} \\
&= \left(\frac{|1-\langle z,w\rangle| + \sqrt{|1-\langle z,w\rangle|^2-(1-|z|^2)(1-|w|^2)}}{(1+|z|)(1+|w|)}\right)^{1/2}.
\end{aligned}
$$

For $\xi, \eta \in \partial H^2_{\mathbb{C}}$ we obtain

$$
\begin{aligned}
\rho_1(\xi,\eta) &= \lim_{z\to\xi, w\to\eta} \exp(-(z|w)_0) \\
&= \sqrt{\frac{|1-\langle \xi,\eta\rangle|}{2}} \qquad (3.22) \\
&= \frac{1}{2}\sqrt{\left||\xi-\eta|^2 - 2\,\mathbf{i}\,\mathrm{Im}\langle \xi,\eta\rangle\right|}.
\end{aligned}
$$

Next we want to use the stereographic projection (3.11) to compare ρ_1 to the Korányi metric $d_{\mathbb{H}}$ in $\mathbb{H}$. Of course, ρ_1 is bounded since $\mathbb{S}^3$ is compact, hence it cannot be globally comparable to $d_{\mathbb{H}}$. However, we will show that it is locally equivalent to the Korányi metric modulo a fixed positive multiplicative factor. In view of left invariance it suffices to compare distances from a fixed point, say, the group identity $o = (0,0,0) \in \mathbb{H}$.

Since $(0,1) \in \partial B(0,1)$ is mapped through (3.11) to o, we substitute $\eta = (0,1)$ in (3.22) to obtain

$$\rho_1(x,(0,1)) = \sqrt{\frac{|1-\xi_2|}{2}}. \tag{3.23}$$

From (3.12) we observe that

$$\xi_2 = \frac{1-|z|^2 + 4\,\mathbf{i}\,x_3}{1+|z|^2 - 4\,\mathbf{i}\,x_3},$$

where $\xi \in \partial B(0,1) \subset \mathbb{C}^2$ is identified with $(z,x_3) \in \mathbb{H}$ via the stereographic projection. Substituting this expression for ξ_2 in (3.23) we obtain

$$\rho_1(\xi,(0,1)) = \frac{||x||_{\mathbb{H}}}{((1+|z|^2)^2 + 16x_3^2)^{1/4}}. \tag{3.24}$$

Note that in a neighborhood of the origin this expression is comparable with the Korányi gauge.[8] The metric ρ_1 is the analog of the standard *chordal metric* $q(x,y) = |x-y|/\sqrt{(1+|x|^2)(1+|y|^2)}$ in $\mathbb{R}^n$.

Next, we show that if the basepoint p in the evaluation of the Gromov product is moved towards a point χ in the boundary at infinity, then the corresponding rescaled metrics $\rho_{\chi,q}$ (as defined in (3.14)) will converge exactly to the Korányi gauge. As this is more easily seen in the Siegel parabolic model, we use the biholomorphic invariance to convert (3.22) to

$$\rho_1((z, 4x_3 + \mathbf{i}\,|z|^2),(0,0)) = \frac{||x||_{\mathbb{H}}}{((1+|z|^2)^2 + 16x_3^2)^{1/4}}. \tag{3.25}$$

Set $p = (0, \mathbf{i}\,s^2), q = (0, \mathbf{i}\,), \chi = (0,\infty)$, $\xi = (0,0)$, and $\eta = (z, 4x_3 + \mathbf{i}\,|z|^2)$ in the definition (3.14).[9] In view of (3.21) we have

$$\exp(d_B^D((p,q)) = \sqrt{\frac{s^2+1+\sqrt{s^4+s^2+1}}{s^2+1-\sqrt{s^4+s^2+1}}} \approx 2s \tag{3.26}$$

for large s. On the other hand, observe that the Gromov product is invariant under the action of isometries of $H^2_{\mathbb{C}}$. Using (3.21) it is immediate to check that the

[8] Compare also the discussion in Section 9.1, specifically, the extremal function in Theorem 9.1.
[9] Note that when viewed in the ball model, the point at infinity χ corresponds to the south pole $(0,-1) \in \partial B(0,1)$.

dilation δ_s is an isometry of the Siegel domain D equipped with the corresponding Bergman (or hyperbolic) metric. Using this fact and (3.25), we compute

$$\begin{aligned}
\exp(-(\xi|\eta)_p) &= \exp\big(-(\delta_{s^{-1}}(\xi)|\delta_{s^{-1}}(\eta))_{\delta_{s^{-1}}(p)}\big) \\
&= \exp\Big(-\Big(\big(\frac{z}{s}, \frac{4x_3}{s^2} + \mathbf{i}\,\frac{|z|^2}{s^2}\big) \mid (0,0)\Big)_{(0,\mathbf{i})}\Big) \\
&= \frac{(|z|^4 + 16x_3^2)^{1/4}}{\big((s^2+|z|^2)^2 + 16x_3^2\big)^{1/4}} \\
&\approx \frac{(|z|^4 + 16x_3^2)^{1/4}}{s}
\end{aligned} \tag{3.27}$$

for s much larger than $|z|$ and $|x_3|$. Combining (3.26) and (3.27) we obtain

$$\rho_{(0,\infty),(0,\mathbf{i})}\big((z, 4x_3 + \mathbf{i}\,|z|^2), (0,0)\big) = \frac{1}{2}(|z|^4 + 16x_3^2)^{1/4}, \tag{3.28}$$

where the rescaled distance $\rho_{(0,\infty),(0,\mathbf{i})}$ is defined as in (3.14).

3.5 Further results: geodesics in the roto-translation space

Geodesics in the roto-translation space $\mathcal{RT}$ can be computed explicitly in terms of Jacobi elliptic functions and elliptic integrals of the second kind. The use of elliptic functions in the study of geodesics in sub-Riemannian structures associated with groups of motions of the classical geometries and Euler's elasticae is well known, see, for example, Jurdjevic [158], [159]. We give here a detailed derivation of these explicit expressions in the roto-translation space.

The roto-translation space is equipped with a group law

$$(x, y, \theta)(x', y', \theta') = (x + x'\cos\theta - y'\sin\theta, y + x'\sin\theta + y'\cos\theta, \theta + \theta'),$$

with respect to which the above vector fields X_1 and X_2 are left invariant (see [68] for a derivation of this group law). We remark that an explicit description of the local approximation of the roto-translation group by the Heisenberg group can be seen in Section 5.5 of [32].

The following theorem provides an explicit description of the normal geodesics in $(\mathcal{RT}, d_{CC})$ starting from the origin $o = (0,0,0)$.

Theorem 3.4. *Let $a, b, c \in \mathbb{R}$ with $|c| \geq |b|$ and $(a,b) \neq (0,0)$. Let*

$$r = \sqrt{a^2+b^2}, \quad R = \sqrt{a^2+c^2}, \quad \text{and} \quad m = r^2/R^2.$$

The curve $(x(t), y(t), \theta(t))$ *given by*

$$\begin{aligned}
x(t) &= \frac{a}{m}\left(t - \frac{E(Rt+u_0|m) - E(u_0|m)}{R}\right) - \frac{b}{m}\left(\frac{\mathrm{dn}(Rt+u_0|m) - \mathrm{dn}(u_0|m)}{R}\right),\\
y(t) &= \frac{b}{m}\left(t - \frac{E(Rt+u_0|m) - E(u_0|m)}{R}\right) + \frac{a}{m}\left(\frac{\mathrm{dn}(Rt+u_0|m) - \mathrm{dn}(u_0|m)}{R}\right),\\
\theta(t) &= \mathrm{am}(Rt+u_0|m) - \arctan(a/b),
\end{aligned}$$

is a geodesic in $\mathcal{RT}$ *starting from the origin* $(0,0,0)$ *with initial velocity* (a,b,c) *and constant speed* $\sqrt{a^2+b^2+c^2}$. *Here*

$$u_0 = F(\arctan(a/b)|m),$$

$F(u|m)$ *and* $E(u|m)$ *are the (incomplete) elliptic integrals of the first and second kind, respectively, while* $\mathrm{am}(u|m)$ *is the amplitude of the Jacobi elliptic functions* $\mathrm{sn}(u|m)$, $\mathrm{cn}(u|m)$ *and* $\mathrm{dn}(u|m)$.

If $(a,b) = (0,0)$ *the corresponding geodesic takes the form* $(x(t), y(t), \theta(t)) = (0,0,ct)$.

Let m be a real parameter with $0 \le m \le 1$. The elliptic integrals of the first and second kind [1, Chapters 16 and 17] are

$$u = F(\varphi|m) = \int_0^{\varphi} \frac{d\theta}{\sqrt{1 - m\sin^2\theta}} \tag{3.29}$$

and

$$E(u|m) = \int_0^{\varphi} \sqrt{1 - m\sin^2\theta}\, d\theta \tag{3.30}$$

respectively. The Jacobi elliptic functions $\mathrm{sn}(u|m) = \sin\varphi$, $\mathrm{cn}(u|m) = \cos\varphi$ and $\mathrm{dn}(u|m) = \Delta(\varphi) := \sqrt{1 - m\sin^2\varphi}$ are defined in terms of, where $\varphi = \mathrm{am}(u|m)$ denotes the *amplitude*. It is traditional to drop the parameter m in this notation, writing $\mathrm{sn}\, u = \mathrm{sn}(u|m)$, etc.

The Jacobi elliptic functions interpolate between standard elliptic and hyperbolic trigonometry, as is made clear in the following table:

	$m = 0$	$m = 1$
$\mathrm{sn}\, u$	$\sin u$	$\tanh u$
$\mathrm{cn}\, u$	$\cos u$	$\mathrm{sech}\, u$
$\mathrm{dn}\, u$	1	$\mathrm{sech}\, u$
$\mathrm{am}\, u$	u	$\mathrm{gd}(u)$
$E(u\|m)$	u	$\tanh u$
$F(\varphi\|m)$	u	$\mathrm{gd}^{-1}(\varphi)$

Here $\mathrm{gd}(u) = 2\arctan(e^u) - \frac{\pi}{2}$ denotes the classical Gudermannian function, and $\mathrm{gd}^{-1}(\varphi) = \log\tan(\frac{\pi}{4} + \frac{\varphi}{2}) = \log(\sec\varphi + \tan z = \varphi)$ denotes its inverse.

Proof of Theorem 3.4. Geodesics in $\mathcal{RT}$ may be computed as solutions to the Hamilton–Jacobi system using the Pontrjagin Maximal Principle [203]. We denote points in the cotangent bundle of $\mathcal{RT}$ by (p, v), where $p = (x, y, \theta)$ and $v = (\xi, \eta, \tau)$, and let

$$\mathbf{a} = \mathbf{a}(p, v) = \langle X_1(p), v\rangle^2 + \langle X_2(p), v\rangle^2 = (\cos\theta \cdot \xi + \sin\theta \cdot \eta)^2 + \tau^2$$

be the Hamiltonian. The Hamiltonian equations $\dot{p} = \frac{1}{2}\nabla_v \mathbf{a}$, $\dot{v} = -\frac{1}{2}\nabla_p \mathbf{a}$ read

$$\begin{aligned} \dot{x} &= \cos^2\theta \cdot \xi + \sin\theta\cos\theta \cdot \eta, & \dot{\xi} &= 0,\\ \dot{y} &= \sin\theta\cos\theta \cdot \xi + \sin^2\theta \cdot \eta, & \dot{\eta} &= 0,\\ \dot{\theta} &= \tau, & \dot{\tau} &= (\sin\theta \cdot \xi - \cos\theta\eta)(\cos\theta \cdot \xi + \sin\theta\eta). \end{aligned}$$

We wish to solve this system with initial conditions $(x(0), y(0), \theta(0)) = (0, 0, 0)$ and $(\xi(0), \eta(0), \tau(0)) = (a, b, c)$. Clearly $\xi(t) = a$ and $\eta(t) = b$, and we reduce to the system

$$\begin{aligned} \dot{x} &= a\cos^2\theta + b\sin\theta\cos\theta = \cos\theta(a\cos\theta + b\sin\theta),\\ \dot{y} &= a\sin\theta\cos\theta + b\sin^2\theta = \sin\theta(a\cos\theta + b\sin\theta),\\ \dot{\theta} &= \tau,\\ \dot{\tau} &= (a\sin\theta - b\cos\theta)(a\cos\theta + b\sin\theta). \end{aligned}$$

Note that $a\dot{x} + b\dot{y} = \dot{x}^2 + \dot{y}^2 = (a\cos\theta + b\sin\theta)^2$ and $\ddot{\theta} = a\dot{y} - b\dot{x}$, whence $\dot{\theta} = ay - bx + c$.

As the case $(a, b) = (0, 0)$ is trivial we assume $(a, b) \neq (0, 0)$. Multiplying the second order autonomous ODE

$$\ddot{\theta} = (a\sin\theta - b\cos\theta)(a\cos\theta + b\sin\theta)$$

by $\dot{\theta}$ and integrating yields

$$\begin{aligned} \frac{1}{2}(\dot{\theta}(t)^2 - c^2) &= \int_0^{\theta(t)} (a\sin\theta - b\cos\theta)(a\cos\theta + b\sin\theta)\, d\theta\\ &= \frac{1}{2}\left(a^2 - (a\cos\theta(t) + b\sin\theta(t))^2\right) \end{aligned}$$

so

$$\dot{\theta} = \sqrt{R^2 - r^2\cos^2(\theta - \varphi)},$$

where we have introduced the notation $r^2 = a^2 + b^2$, $\varphi = \arctan(b/a)$ and $R^2 = a^2 + c^2$. Thus

$$t = \int_0^{\theta(t)} \frac{d\theta}{\sqrt{R^2 - r^2\cos^2(\theta - \varphi)}},$$

i.e.,

$$Rt = \int_{\frac{\pi}{2}-\varphi}^{\frac{\pi}{2}-\varphi+\theta(t)} \frac{d\theta}{\sqrt{1 - m\sin^2\theta}},$$

where $m = r^2/R^2$. From (3.29) we see that

$$F\left(\frac{\pi}{2} - \varphi + \theta(t)|m\right) = Rt + F\left(\frac{\pi}{2} - \varphi|m\right) \tag{3.31}$$

or equivalently

$$\theta(t) = \varphi - \frac{\pi}{2} + \mathrm{am}\left(Rt + F\left(\frac{\pi}{2} - \varphi|m\right)\right)$$

as asserted in the theorem.

Next, we compute $x(t)$ and $y(t)$. Denoting

$$u_0 = F\left(\frac{\pi}{2} - \varphi|m\right) = F(\arctan(a/b)|m) \quad \text{and} \quad u_t = F\left(\frac{\pi}{2} - \varphi + \theta(t)|m\right)$$

we rewrite (3.31) in the form

$$u_t = Rt + u_0.$$

From the definitions of sn and cn we have $\operatorname{sn} u_t = \cos(\varphi - \theta(t))$ and $\operatorname{cn} u_t = \sin(\varphi - \theta(t))$, while $\operatorname{sn} u_0 = \cos\varphi = \frac{a}{r}$ and $\operatorname{cn} u_0 = \sin\varphi = \frac{b}{r}$. Then

$$\cos\theta(t) = \frac{a \operatorname{sn} u_t + b \operatorname{cn} u_t}{r} \quad \text{and} \quad \sin\theta(t) = \frac{b \operatorname{sn} u_t - a \operatorname{cn} u_t}{r},$$

furthermore $a\cos\theta(t) + b\sin\theta(t) = r \operatorname{sn} u_t$. From the ODE we see that

$$\begin{aligned}
\dot{x} &= a \operatorname{sn}^2 u_t + b \operatorname{sn} u_t \operatorname{cn} u_t = a \operatorname{sn}^2(Rt + u_0) + b \operatorname{sn}(Rt + u_0) \operatorname{cn}(Rt + u_0), \\
\dot{y} &= b \operatorname{sn}^2 u_t - a \operatorname{sn} u_t \operatorname{cn} u_t = b \operatorname{sn}^2(Rt + u_0) - a \operatorname{sn}(Rt + u_0) \operatorname{cn}(Rt + u_0).
\end{aligned}$$

Integrating using the calculus of the Jacobi elliptic functions we obtain the stated formulas for $x(t)$ and $y(t)$. □

3.6 Notes

Notes for Section 3.1**.** Jet spaces play a crucial role in several fields of mathematics and in applications. A nice description of their relevance for sub-Riemannian geometry can be found in Montgomery's book [203] (see Section 6.5 of [203]). The role of jet spaces in sub-Riemannian geometric function theory was recently investigated by Warhurst [256], [257], who studied them as examples of *non-rigid* Carnot groups, i.e., groups where not all quasiconformal deformations are conformal. Jet spaces are essentially the *only* known examples of non-rigid spaces, as they include Euclidean spaces and the Heisenberg groups. (Complexified Heisenberg groups are another class of non-rigid groups, although they are more properly classified as *semi-rigid*; all contact transformations are in fact holomorphic.) See [220] and Section 6.5 for more on rigidity.

Notes for Section 3.2**.** The local equivalence between the roto-translation goup and the Heisenberg group relies on the contact Darboux Theorem. This theorem is discussed in any of a number of sources, for example [4]. The weighted Taylor approximation of $\mathcal{RT}$ by $\mathbb{H}$ is described in detail (and in greater generality) in Bellaiche [32].

Notes for Subsection 3.2.1**.** As demonstrated in the example, the application of sub-Riemannian geometry to optimal control primarily involves the geodesic problem. In this context we would be remiss if we failed to mention a mistake in the literature which is often confusing for newcomers to the area. At an early stage in the differential geometric analysis of sub-Riemannian spaces, the analog of the Hamilton–Jacobi equations for the geodesic problem was derived. On more than one occasion, a proof was announced for the assertion that geodesics in a wide class of sub-Riemannian spaces were smooth. A variety of results were built upon this premise. However, as R. Montgomery first pointed out [201], there are examples of so-called *abnormal* geodesics which do not satisfy the geodesic equations. In the intervening years, various authors have provided numerous examples of abnormal minimizers, even in relatively simple sub-Riemannian spaces. The interested reader should see, for example, [178, 202].

There is an enormous literature in the study of sub-Riemannian geometries in the context of optimal control and path planning. While not attempting a comprehensive listing, we do point out several general references for the interested reader. One of the first connections between control theory and sub-Riemannian geometry was made by Brockett [45]. The book of Sontag [242] and the article of Agrachev [5] are other useful references. Recently, Laumond, Sekhavat and Lamiraux [175] compiled a nice survey article outlining path planning and control methods for mobile robots. Sussman [244] and Bloch [36] have written related surveys, while Bonnard and Chyba [39] have authored a book concerning the role of singular geodesics in control theory.

Notes for Subsection 3.2.2**.** The study of the geometry of the roto-translation group $\mathcal{RT}$ as a model for the function of the first layer V1 of the mammalian visual cortex is still in its infancy.

Early research on the structure and neurowiring of the cortex was done by Hubel and Weisel [150, 151]. The long range horizontal connectivity was established by Gilbert et al in 1996 [124]. The use of the sub-Riemannian contact geometry of the first jet space $J^1(\mathbb{R}, \mathbb{R})$ in the study of the neurogeometry of the visual cortex was proposed by Petitot. Quoting from his survey [227]:

> "*Jets are feature detectors specialized in the detection of tangents. The fact that* $V1$ *can be viewed as* $J^1(\mathbb{R}, \mathbb{R})$ *explains why* $V1$ *is functionally relevant for contour integration. On a 2-dimensional manifold* R*,*[10] *to determine the direction* p *of the tangent to a contour at a point* a *requires*

[10] For the sake of simplicity, the reader is invited to think of R as the retina.

to compare the values of the curve within a neighborhood of that point. But the system can access directly this geometrical information as a single numerical value in the 3*-dimensional jet space. This spares a local computation which would be very expensive in terms of wiring.*"

Building on earlier work of Hoffman [148] and Petitot and Tondut [227, 228], Citti and Sarti [68] showed that the role of V1 in completing missing or occluded image data from the retina can be modeled by solving the sub-Riemannian minimal surface problem with Dirichlet boundary conditions in $\mathcal{RT}$. In the same paper the authors introduce a digital impainting algorithm in digital image processing which mimics the way V1 processes data. For more results in this vein see [67], [238] and [66]. Recent work of Hladky and Pauls [146] discusses the nature of smooth sub-Riemannian minimal surfaces in the roto-translation space, showing that such surfaces are ruled and providing algorithms for constructing such surfaces. In a subsequent paper [145], these algorithms are converted to discrete algorithms suitable for solving occlusion problems for digital image reconstruction.

For an overview of the mathematics of this model see [68, 148, 227, 228] and for the neurobiological aspects, see, for example, [96, 150, 151, 236, 258].

Notes for Section 3.3**.** The local approximation of strictly pseudo-convex domains in $\mathbb{C}^n$ with their osculating Heisenberg groups goes back to the work of Folland–Kohn [101] and Folland–Stein [103]. For a detailed list of references see [243]. Our exposition follows closely the original paper of Korányi and Reimann [168]; note however that our slightly different presentation for the horizontal distribution adjusts some of the formulas. The results of this section extend *mutatis mutandis* to the higher-dimensional Heisenberg groups. See [171] for the details.

Notes for Section 3.4**.** Our main references for this section are the monographs [243, Chapter 12] and [43, Chapter II.10], as well as [168]. Another treatment is contained in [126]. The parabolic version (3.14) of the visual metric on the boundary of a Gromov hyperbolic space was introduced in Lemma 2.1 of [38]. A clear exposition of the Bergman metric, including the explicit computation for the unit ball, can be found in [173, §1.4].

The metric ρ_1 on the boundary of $H^2_{\mathbb{C}}$ in (3.22) was first introduced by Mostow [212, pp. 149–151]; see also [172] for related discussion. The verification that ρ_1 is a metric can be seen in [71] or [168, p. 321]. The proof of the triangle inequality is similar to the proof of (2.12) for the Korányi gauge, using instead the representation on the final line in (3.22).

To show that that $d([x],[y])$ in (3.17) is a metric, we note that the reverse Schwartz inequality [43, Lemma 10.3] yields $Q(x,y)Q(y,x) \geq Q(x,x)Q(y,y)$ for all x, y, with equality if and only if $x = y$. Thus $d([x],[y]) = 0$ if and only if $[x] = [y]$. The triangle inequality is proved in [43, Corollary 10.9]. The fact that the metric space $(H^2_{\mathbb{C}}, d)$ is Gromov hyperbolic is shown in [43, Theorem 10.10].

In this context we also would like to recall a related beautiful result of Balogh and Bonk [21]. Let $\Omega \subset \mathbb{C}^2$ be a bounded strictly pseudoconvex domain with C^2

boundary, and denote by δ_Ω the *signed distance function* from $\partial\Omega$. It can be shown that $\delta_\Omega \in C^2$ in a neighborhood of $\partial\Omega$, and as such can be used as a local defining function for Ω. For sufficiently small ϵ_0, and all $z \in \Omega$ with $\delta_\Omega(z) < \epsilon_0$, let $\pi(z) \in \partial\Omega$ be the closest point in $\partial\Omega$. Let $H\partial\Omega$ be the horizontal subbundle of $T\partial\Omega$ as defined in Section 3.3. We denote by $N\partial\Omega$ the orthogonal complement (with respect to the Hermitian norm in $\mathbb{C}^2$) of $H\partial\Omega$ in $T\partial\Omega$. Any tangent vector $Z \in T_z\partial\Omega$, $z \in \partial\Omega$, will have a unique decomposition $Z = Z_H + Z_N$ with $Z_H \in H_z\partial\Omega$ and $Z_N \in N_z\partial\Omega$. Using the Bergman metric (b_{ij}) on Ω (see (3.19)), we introduce the norm $|Z|_B^2 = \sum_{ij} b_{ij}(z) Z_i \overline{Z_j}$, for $z \in \Omega$ and $Z \in T_z\Omega$. Let

$$L(z, V) = \sum_{ij} \partial^2_{z_i \overline{z_j}} \delta(z) P_i \overline{P_j}$$

be the Levi form of $\partial\Omega$ and define, for all $z \in \partial\Omega$ and $P \in H_z\partial\Omega$, the norm $|P|_H(z) = \sqrt{L(z, P)}$. Finally, let d_H be the Carnot–Carathéodory distance on $\partial\Omega$ with respect to the horizontal distribution $H\partial\Omega$ and the norm $|\cdot|_H$. In the case when $\Omega = B(0,1)$ is the unit ball, the result of Balogh and Bonk (Proposition 1.2 in [21]) reads as follows: *For all $\epsilon > 0$, there exists $\epsilon_0 > 0$, $C \geq 0$ such that*

$$\begin{aligned}(1 - C\sqrt{\delta(z)}) &\left(\frac{|Z_N|^2}{4\delta(z)^2} + (1-\epsilon) \frac{L(\pi(z), Z_H)}{\delta(z)} \right)^{\frac{1}{2}} \\ &\leq |Z|_B \leq (1 + C\sqrt{\delta(z)}) \left(\frac{|Z_N|^2}{4\delta(z)^2} + (1+\epsilon) \frac{L(\pi(z), Z_H)}{\delta(z)} \right)^{\frac{1}{2}} \qquad (3.32)\end{aligned}$$

for all $Z \in \mathbb{C}^2$ and $z \in \Omega$ with $\delta(z) \leq \epsilon_0$. Here the splitting $Z = Z_H + Z_N$ is understood to be carried out at the point $\pi(z) \in \partial\Omega$. The same estimate holds for an arbitrary bounded strictly pseudoconvex C^2 domain $\Omega \subset \mathbb{C}^n$, provided the Bergman metric is replaced by the Kobayashi metric (a Finsler metric). See also [58] and other references in [21] for related results in the C^∞ category.

The inequalities in (3.32) give a precise description of the asymptoptic behavior of the Bergman (or Kobayashi) metrics with respect to the Levi form (which in some sense is the infinitesimal Carnot–Carathéodory metric) of $\partial\Omega$ as one approaches the boundary. This result is the starting point for a comparison theorem between the corresponding (integrated) distances [21, Theorem 1.4]. As a consequence, one finds that any bounded strictly pseudoconvex domain Ω equipped with the Kobayashi metric becomes Gromov hyperbolic, and the Carnot–Carathéodory metric d_H on $\partial\Omega$ is a visual metric for the corresponding Gromov boundary $\partial_\infty\Omega$.

Chapter 4

Horizontal Geometry of Submanifolds

This chapter is devoted to the study of the sub-Riemannian geometry of codimension 1 smooth submanifolds of the Heisenberg group.

In Section 4.1 we show that any two Riemannian extensions of a fixed sub-Riemannian metric give rise to the same horizontal part of the Levi-Civita connection. This fact is at the basis of our approach as it allows us to define intrinsic notions of horizontal curvature tensors. In Section 4.2 we explicitly calculate the basic differential geometric machinery on such submanifolds, relative to the Riemannian metrics g_L. We focus in particular on the dependence of this machinery on the parameter L, which governs how the relevant quantities diverge or degenerate in the sub-Riemannnian limit. In Section 4.3 we introduce the sub-Riemannian notion of horizontal mean curvature as a limit of the corresponding Riemannian objects. In Section 4.3.2 we propose an equivalent, extrinsic definition of horizontal second fundamental form, and show that in $\mathbb{H}$ the horizontal mean curvature of a surface S can be computed by lifting the classical planar curvature to a Legendrian foliation of S. In all of the preceding discussion we focus our attention on noncharacteristic points of S, e.g., points $x \in S$ where the tangent space T_xS does not coincide with the horizontal plane $H(x)$. In Section 4.4 we show via a few examples the pathologies that can arise at characteristic points, and present some crucial results, due to Cheng, Hwang, Malchiodi and Yang, concerning the fine behavior of the Legendrian foliation in neighborhoods of the characteristic locus.

4.1 Invariance of the Sub-Riemannian Metric with respect to Riemannian extensions

One of our main tools in the investigation of sub-Riemannian geometry is the use of Riemannian completions of the sub-Riemannian inner product on the horizontal bundle. We do this mainly to exploit properties of the Levi-Civita connection. For example, in the previous chapter we used the family of metrics, g_L, and the corresponding connections to compute the curvature blow-up of the Riemannian approximants to $\mathbb{H}$. As we have mentioned before, our approach is to define sub-Riemannian objects as limits of horizontal objects in $(\mathbb{R}^3, g_L)$. At the heart of this approach is the fact that the intrinsic horizontal geometry does not change with L. In general, this will not be the case, as different choices of Riemannian extension will certainly affect the horizontal component, however, this does not happen for *natural* extensions, i.e., extensions in which the different layers of the Lie algebra stratification are orthogonal.

Proposition 4.1. *Let $\langle\cdot,\cdot\rangle$ be a smoothly varying inner product on the horizontal subbundle H of $\mathbb{H}^n$. Denote by V the vector bundle obtained by left translation of the center of the group. If g_1, g_2 are Riemannian metrics in $\mathbb{H}^n$ which make H and V orthogonal and $g_i|_H(\cdot,\cdot) = \langle\cdot,\cdot\rangle$, $i = 1, 2$, then the associated Levi-Civita connections, ∇_1, ∇_2, to g_1, g_2 coincide when projected to H. In other words, if U, V, W are sections of H, then*

$$g_1(\nabla_{1U}V, W) = g_2(\nabla_{2U}V, W).$$

Proof. The Kozul identity (2.30) yields

$$\begin{aligned} g_i(\nabla_{iU}V, W) = \frac{1}{2}\{&Ug_i(V,W) + Vg_i(W,U) - Wg_i(U,V) \\ &- g_i(W,[V,U]) - g_i([V,W],U) - g_i(V,[U,W])\}. \end{aligned} \tag{4.1}$$

Consider the first three terms on the right-hand side of (4.1). Since $g_1 = g_2$ when applied to horizontal vectors, these terms are identical for $i = 1, 2$. Moreover, the computation of the brackets does not depend on the choice of metric and we claim that the last three terms do not depend on the choice of extension. For example, if $U = \sum_l u_l X_l$, $V = \sum_l v_l X_l$, $W = \sum_l w_l X_l$, then

$$[V,U] = \sum_{l,j} (v_l X_l(u_j) - u_l X_l(v_j))X_j + u_l v_j [X_j, X_l].$$

By the bracket generating property of the horizontal layer, $[X_j, X_l] \in V$; since moreover H and V are orthogonal for both g_1 and g_2, we have

$$g_i(W,[V,U]) = \sum_{l,j} w_j(v_l X_l(u_j) - u_l X_l(v_j)).$$

In a similar fashion, the remaining terms in (4.1) do not depend on the choice of extension. □

The same result holds for any Carnot group with the same proof. In this more general setting the statement is as follows:

Proposition 4.2. *Let G be a Carnot group, H the subbundle of TM generated by left translation of the first layer of the grading, V a subbundle of TM complementary to H, and $\langle \cdot, \cdot \rangle$ a smoothly varying inner product on H. If g_1 and g_2 are Riemannian metrics which make H and V orthogonal and $g_i(\cdot, \cdot)|_H = \langle \cdot, \cdot \rangle$, $i = 1, 2$, then the associated Levi-Civita connections, ∇_1, ∇_2, coincide when projected to H:*

$$g_1(\nabla_{1U} V, W) = g_2(\nabla_{2U} V, W)$$

whenever U, V, W are sections of H.

4.2 The second fundamental form in $(\mathbb{R}^3, g_L)$

Consider a C^2 surface

$$S = \left\{ x \in \mathbb{R}^3 : u(x) = 0 \right\}, \tag{4.2}$$

which is regular, i.e., $u \in C^2(\mathbb{R}^3)$ has nonvanishing gradient along S. In this section, we will compute the second fundamental form II^L of S with respect to the Riemannian metric g_L on $\mathbb{R}^3$ at points where the horizontal gradient $(X_1 u, X_2 u)$ is nonvanishing. To illustrate the complexity of a direct approach, we compute here using an explicit orthonormal frame adapted to the submanifold S and the grading of the Lie algebra. In Section 4.3.1, we will give another derivation via a more intrinsic approach.

We require additional notation. Let $X_i, \tilde{X}_i$, $i = 1, 2, 3$, be as in Section 2.4.2 and set $p = X_1 u$, $q = X_2 u$ and $r = \tilde{X}_3 u$. To simplify the upcoming formulas, we set

$$l = \sqrt{p^2 + q^2}, \quad l_L = \sqrt{p^2 + q^2 + r^2},$$

and $\overline{p} = p/l$, $\overline{q} = q/l$, $\overline{p}_L = p/l_L$, $\overline{q}_L = q/l_L$, $\overline{r}_L = r/l_L$. Observe that the Riemannian normal to S is

$$\nu_L = \overline{p}_L X_1 + \overline{q}_L X_2 + \overline{r}_L \tilde{X}_3,$$

while the tangent space to S is spanned by the orthonormal basis

$$e_1 = \overline{q} X_1 - \overline{p} X_2, \quad \text{and} \quad e_2 = \overline{r}_L \overline{p} X_1 + \overline{r}_L \overline{q} X_2 - \frac{l}{l_L} \tilde{X}_3. \tag{4.3}$$

The second fundamental form of S is

$$II^L = \begin{pmatrix} \langle \nabla_{e_1} \nu_L, e_1 \rangle_L & \langle \nabla_{e_1} \nu_L, e_2 \rangle_L \\ \langle \nabla_{e_2} \nu_L, e_1 \rangle_L & \langle \nabla_{e_2} \nu_L, e_2 \rangle_L \end{pmatrix},$$

where ∇ denotes the Levi-Civita connection associated to g_L (see Section 2.4.2). The principal result of this section is the following.

Theorem 4.3. *Let $S \subset (\mathbb{R}^3, g_L)$ be a regular C^2 surface defined as in* (4.2)*. Relative to the orthonormal frame $\{e_1, e_2\}$ defined in* (4.3)*, the second fundamental form of S is*

$$II^L = \begin{pmatrix} \frac{l}{l_L}\left(X_1\overline{p} + X_2\overline{q}\right) & -\frac{l_L}{l}\langle e_1, \nabla_0 \overline{r}_L\rangle_L - \frac{\sqrt{L}}{2} \\ -\frac{l_L}{l}\langle e_1, \nabla_0 \overline{r}_L\rangle_L - \frac{\sqrt{L}}{2} & -\frac{l^2}{l_L^2}\langle e_2, \nabla_0 (\frac{r}{l})\rangle_L + \tilde{X}_3\overline{r}_L \end{pmatrix}. \tag{4.4}$$

In particular, the mean curvature $\mathcal{H}_L = \operatorname{Trace} II^L$ of S is

$$\frac{l}{l_L}(X_1\overline{p} + X_2\overline{q}) - \frac{l^2}{l_L^2}\left\langle e_2, \nabla_0\left(\frac{r}{l}\right)\right\rangle_L + \tilde{X}_3\overline{r}_L, \tag{4.5}$$

while the Gauss curvature $\mathcal{K}_L = \det II^L$ is

$$\frac{l}{l_L}\left(X_1\overline{p} + X_2\overline{q}\right)\left(-\frac{l^2}{l_L^2}\left\langle e_2, \nabla_0\left(\frac{r}{l}\right)\right\rangle_L + \tilde{X}_3\overline{r}_L\right) - \left(\frac{l_L}{l}\langle e_1, \nabla \overline{r}_L\rangle_L + \frac{\sqrt{L}}{2}\right)^2. \tag{4.6}$$

To prove Theorem 4.3, we will compute each of the entries in II^L in turn.

$\underline{\langle \nabla_{e_1}\nu_L, e_1\rangle_L = -\langle \nabla_{e_1} e_1, \nu_L\rangle_L}$**.** Using the definition of the connection, the identities in (2.31) and grouping terms, we have

$$\begin{aligned} \nabla_{e_1} e_1 &= \overline{q}(X_1\overline{q}\, X_1 - X_1\overline{p}\, X_2 - \overline{p}\nabla_{X_1}X_2) - \overline{p}(X_2\overline{q}\, X_1 + \overline{q}\nabla_{X_2}X_1 - X_2\overline{p}\, X_2) \\ &= (\overline{q}X_1\overline{q} - \overline{p}X_2\overline{q})\, X_1 - (\overline{q}X_1\overline{p}) - \overline{p}X_2\overline{p})\, X_2. \end{aligned}$$

Since $\overline{p}^2 + \overline{q}^2 = 1$, we have $\overline{p}X_i\overline{p} + \overline{q}X_i\overline{q} = 0$ and $-\langle \nabla_{e_1} e_1, \nu_L\rangle_L$ simplifies to give the desired expression.

In Lemma 4.5, we will identify the limit of $\langle \nabla_{e_1}\nu_L, e_1\rangle_L$ as $L \to \infty$ as the sub-Riemannian *horizontal mean curvature* of S.

$\underline{\langle \nabla_{e_1}\nu_L, e_2\rangle_L = -\langle \nabla_{e_1} e_2, \nu_L\rangle_L}$**.** Using the definition of the connection, the identities in (2.31) and grouping terms, we have

$$\begin{aligned} \nabla_{e_1} e_2 = {} & \overline{q}\left(X_1(\overline{pr}_L)\, X_1 + X_1(\overline{qr}_L)\, X_2 + \overline{qr}_L\nabla_{X_1}X_2 - X_1\left(\frac{l}{l_L}\right)\tilde{X}_3 - \frac{l}{l_L}\nabla_{X_1}\tilde{X}_3\right) \\ & - \overline{p}\left(X_2(\overline{pr}_L)\, X_1 + \overline{pr}_L\nabla_{X_2}X_1 + X_2(\overline{qr}_L)\, X_2 - X_2\left(\frac{l}{l_L}\right)\tilde{X}_3 - \frac{l}{l_L}\nabla_{X_2}\tilde{X}_3\right) \\ = {} & \left(\overline{q}X_1(\overline{pr}_L) - \overline{p}X_2(\overline{pr}_L) + \frac{\overline{p}_L\sqrt{L}}{2}\right) X_1 + \left(\overline{q}X_1(\overline{qr}_L) - \overline{p}X_2(\overline{qr}_L) + \frac{\overline{q}_L\sqrt{L}}{2}\right) X_2 \\ & + \left(\frac{\overline{r}_L\sqrt{L}}{2} - \overline{q}X_1\left(\frac{l}{l_L}\right) + \overline{p}X_2\left(\frac{l}{l_L}\right)\right)\tilde{X}_3. \end{aligned}$$

Next, we compute the inner product of this with ν_L. Using the product rule and the identity $\overline{q}_L\overline{p} = \overline{p}_L\overline{q}$, we obtain

$$\begin{aligned}
\langle \nabla_{e_1} e_2, \nu_L\rangle_L &= \overline{qp}_L(\overline{p}X_1\overline{r}_L + \overline{r}_L X_1\overline{p}) - \overline{pp}_L(\overline{p}X_2\overline{r}_L + \overline{r}_L X_2\overline{p}) \\
&\quad + \overline{q}_L\overline{q}(\overline{q}X_1\overline{r}_L + \overline{r}_L X_1\overline{q}) - \overline{pq}_L(\overline{q}X_2\overline{r}_L + \overline{r}_L X_2\overline{q}) \\
&\quad + \frac{\sqrt{L}}{2}\left(\overline{p}_L^2 + \overline{q}_L^2 + \overline{r}_L^2\right) + \overline{r}_L\overline{p}X_2\left(\frac{l}{l_L}\right) - \overline{r}_L\overline{q}X_1\left(\frac{l}{l_L}\right) \\
&= \overline{qp}_L\overline{p}X_1\overline{r}_L - \overline{p}^2\overline{p}_L X_2\overline{r}_L + \overline{q}_L\overline{q}^2 X_1\overline{r}_L - \overline{pq}_L\overline{q}X_2\overline{r}_L \\
&\quad + \frac{\sqrt{L}}{2}\left(\overline{p}_L^2 + \overline{q}_L^2 + \overline{r}_L^2\right) + \overline{r}_L\overline{p}X_2\left(\frac{l}{l_L}\right) - \overline{r}_L\overline{q}X_1\left(\frac{l}{l_L}\right).
\end{aligned}$$

The identities $\overline{p}_L^2 + \overline{q}_L^2 + \overline{r}_L^2 = 1$ and $\overline{p}^2 + \overline{q}^2 = 1$ yield

$$\begin{aligned}
\langle \nabla_{e_1} e_2, \nu_L\rangle_L &= \frac{l}{l_L}\overline{q}X_1\overline{r}_L - \frac{l}{l_L}\overline{p}X_2\overline{r}_L + \overline{r}_L\overline{p}X_2\left(\frac{l}{l_L}\right) - \overline{r}_L\overline{q}X_1\left(\frac{l}{l_L}\right) + \frac{\sqrt{L}}{2} \\
&= \frac{l}{l_L}\langle e_1, \nabla_0\overline{r}_L\rangle_L - \overline{r}_L\left\langle e_1, \nabla_0\left(\frac{l}{l_L}\right)\right\rangle_L + \frac{\sqrt{L}}{2}.
\end{aligned} \tag{4.7}$$

Finally we use the identity $(l/l_L - l_L/l)\nabla_0\overline{r}_L = \overline{r}_L\nabla_0(l/l_L)$ in (4.7).

As the second fundamental form II^L is symmetric, we expect to obtain $\langle \nabla_{e_2}\nu_L, e_1\rangle_L = \langle \nabla_{e_1}\nu_L, e_2\rangle_L$. For the sake of completeness we verify the value of $\langle \nabla_{e_2}\nu_L, e_1\rangle$ separately.

$\underline{\langle \nabla_{e_2}\nu_L, e_1\rangle_L = -\langle \nabla_{e_2} e_1, \nu_L\rangle_L}$. Using the definition of the connection, the identities in (2.31) and grouping terms as in the previous computation, we have

$$\begin{aligned}
\nabla_{e_2} e_1 &= \left(\overline{pr}_L X_1(\overline{q}) + \overline{qr}_L X_2(\overline{q}) - \frac{l}{l_L}\tilde{X}_3(\overline{q}) + \frac{\overline{p}_L\sqrt{L}}{2}\right) X_1 \\
&\quad + \left(-\overline{pr}_L X_1(\overline{p}) - \overline{qr}_L X_2(\overline{p}) + \frac{l}{l_L}\tilde{X}_3(\overline{p}) + \frac{\overline{q}_L\sqrt{L}}{2}\right) X_2 - \frac{\overline{r}_L\sqrt{L}}{2}\tilde{X}_3.
\end{aligned}$$

Next, we compute the inner product of this with ν_L. In the X_1 and X_2 terms, we use the quotient rule for $\tilde{X}_3\overline{p} = \tilde{X}_3(p/l)$ and $\tilde{X}_3\overline{q} = \tilde{X}_3(q/l)$ and the commutation relations $\tilde{X}_3 p = X_1 r$, $\tilde{X}_3 q = X_2 r$. The result is $\langle \nabla_{e_2} e_1, \nu_L\rangle_L =$

$$\overline{p}_L\overline{r}_L(\overline{p}X_1\overline{q} + \overline{q}X_2\overline{q}) - \overline{q}_L\overline{r}_L(\overline{p}X_1\overline{p} + \overline{q}X_2\overline{p}) + \frac{qX_1 r - pX_2 r - r^2\sqrt{L}}{l_L^2} + \frac{\sqrt{L}}{2}.$$

In the penultimate term, use the commutation relation $r\sqrt{L} = X_1 q - X_2 p$ together with the product rules for $X_2 p = X_2(l_L\overline{p}_L)$, $X_1 q = X_1(l_L\overline{q}_L)$ and $X_i r = X_i(l_l\overline{r}_L)$.

In the first two terms, use the quotient rules for $X_i\overline{p} = X_i(\frac{\overline{p}_L}{l/l_L})$ and $X_i\overline{q} = X_i(\frac{\overline{q}_L}{l/l_L})$. The result is

$$\begin{aligned}\langle \nabla_{e_2} e_1, \nu_L\rangle_L &= \overline{r}_L(\overline{p}^2 X_1\overline{q}_L + \overline{p}\,\overline{q}X_2\overline{q}_L - \overline{p}\,\overline{q}X_1\overline{p}_L - \overline{q}^2X_2\overline{p}_L)\\ &\quad + \overline{q}_LX_1\overline{r}_L - \overline{p}_LX_2\overline{r}_1 - \overline{r}_LX_1\overline{q}_L + \overline{r}_LX_2\overline{p}_L + \frac{\sqrt{L}}{2}.\end{aligned}$$

Using the identities $\overline{p}^2 + \overline{q}^2 = 1$, $\overline{r}_L\overline{p} = (r/l)\overline{p}_L$, $\overline{r}_L\overline{q} = (r/l)\overline{q}_L$, $\overline{p}_L^2 + \overline{q}_L^2 + \overline{r}_L^2 = 1$ and $\overline{p}_LX_i\overline{p}_L + \overline{q}_LX_i\overline{q}_L = -\overline{r}_LX_i\overline{r}_L$ and grouping terms, we obtain

$$\begin{aligned}\langle \nabla_{e_2} e_1, \nu_L\rangle_L &= \left(\overline{q}_L + \overline{q}_L\frac{r^2}{l^2}\right)X_1\overline{r}_L - \left(\overline{p}_L + \overline{p}_L\frac{r^2}{l^2}\right)X_2\overline{r}_L + \frac{\sqrt{L}}{2}\\ &= \frac{l_L}{l}\langle e_1, \nabla_0\overline{r}_L\rangle_L + \frac{\sqrt{L}}{2}.\end{aligned}$$

<u>$\langle \nabla_{e_2}\nu_L, e_2\rangle_L = -\langle \nabla_{e_2}e_2, \nu_L\rangle_L$.</u> Using the definition of connection, the identities in (2.31) and grouping terms, we have:

$$\begin{aligned}\nabla_{e_2}e_2 =& \overline{pr}_L\left(X_1(\overline{pr}_L)X_1 + X_1(\overline{qr}_L)X_2 + \overline{qr}_L\nabla_{X_1}X_2 - X_1\left(\frac{l}{l_L}\right)\tilde{X}_3 - \frac{l}{l_L}\nabla_{X_1}\tilde{X}_3\right)\\ &+\overline{qr}_L\left(X_2(\overline{pr}_L)X_1 + \overline{pr}_L\nabla_{X_2}X_1 + X_2(\overline{qr}_L)X_2 - X_1\left(\frac{l}{l_L}\right)\tilde{X}_3 - \frac{l}{l_L}\nabla_{X_2}\tilde{X}_3\right)\\ &-\frac{l}{l_L}\left(\tilde{X}_3(\overline{pr}_L)X_1 + \overline{pr}_L\nabla_{\tilde{X}_3}X_1 + \tilde{X}_3(\overline{qr}_L)x_2 + \overline{qr}_L\nabla_{\tilde{X}_3}X_2 - \tilde{X}_3\left(\frac{l}{l_L}\right)\tilde{X}_3\right)\\ =& \left(\overline{pr}_LX_1(\overline{pr}_L) + \overline{qr}_LX_2(\overline{pr}_L) - \frac{l}{l_L}\tilde{X}_3(\overline{pr}_L) - \overline{q}_L\overline{r}_L\sqrt{L}\right)X_1\\ &+\left(\overline{pr}_LX_1(\overline{qr}_L) + \overline{qr}_LX_2(\overline{qr}_L) - \frac{l}{l_L}\tilde{X}_3(\overline{qr}_L) + \overline{p}_L\overline{r}_L\sqrt{L}\right)X_2\\ &+\left(-\overline{pr}_LX_1\left(\frac{l}{l_L}\right) - \overline{qr}_LX_2\left(\frac{l}{l_L}\right) + \frac{l}{l_L}\tilde{X}_3\left(\frac{l}{l_L}\right)\right)\tilde{X}_3.\end{aligned}$$

Taking the inner product with ν_L yields

$$\begin{aligned}\langle \nabla_{e_2} e_2, \nu_L\rangle_L &= \overline{p}_L\left(\overline{pr}_LX_1(\overline{pr}_L) + \overline{qr}_LX_2(\overline{pr}_L) - \frac{l}{l_L}\tilde{X}_3(\overline{pr}_L)\right)\\ &\quad + \overline{q}_L\left(\overline{pr}_LX_1(\overline{qr}_L) + \overline{qr}_LX_2(\overline{qr}_L) - \frac{l}{l_L}\tilde{X}_3(\overline{qr}_L)\right)\\ &\quad + \overline{r}_L\left(-\overline{pr}_LX_1\left(\frac{l}{l_L}\right) - \overline{qr}_LX_2\left(\frac{l}{l_L}\right) + \frac{l}{l_L}\tilde{X}_3\left(\frac{l}{l_L}\right)\right).\end{aligned}$$

To simplify this, first use the product rule for the terms involving $X_i(\overline{pr}_L)$ and $X_i(\overline{qr}_L)$ together with the identities $\overline{p}X_i\overline{p}+\overline{q}X_i\overline{q} = 0, \overline{r}_L\nabla_0(\frac{l}{l_L}) = \left(\frac{l}{l_L} - \frac{l_L}{l}\right)\nabla_0\overline{r}_L$

and $\overline{p}^2 + \overline{q}^2 = 1$. Under these simplifications, terms involving $X_i(\overline{p})$ and $X_i(\overline{q})$ cancel and one is left with terms involving components of $\nabla\overline{r}_L$:

$$\langle \nabla_{e_2} e_2, \nu_L \rangle_L = \overline{p}_L \left(\overline{p}_L + \overline{p}\frac{r}{l} \right) X_1(\overline{r}_L) + \overline{q}_L \left(\overline{q}_L + \overline{q}\frac{r}{l} \right) X_2(\overline{r}_L) - \tilde{X}_3(\overline{r}_L).$$

We conclude by rewriting the expression $X_i(\overline{r}_L)$ in terms of $X_i\left(\frac{r}{l}\right)$.

4.3 Horizontal geometry of hypersurfaces in $\mathbb{H}$

In this section we want to examine the behavior of the second fundamental form (4.4) and curvatures (4.5), (4.6) as $L \to \infty$. As expected, the horizontal components have well-defined limits which are natural candidates for sub-Riemannian analogs of these classical differential geometric quantities. The vertical components are unbounded, corresponding to the blow-up of the curvature of $(\mathbb{H}, g_L)$.

Before initiating this analysis we take a moment to introduce a fundamental notion in the study of sub-Riemannian submanifold geometry.

Definition 4.4. Let $S \subset \mathbb{H}$ be a C^1 surface defined as in (4.2). The *characteristic set* of S is the closed set

$$\Sigma(S) = \{x \in S : \nabla_0 u(x) = 0\}. \tag{4.8}$$

In other words, $\Sigma(S)$ is the set of points where the tangent space is purely horizontal.

Note that $\Sigma(S)$ is nowhere dense in S, as follows from the Frobenius integrability theorem. In fact, the surface measure of $\Sigma(S)$ is equal to zero. For more precise statements on the size of $\Sigma(S)$, see the notes to this chapter.

Note that r, $\overline{r}_L$ and $\tilde{X}_3$ all converge to zero as $L \to \infty$ to zero at rates on the order of $L^{-1/2}$. On the other hand, $\overline{q}_L \to \overline{q}$, $\overline{p}_L \to \overline{p}$, $l_L \to l$, and $e_2 \to 0$. Hence the Riemannian normal ν_L converges to the so-called *horizontal normal*

$$\nu_H = \sum_{i=1}^{2} \frac{X_i u}{|\nabla_0 u|} X_i \in L^\infty(S \setminus \Sigma(S)) \tag{4.9}$$

in the complement of the characteristic set. Note that ν_H is simply the projection of ν_L onto the horizontal subbundle. The vectors ν_H and e_1 (see (4.3)) form an orthonormal frame of the horizontal subbundle.

Direct computation shows that the Gauss curvature $\mathcal{K}_L$ diverges as $L \to \infty$ (similarly to the behavior of the sectional, Ricci and scalar curvatures discussed in Section 2.4.2). Indeed

$$\lim_{L\to\infty} \frac{\mathcal{K}_L}{L} = -\frac{1}{4}.$$

This reflects the fact that $\nu_L \to \overline{p}\ X_1 + \overline{q}\ X_2$ as $L \to \infty$, i.e., the tangent plane to S tends, as $L \to \infty$, towards a vertical plane. As (2.33) and (2.34) show, the curvature of such planes computed with respect to g_L equals $-L/4$.

Surprisingly, while the Gauss curvature does not have a limit as $L \to \infty$, the mean curvature presents a rather different behavior. The following lemma is an immediate consequence of (4.5).

Lemma 4.5. *Let S be a C^2 regular surface defined as in* (4.2). *Then*

$$\lim_{L\to\infty} \operatorname{Trace} II^L = X_1\overline{p} + X_2\overline{q} \tag{4.10}$$

at noncharacteristic points.

Definition 4.6. Let $S \subset \mathbb{H}$ be a C^2 regular surface, given as a level set of a function u. The *horizontal mean curvature* of S at a noncharacteristic point is[1]

$$\mathcal{H}_0 = X_1\overline{p} + X_2\overline{q},$$

where $\overline{p} = p/l$, $\overline{q} = q/l$, $l = \sqrt{p^2+q^2}$, and $(p,q) = (X_1u, X_2u)$.

We can write the horizontal mean curvature $\mathcal{H}_0$ in several ways. First,

$$\mathcal{H}_0 = \sum_{i=1}^{2} X_i \left(\frac{X_i u}{|\nabla_0 u|} \right). \tag{4.11}$$

A direct computation shows that the horizontal mean curvature can also be expressed via the identity

$$\mathcal{H}_0\, |\nabla_0 u| = \mathcal{L}u - \frac{\mathcal{L}_\infty u}{|\nabla_0 u|^2}, \tag{4.12}$$

where $\mathcal{L}u = X_1^2 u + X_2^2 u$ is the Heisenberg Laplacian of u and

$$\mathcal{L}_\infty u = \sum_{i,j=1}^{2} X_i u X_j u X_i X_j u$$

is the Heisenberg *infinite Laplacian* of $u : \mathbb{H} \to \mathbb{R}$. Finally, if $u(x) = x_3 - f(|z|)$ and we let $r = |z|$, then

$$\mathcal{H}_0 = -\frac{\frac{1}{4}r^2 f'' + \frac{(f')^3}{r}}{((f')^2 + \frac{1}{4}r^2)^{\frac{3}{2}}}. \tag{4.13}$$

In Section 6.4 we will see a derivation of the horizontal mean curvature as a *first variation* of the perimeter functional among all horizontal perturbations.

[1] We note that some authors define the mean curvature as $\frac{1}{2}(X_1\overline{p} + X_2\overline{q})$.

Example 4.7. Both the plane $\{x_3 = 0\}$, and the saddle surfaces $\{x_3 = \pm\frac{x_1x_2}{2}\}$ in $\mathbb{H}$ have vanishing horizontal mean curvature away from the characteristic locus $\Sigma = \{o\}$.

Definition 4.8. A C^2 regular surface $S \subset \mathbb{H}$ is a *horizontal minimal surface* if it has vanishing horizontal mean curvature along its noncharacteristic locus.

Lemma 4.5 immediately implies:

Theorem 4.9. *Let $S \subset \mathbb{H}$ be a C^2 regular surface and let II^L be its second fundamental form computed with respect to g_L. If*

$$\lim_{L\to\infty} \operatorname{Trace} II^L = 0$$

then S is a horizontal minimal surface.

Example 4.10. The horizontal mean curvature of the Euclidean sphere $\{(z, x_3) : |z|^2 + x_3^2 = R^2\}$ diverges near the characteristic locus $\{(0, 0, \pm R)\}$ at a rate proportional to $|z|^{-1}$. In fact,

$$\mathcal{H}_0 = \frac{2(4 + R^2)}{|z|(4 + x_3^2)^{3/2}}.$$

A similar phenomenon holds for the Euclidean paraboloid (and sub-Riemannian *cone*) $P_\alpha = \{(z, x_3) : x_3 = \alpha|z|^2\}$, whose horizontal mean curvature diverges near the characteristic locus $\{o\}$ at a rate proportional to $|z|^{-1}$:

$$\mathcal{H}_0 = -\frac{4\alpha}{\sqrt{1 + 16\alpha^2}} \cdot \frac{1}{|z|}.$$

Example 4.11. The horizontal mean curvature of the Korányi sphere $\{(z, x_3) : |z|^4 + 16x_3^2 = R^4\}$, on the other hand, tends to zero near the characteristic locus $\{(0, 0, \pm R)\}$ at a rate proportional to $|z|$. In fact,

$$\mathcal{H}_0 = \frac{3|z|}{R^2}.$$

Example 4.12. The horizontal mean curvature of the CC sphere $\partial B_{cc}(o, R) = \{x \in \mathbb{H} : d(x, o) = R\}$ can be computed via (4.13). From (2.23) one easily deduces the parametric representation $|z| = A(c) := (2/c)\sin(cR/2)$, $x_3 = B(c) := (cR - \sin(cR))/(2c^2)$ for $\partial B_{cc}(o, R)$. With $x_3 = f(|z|)$, $f = B \circ A^{-1}$, a simple computation gives

$$\mathcal{H}_0 = \frac{1}{2} \cdot \frac{c/2}{\sin(cR/2)} \cdot \frac{\sin(cR) - cR\cos(cR)}{\sin(cR/2) - (cR/2)\cos(cR/2)}.$$

It can be shown that $\mathcal{H}_0 \sim |z|^{-1}$ as $x = (z, x_3)$ approaches the x_3-axis.

The following example is crucial in the study of the isoperimetric profile of $\mathbb{H}$ (see Chapter 8).

Example 4.13. Choosing

$$f(r) = f_R(r) = \pm\frac{1}{4}(r\sqrt{R^2 - r^2} + R^2 \arccos r/R),$$

one easily computes that the horizontal mean curvature of the boundaries of the bubble sets $\mathcal{B}(o,R)$ defined in Section 2.3 is equal to the constant $2/R$.

4.3.1 Horizontal geometry of hypersurfaces in $\mathbb{H}^n$

In this subsection, we repeat the analysis of the preceding sections for hypersurfaces in the higher-dimensional Heisenberg groups. Again, we study the limit as $L \to \infty$ of the horizontal part of the second fundamental form of a hypersurface S in $(\mathbb{H}^n, g_L)$. In contrast with the previous section, where we computed very explicitly using a specific frame, we use here only basic properties of the Levi-Civita connection to accomplish our analysis.

Let $S = \{x \in \mathbb{H}^n : u(x) = 0\}$ be a C^2 regular hypersurface. The characteristic set $\Sigma(S)$ is defined as before: it consists of all points $x \in S$ where the horizontal space $H(x)$ and the tangent space T_xS agree.

Consider left invariant vector fields $\tilde{X}_1, \dots, \tilde{X}_{2n+1}$ and a Riemannian metric g_L in $\mathbb{R}^{2n+1}$ as in Section 2.4.5. Let $|\nabla_L u|^2 = \sum_{i=1}^{2n+1}(\tilde{X}_i u)^2$ and observe that the vector

$$\nu_L = \frac{1}{|\nabla_L u|}\sum_{i=1}^{2n+1} \tilde{X}_i u \tilde{X}_i$$

is the unit normal to S in $(\mathbb{H}^n, g_L)$.

As before, we restrict attention to noncharacteristic points. Let

$$\nu_H := \lim_{L\to\infty} \nu_L = \frac{1}{|\nabla_0 u|}\sum_{i=1}^{2n} X_i u X_i.$$

Then

$$\nu_L = \alpha_L \nu_H + \beta_L \tilde{X}_{2n+1}, \tag{4.14}$$

where $\alpha_L = \langle \nu_L, \nu_H\rangle_L$ and $\beta_L = \langle \nu_L, \tilde{X}_{2n+1}\rangle_L$.

Lemma 4.14. *$\lim_{L\to\infty} \alpha_L = 1$ and $\beta_L = O(1/\sqrt{L})$ on $S \setminus \Sigma(S)$.*

Proof. $\alpha_L = \langle \nu_L, \nu_H\rangle_L = \frac{|\nabla_0 u|}{|\nabla_L u|}$ and $\beta_L = \langle \nu_L, \tilde{X}_{2n+1}\rangle_L = \frac{\tilde{X}_{2n+1}u}{|\nabla_L u|} = O(1/\sqrt{L})$. □

Set

$$e_{2n} = \beta_L \nu_H - \alpha_L \tilde{X}_{2n+1}$$

and observe that e_{2n} is unit and orthogonal to ν_L, hence tangent to S. Choose horizontal tangent vector fields $e_1, \ldots, e_{2n-1}$ so that $\{e_1, \ldots, e_{2n-1}, e_{2n}\}$ is an orthonormal frame of TS. The second fundamental form of S in $(\mathbb{H}^n, g_L)$ has entries

$$II_{ij}^L = \langle \nabla_{e_i} \nu_L, e_j \rangle_L$$

for $i, j = 1, \ldots, 2n$. In view of (4.14) we can write

$$II_{ij}^L = \alpha_L h_{ij}^L + \beta_L v_{ij}^L \tag{4.15}$$

for $i, j = 1, \ldots, 2n-1$, where

$$h_{ij}^L = \langle \nabla_{e_i} \nu_H, e_j \rangle_L, \qquad i, j = 1, \ldots, 2n-1, \tag{4.16}$$

is the so-called *horizontal second fundamental form*, and

$$v_{ij}^L = \langle \nabla_{e_i} \tilde{X}_{2n+1}, e_j \rangle_L, \qquad i, j = 1, \ldots, 2n-1$$

is its vertical complement.

Remark 4.15. In view of Proposition 4.1 it is clear that the terms h_{ij}^L are actually independent of L. We therefore omit the superscript L in what follows. Note also that in general the coefficients v_{ij}^L do not vanish as $L \to \infty$.

Proposition 4.16. *The matrix (v_{ij}^L) is anti-symmetric.*

The proof is an easy consequence of the following elementary lemma.

Lemma 4.17.

$$\left\langle \nabla_U V, \tilde{X}_{2n+1} \right\rangle_L = -\frac{1}{2} \left\langle [V, U], \tilde{X}_{2n+1} \right\rangle_L \tag{4.17}$$

for all orthonormal horizontal vectors U, V.

Proof. (4.17) is a direct consequence of basic properties of the Levi-Civita connection. Here we present a proof using coordinate frames. Write $U = \sum_{i=1}^{2n} a_i \tilde{X}_i$ and $V = \sum_{i=1}^{2n} b_i \tilde{X}_i$. Observe that

$$[U, \tilde{X}_{2n+1}] = \sum_{l=1}^{2n} \left(a_l [\tilde{X}_l, \tilde{X}_{2n+1}] - (\tilde{X}_{2n+1} a_l) \tilde{X}_l \right) = -\sum_{l=1}^{2n} (\tilde{X}_{2n+1} a_l) \tilde{X}_l,$$

while $[V, \tilde{X}_{2n+1}]$ is given by the same expression with b_l replacing a_l. A direct computation yields

$$\langle [U, \tilde{X}_{2n+1}], V \rangle_L = -\sum_{l=1}^{2n} b_l (\tilde{X}_{2n+1} a_l) \quad \text{and} \quad \langle [V, \tilde{X}_{2n+1}], U \rangle = -\sum_{l=1}^{2n} a_l (\tilde{X}_{2n+1} b_l).$$

Thus

$$\langle [U, \tilde{X}_{2n+1}], V\rangle_L + \langle [V, \tilde{X}_{2n+1}], U\rangle_L = -\sum_{l=1}^{2n} \left(a_l \tilde{X}_{2n+1} b_l + b_l \tilde{X}_{2n+1} a_l \right)$$
$$= -\tilde{X}_{2n+1}\langle U, V\rangle_L = 0.$$

The result now follows from the orthogonality of U and V and the Kozul identity. □

Theorem 4.18. *Let $S \subset \mathbb{H}^n$ be a C^2 regular hypersurface. Then*

$$\lim_{L\to\infty} II^L_{ij} = \frac{h_{ij} + h_{ji}}{2} \tag{4.18}$$

for $i, j = 1, \ldots, 2n-1$, at noncharacteristic points.

Proof. Since II^L is symmetric, (4.15) and Proposition 4.16 yield

$$II^L_{ij} = \frac{II^L_{ij} + II^L_{ji}}{2} = \alpha_L \frac{h_{ij} + h_{ji}}{2} + \beta_L \frac{v^L_{ij} + v^L_{ji}}{2} = \alpha_L \frac{h_{ij} + h_{ji}}{2}. \tag{4.19}$$

(4.18) now follows from (4.19) and Lemma 4.14. □

In other words, the second fundamental form II^L converges as $L \to \infty$ to the *symmetrized horizontal second fundamental form*

$$(II^0)^* = (h^*_{ij}),$$

where $h^*_{ij} = \frac{1}{2}(h_{ij} + h_{ji})$. It is now natural to introduce some sub-Riemannian analogs for classical notions of curvature.

Definition 4.19. Let $S \subset \mathbb{H}^n$ be a C^2 regular hypersurface and denote by $(II^0)^*$ its symmetrized horizontal second fundamental form, defined at noncharacteristic points. The *horizontal principal curvatures* of S are $k_i = h_{ii}$, $i = 1, \ldots, 2n-1$, the *horizontal mean curvature* of S is

$$\mathcal{H}_0 = \mathrm{Trace}(II^0)^* = \sum_{i=1}^{2n-1} k_i,$$

and the *horizontal Gauss curvature* of S is $\mathcal{K}_0 = \det(II^0)^*$.

With this notation in place, we can define the analogue of constant mean curvature (and hence minimal) surfaces:

Definition 4.20. A C^2 regular hypersurface $S \subset \mathbb{H}^n$ is called a *horizontal constant mean curvature surface* (CMC) if $\mathcal{H}_0$ is constant along the noncharacteristic locus, and is called a *horizontal minimal surface* if $\mathcal{H}_0 = 0$ along the noncharacteristic locus. We will refer to the class of CMC surfaces with horizontal mean curvature ρ with the notation $CMC(\rho)$.

Lemma 4.5 extends to $\mathbb{H}^n$ as follows:

Corollary 4.21. *Let $S \subset \mathbb{H}^n$ be a C^2 regular hypersurface. Then $\mathcal{H}_L \to \mathcal{H}_0$ at noncharacteristic points.*

Remark 4.22. Observe that

$$\mathcal{H}_L = \mathrm{div}_{g_L}(\nu_L) = \sum_{i=1}^{2n+1} \tilde{X}_i \left(\frac{\tilde{X}_i u}{|\nabla_L u|} \right). \tag{4.20}$$

A direct computation shows that

$$\mathcal{H}_L\,|\nabla_L u| = \sum_{i,j=1}^{2n} \left(\delta_{ij} - \frac{\tilde{X}_i u \tilde{X}_j u}{|\nabla_L u|^2} \right) \tilde{X}_i \tilde{X}_j u.$$

It is clear that both sides of this equation converge (at all points, characteristic or not), yielding

$$\lim_{L\to\infty} \mathcal{H}_L\,|\nabla_L u| = \begin{cases} \mathcal{H}_0 |\nabla_0 u| = \mathcal{L}u - \frac{\mathcal{L}_\infty u}{|\nabla_0 u|^2} & \text{on } S \setminus \Sigma(S), \\ \mathcal{L}u & \text{on } \Sigma(S), \end{cases} \tag{4.21}$$

where $\mathcal{L}u = \sum_{i=1}^{2n} X_i^2 u$ and $\mathcal{L}_\infty u = \sum_{i,j=1}^{2n} X_i u X_j u X_i X_j u$ denote the sub-Laplacian and infinite sub-Laplacian in $\mathbb{H}^n$, respectively. In view of this consideration we can extend the function $\mathcal{H}_0 |\pi_H(\nu_1)|$ from $S \setminus \Sigma(S)$ to all of S as a continuous function. Here we denote by π_H the orthogonal projection of Lie algebra vectors onto the horizontal bundle.

4.3.2 Horizontal second fundamental form and the Legendrian foliation

In this section we want to give a more extrinsic definition of the horizontal second fundamental form and relate it to Legendrian foliations.

Let $S = \{x \in \mathbb{H}^n : u(x) = 0\}$ be a C^2 hypersurface with characteristic set $\Sigma(S)$. For every point $x \in S \setminus \Sigma(S)$ the intersection of the horizontal plane $H(x)$ with the tangent space $T_x S$ defines a $(2n-1)$-dimensional horizontal tangent space $H_x S$; we denote by HS the corresponding *horizontal tangent subbundle.* We recall that ν_H denotes the horizontal normal to S, and choose a g_1-orthonormal frame $\{e_1, \dots, e_{2n-1}\}$ for HS. The vectors $\{e_1, \dots, e_{2n-1}, \nu_H\}$ form a g_1-orthonormal frame for $H\mathbb{H}^n|_S$. For any $x \in S$, let $\Pi_i(x)$ be the 2-plane spanned by the vectors $e_i(x)$ and $\nu_H(x)$, and define the curve $\gamma_{i,x} = S \cap \Pi_i(x)$ with $\gamma_{i,x}(0) = x$ and $\gamma'_{i,x}(0) = e_i(x)$. Note that $\gamma'_{i,x}$ is not necessarily horizontal away from zero. Let us parametrize the horizontal component of $\gamma_{i,x}$ by arc length in the g_1-metric, i.e., we require that $\langle \pi\gamma'_{i,x}, \pi\gamma'_{i,x} \rangle_1 = 1$.

Proposition 4.23. *Let $S \subset \mathbb{H}^n$ be a C^2 regular hypersurface and denote by (h_{ij}) its horizontal second fundamental form as defined in* (4.16). *Then*

$$h_{ij}(x) = \left\langle \frac{d}{ds}\nu_H(\gamma_{i,x}(s))\Big|_{s=0}, e_j \right\rangle_1 \tag{4.22}$$

at noncharacteristic points. In particular,

$$\mathcal{H}_0 = \sum_{i=1}^{2n-1} k_i = \sum_{i=1}^{2n-1} \left\langle \frac{d}{ds}\nu_H(\gamma_{i,x}(s))\Big|_{s=0}, e_i \right\rangle_1. \tag{4.23}$$

This follows immediately from the definition of h_{ij}.

In $\mathbb{H}$, HS is a line bundle, and the corresponding flow lines are Legendrian curves $\gamma_{1,x} = \gamma = (\gamma_1, \gamma_2, \gamma_3)$. We call this family of curves the *Legendrian foliation* of S. As above we assume that (γ_1, γ_2) is parameterized by arc length. Since the metric induced by g_1 on HS is the pull-back of the usual Euclidean metric in the plane, it is easy to see that $\vec{n} = \pi\nu_H$ is a unit normal for the planar curve (γ_1, γ_2), and $\pi e_1 = (\gamma_1', \gamma_2')(0) = \mathbf{i}\,\vec{n}$. In terms of a defining function u for S,

$$\gamma' = (\overline{q}X_1 - \overline{p}X_2)\circ\gamma = (X_2uX_1 - X_1uX_2)|\nabla_0 u|^{-1}\circ\gamma. \tag{4.24}$$

The second fundamental form $II = (II_{11})$ takes a very simple form

$$II_{11} = \left\langle \frac{d}{ds}\nu_H(\gamma_{1,x}(s))\Big|_{s=0}, \gamma_{1,x}'(0) \right\rangle_1 = \left\langle (\vec{n}\circ\pi\circ\gamma_{1,x})'(0), \mathbf{i}\,(\vec{n}\circ\pi)(x) \right\rangle = k,$$

where we have denoted by k the Euclidean curvature of the planar curve γ.

Alternatively, we can follow a more explicit approach: The curvature vector $\vec{k} = k\,\mathbf{i}\,(\gamma_1', \gamma_2')$ of (γ_1, γ_2) is given by

$$\begin{aligned}
\vec{k} &= -(\gamma_1'', \gamma_2'') = -\left(\left(\frac{X_2u}{|\nabla_0 u|}, \frac{-X_1u}{|\nabla_0 u|}\right)\circ\gamma\right)' \\
&= \frac{1}{|\nabla_0 u|}\left((X_2uX_1 - X_1uX_2)\left(\frac{-X_2u}{|\nabla_0 u|}\right) + (X_2uX_1 - X_1uX_2)\left(\frac{X_1u}{|\nabla_0 u|}\right)\right)\circ\gamma \\
&= \frac{(X_2u)^2X_1X_1u - X_1uX_2u(X_1X_2u + X_2X_1u) + (X_1u)^2X_2X_2u}{|\nabla_0 u|^4}(X_1u, X_2u)\circ\gamma \\
&= \left(\mathcal{L}u - \frac{\mathcal{L}_\infty u}{|\nabla_0 u|^2}\right)\frac{(X_1u, X_2u)}{|\nabla_0 u|^2}\circ\gamma \\
&= \mathcal{H}_0\pi\nu_H\circ\gamma.
\end{aligned}$$

In conclusion, we have proved the following:

Proposition 4.24. *Let S be a $C^{1,1}$ surface in $\mathbb{H}$, and let $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ be a curve in the Legendrian foliation of $S \setminus \Sigma(S)$. Then the curvature k of γ at $(\gamma_1(t), \gamma_2(t))$ equals the horizontal mean curvature $\mathcal{H}_0$ of S at $(\gamma_1(t), \gamma_2(t), \gamma_3(t))$.*

Remark 4.25. If we denote by $(\cos\theta(s), \sin\theta(s))$ the value of $\pi\nu_H$ at the point corresponding to $\gamma(s)$, then $(\gamma_1'(s), \gamma_2'(s)) = (\sin\theta(s), -\cos\theta(s))$ and

$$\mathcal{H}_0 \pi\nu_H = -(\gamma_1'', \gamma_2'')(s) = -\theta'(s)\pi\nu_H,$$

hence $\mathcal{H}_0 = -\theta'(s)$.

Remark 4.26. Proposition 4.24 can be used to give an alternate derivation of the examples in Section 4.3. For example, the fact that the Legendrian foliation of the bubble sets $\mathcal{B}(o,R)$ consists of horizontal lifts of circles of diameter R guarantees that the horizontal mean curvature is constantly equal to $2/R$.

The Legendrian foliation of the CC sphere $\partial B_{cc}(o,R) = \{x = (z,x_3) : d(x,o) = R\}$ can also be explicitly computed. Recalling that $\partial B_{cc}(o,R)$ is given in parametric form as $|z| = (2/c)\sin(cR/2)$, $x_3 = (cR - \sin(cR))/(2c^2)$, it is simple to verify that the curve

$$\gamma_\theta(s) = \left(\frac{2}{s}\sin(sR/2)e^{\mathbf{i}\,\psi(sR/2)+\mathbf{i}\,\theta}, (sR - \sin(sR))/(2s^2)\right)$$

(where ψ is determined by the condition $\psi'(u) = \cot u/u - \cot^2 u$), lies on the surface of the CC sphere $\partial B_{cc}(o,R)$ and is horizontal. Computing the standard curvature of the planar projection $\pi\gamma_\theta$ at the point $\pi\gamma_\theta(c)$ reproduces the formula

$$\mathcal{H}_0 = \frac{1}{2}\cdot\frac{c/2}{\sin(cR/2)}\cdot\frac{\sin(cR) - cR\cos(cR)}{\sin(cR/2) - (cR/2)\cos(cR/2)}$$

from Example 4.12.

The case of the Heisenberg cone $P_\alpha = \{(z,x_3) : x_3 = \alpha|z|^2\}$ is also interesting. For fixed $y = (w, y_3) \in P_\alpha$ with $w \neq 0$, the curve

$$\gamma_y(s) = \left(swe^{4\,\mathbf{i}\,\alpha\log s}, \frac{1}{4}s^2 y_3\right) \tag{4.25}$$

lies on the surface P_α and is horizontal. Again, the horizontal mean curvature of P_α, computed in Example 4.10, can be reproduced by computing the curvature of the projection $\pi\gamma_y$. This example will resurface in the context of the *horizontal polar coordinate decomposition* of $\mathbb{H}$ in Section 5.4.

4.4 Analysis at the characteristic set and fine regularity of surfaces

In this chapter, we have investigated submanifolds in the Heisenberg group using Riemannian submanifold analysis on a sequence of Riemannian approximants. We remind the reader that all of the computations up to this point have had a common assumption: we consider only noncharacteristic points of a smooth submanifold. This assumption is necessary as a quick perusal of the previous computations shows that even if all of the quantities are well defined at characteristic points in the approximants (which is not necessarily the case), they may degenerate badly in the limit. However, this is only symptomatic of the true problem: characteristic points behave like points of low (or no) regularity with respect to the sub-Riemannian structure. To illustrate this, we will investigate two elementary examples.

Example 4.27. Consider one of the simplest surfaces in the Heisenberg group, the plane $S = \{x_3 = 0\}$. The unit horizontal normal to this plane is given by

$$\nu_H = -\frac{x_2}{\sqrt{x_1^2 + x_2^2}}\, X_1 + \frac{x_1}{\sqrt{x_1^2 + x_2^2}}\, X_2.$$

Note that ν_H is not well defined at o and that this point is the only characteristic point of S.

Considering $\{X_1, X_2\}$ as a basis for a two-dimensional vector space, we can visualize this vector field as a vector field on $\mathbb{R}^2 \setminus \{(0,0)\}$. Note that we cannot define a reasonable value for ν_H at $(0,0)$ as the limit does not exist. It is in this sense that the plane has a non-smooth point with respect to the sub-Riemannian structure at the origin; its horizontal normal has a nonremovable discontinuity. We note in passing this vector field has much in common with the (Euclidean) normal field of the helicoid, $\{x_3 = \arctan(x_2/x_1)\}$, given by

$$\left(\frac{-x_2}{x_1^2 + x_2^2}, \frac{x_1}{x_1^2 + x_2^2}, -1 \right).$$

Example 4.28. Next, we consider another simple example, the surface $S = \{x_3 = \frac{x_1 x_2}{2}\}$. The unit horizontal normal to this surface is

$$\nu_H = \operatorname{sign}(x_2)\, X_1$$

Here, the characteristic locus is given by $x_2 = 0$ and, again, we observe that the unit horizontal normal is discontinuous along this line.

We note that this surface has piecewise constant horizontal normal and hence, from a sub-Riemannian point of view, is analogous to the union of pieces of two different planes in $\mathbb{R}^3$. Thus, in this case, the characteristic locus behaves similarly to a locus of discontinuity of a surface in $\mathbb{R}^3$.

Even from these two simple examples, we see that characteristic points introduce a number of serious pathologies in the study of submanifold geometry. Near characteristic points a surface S is locally in the form of a so-called "t-graph" $x_3 = u(x_1, x_2)$. In this section we study the behavior of the horizontal normal and Legendrian foliation near the characteristic locus for surfaces in $\mathbb{H}$. The descriptions which we obtain will be of key importance in later chapters, see specifically 8.5 for an application to Pansu's isoperimetric problem.

Without loss of generality we may assume that in a neighborhood of each characteristic point the surface is a graph over the z-plane:

$$\mathcal{G}_u = \{(z, u(z)) : z \in \Omega\} \tag{4.26}$$

where u is a C^2 function defined on an open set $\Omega \subset \mathbb{R}^2$.

First we set some notation. For u and $\mathcal{G}_u$ as in (4.26) we define the *singular set* of u,

$$S_u = \{z \in \Omega : \nabla_0 U(z, \cdot) = 0\} = \pi(\Sigma(\mathcal{G}_u)), \tag{4.27}$$

where $U(z, x_3) = u(z) - x_3$ and π denotes the projection from $\mathbb{H}$ to the z-plane. For every curve $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ in the Legendrian foliation of $\mathcal{G}_u$ we will call its projection (γ_1, γ_2) a *characteristic curve*. We will denote by

$$D_0^2 U = \begin{pmatrix} X_1^2 U & X_1 X_2 U \\ X_2 X_1 U & X_2^2 U \end{pmatrix} = \begin{pmatrix} \partial^2_{x_1} u & \partial_{x_1,x_2} u - \frac{1}{2} \\ \partial_{x_1,x_2} u + \frac{1}{2} & \partial^2_{x_2} u \end{pmatrix}$$

the *horizontal Hessian* of U. Note that this is always a non-zero matrix as the two entries on the anti-diagonal cannot vanish simultaneously. For each $a = (a_1, a_2) \in \mathbb{S}^1$ we set

$$F_a(z) = a_1 X_1 U(z, \cdot) + a_2 X_2 U(z, \cdot) = a_1 \Big(\partial_{x_1} u - \frac{x_2}{2} \Big) + a_2 \Big(\partial_{x_2} u + \frac{x_1}{2} \Big),$$

and

$$\gamma_a = \Big\{ z \in \Omega : F_a(z) = 0 \Big\} = S_u \cup \Big\{ z \in \Omega : \nu_H(z, u(z)) \perp a \Big\}, \tag{4.28}$$

where ν_H denotes the horizontal normal to $\mathcal{G}_u$. In other words, γ_a contains all (projections of) characteristic points plus points at which the Legendrian foliation is tangent to a. A simple computation shows that

$$\nabla F_a = a D_0^2 U(z, \cdot). \tag{4.29}$$

We treat the cases of non-isolated and isolated points in $\Sigma(\mathcal{G}_u)$ separately.

4.4.1 The Legendrian foliation near non-isolated points of the characteristic locus

First, we treat the case of non-isolated points in the characteristic locus.

Lemma 4.29. *Each $p \in S_u$ has a neighborhood U_p such that $S_u \cap U_p$ lies in a C^1 curve of the form γ_a for some $a \in \mathbb{S}^1$. Moreover, if p is not isolated in S_u, then* $\det D_0^2 U(p, \cdot) = 0$.

Proof. Set $A = D_0^2 U(z, \cdot)$. In view of (4.4) there is at most one $c \in \mathbb{S}^1$ (modulo choice of sign) for which $cA = 0$. From (4.29) we deduce that $\nabla F_a(p) \neq 0$. The implicit function theorem then implies that γ_a is a C^1 curve in a neighborhood U_p of p, proving the first claim.

If p is not isolated in S_u, choose $\{p_j\} \subset S_u \subset \gamma_a$ with $\lim_{j \to \infty} p_j = p$. Let γ_a be parametrized by arc length with $\gamma_a(0) = p$ and choose $s_j \in \mathbb{R}$ such that $\gamma_a(s_j) = p_j$. By passing to a subsequence if necessary, we may assume that (s_j) is increasing. Since $\nabla_0 U(p_j,\cdot)=0$, the mean value theorem ensures that there exist numbers $\bar{s}_j, \bar{t}_j$ in the interval (s_j, s_{j+1}) such that

$$\frac{d}{ds}X_1U(\gamma_a(s),\cdot)\bigg|_{s=\bar{s}_j} = \frac{d}{ds}X_2U(\gamma_a(s),\cdot)\bigg|_{s=\bar{t}_j} = 0. \tag{4.30}$$

Letting $j\to\infty$ and using (4.29) and the smoothness assumption on u, we have

$$\gamma_a'(0)\cdot A = 0. \tag{4.31}$$

Since $\gamma_a'(0)$ is a unit vector, we conclude that $\det A = 0$. □

Remark 4.30. In fact p is a non-isolated characteristic point if and only if $D_0^2U(p,\cdot)$ has zero determinant.

With additional assumptions on the blow-up of the horizontal mean curvature near $\Sigma(\mathcal{G}_u)$ one can show that S_u is not only contained in a C^1 curve, but, near non-isolated characteristic points, is precisely a C^1 curve.

Theorem 4.31 (Cheng–Hwang–Malchiodi–Yang). *Let $p\in S_u$ be (the projection of) a non-isolated characteristic point. If the horizontal mean curvature $\mathcal{H}_0$ of $\mathcal{G}_u$ satisfies*

$$|\mathcal{H}_0(z,u(z))| = O(\operatorname{dist}(p,z)^{-1})$$

as $z\notin S_u$ approaches p, then there exists a neighborhood U_p of p such that $S_u\cap U_p$ is exactly a C^1 curve γ_a, for some $a\in\mathbb{S}^1$.

Proof. By Remark 4.30, $\det A=0$ for $A=D_0^2U(p,\cdot)$. Since $A\neq 0$ there exists a unique vector $c\in\mathbb{S}^1$ (modulo sign) such that $cA=0$. For all distinct $a,b\in\mathbb{S}^1\setminus\{\pm b\}$, $\nabla F_a(p)$ is a non-zero multiple of $\nabla F_b(p)$. By the Implicit Function Theorem, p has a neighborhood U_p containing a pair of C^1 curves γ_a,γ_b, defined as in (4.28), both passing through the point p, with a common tangent at p, and coinciding exactly on $S_u\cap U_p$. Again we assume that these curves are parameterized by arc length. Denote by s,t the arc length parameters on γ_a,γ_b. Observe that the distance from p is a monotone increasing function for $s,t>0$, and is monotone decreasing for $s,t<0$. The curves γ_a and γ_b are formed as the union of disjoint arcs $\gamma_a|(s_j^\pm,\tilde{s}_j^\pm)$ and $\gamma_b|(t_j^\pm,\tilde{t}_j^\pm)$ in $\Omega\setminus S_u$ and arcs $\gamma_a|[\tilde{s}_j^+,s_{j-1}^+]=\gamma_b|[\tilde{t}_j^+,t_{j-1}^+]$ and $\gamma_a|[\tilde{s}_j^-,s_{j+1}^-]=\gamma_b|[\tilde{t}_j^-,t_{j+1}^-]$ in S_u. More precisely,[2] for $s_i^-<\tilde{s}_i^-<0$, $\tilde{s}_i^-\le s_{i+1}^-$, and $0<s_i^+<\tilde{s}_i^+\le s_{i-1}$ (and similarly for the t parameter), we have

$$(\gamma_a\cap U)\setminus S_u = \bigcup_{j=1}^\infty \gamma_a|(s_j^+,\tilde{s}_j^+)\cup\gamma_a|(s_j^-,\tilde{s}_j^-)$$

and

$$(\gamma_b\cap U)\setminus S_u = \bigcup_{j=1}^\infty \gamma_b|(t_j^+,\tilde{t}_j^+)\cup\gamma_b|(t_j^-,\tilde{t}_j^-).$$

[2] One allows for equality $\tilde{s}_i^\pm = s_{i\pm1}^\pm$ in case there are only a finite number of arcs.

We will consider $s, t > 0$, the remaining case being dealt with in similar fashion. For $i > 1$, set $p_i = \gamma_a(s_i^+) = \gamma_b(t_i^+)$ and $\tilde{p}_i = \gamma_a(\tilde{s}_i^+) = \gamma_b(\tilde{t}_i^+)$.[3]

Next we show that γ_a and γ_b must meet in a sequence of points converging to p. If not, the two curves are disjoint in $B(p,\epsilon) \setminus \{p\}$ for some $\epsilon > 0$. Consider a sequence $r_i \to 0$ and the sets Ω_i whose boundary is formed by portions of γ_a, γ_b (with $s, t > 0$) and $\partial B(p, r_i)$. Such sets are contained in *fan-shaped* regions with vertex p and aperture $\theta_i \to 0$ (since $\gamma_a'(0) = \gamma_b'(0)$). If we denote by $\vec{n}_i$ the outer normal to Ω_i then

$$\int_{\partial\Omega_i} \nu_H \cdot \vec{n}_i \, ds = \int_{\Omega_i} \operatorname{div} \nu_H \, dx_1 dx_2 = \int_{\Omega_i} \mathcal{H}_0 \, dx_1 dx_2 \le C\theta_i r_i. \tag{4.32}$$

On the other hand, along γ_a (resp. γ_b) the vector ν_H is constantly equal to $\nu_H(a)$ (resp. $\nu_H(b)$), and orthogonal to a (resp. b). The vector $\vec{n}_i$ converges (in the C^1 norm on Ω_i) to a vector $\vec{n}(p)$ perpendicular to $\gamma_a'(0)$. Since $a \ne b$, $c_1 = \nu_H(a) \cdot \vec{n}(p)$ and $c_2 = \nu_H(b) \cdot \vec{n}(p)$ are not equal. Then there exist $0 < \delta_i = o(1)$ and $C > 0$ such that the estimate

$$\left| \int_{\partial\Omega_i} \nu_H \cdot \vec{n}_i \, ds \right| \ge C(|c_1 - c_2| - \delta_i) r_i \tag{4.33}$$

holds for large i, reaching a contradiction with (4.32) in the limit as $i \to \infty$.

Next we show that, although γ_a and γ_b intersect in arbitrarily small neighborhoods of p, the sequence $\{p_i\}$ does not converge to p as $i \to \infty$. Assuming this fact temporarily, we conclude the proof. Indeed, since such points mark the arcs in which $\gamma_a \not\subset S_u$, we have $S_u \cap B(p,\epsilon) \supset \gamma_a([0, \bar{s}^+))$ for some $\bar{s}^+$ and some small $\epsilon > 0$. This suffices to complete the proof.

Suppose that $p_i \to p$. Consider a sequence of regions Ω_i surrounded by γ_a and γ_b from p_i to $\tilde{p}_i$. As in the previous construction, the Ω_i are also contained in *fan-shaped* regions with vertex p and aperture $\theta_i \to 0$. If we denote as before by $\vec{n}_i$ the outer normal to Ω_i then we easily obtain the estimate

$$\begin{aligned} \int_{\partial\Omega_i} \nu_H \cdot \vec{n}_i \, ds = \int_{\Omega_i} \operatorname{div} \nu_H \, dx_1 dx_2 &= \int_{\Omega_i} \mathcal{H}_0 \, dx_1 dx_2 \\ &\le C\theta_i \big| |p_i - p| - |\tilde{p}_i - p| \big|. \end{aligned} \tag{4.34}$$

Arguing as above, and for c_1, c_2, C, δ_i defined earlier, we estimate

$$\left| \int_{\partial\Omega_i} \nu_H \cdot \vec{n}_i \, ds \right| \ge \frac{|c_1 - c_2|}{2} \big| |p_i - p| - |\tilde{p}_i - p| \big| \tag{4.35}$$

for large i, thus reaching a contradiction with (4.34) in the limit as $i \to \infty$. Modulo the earlier discussion, this completes the proof of Theorem 4.31. □

[3] In the case $i = 1$ one may possibly have $\gamma_a(s_i^+) \ne \gamma_b(t_i^+)$ and we set $p_1 = \gamma_a(s_1^+)$ and $\tilde{p}_1 = \gamma_a(\tilde{s}_1^+)$.

Next, let $\gamma \subset S_u$ be a C^1 curve and let $p \in \gamma$ be the projection of a characteristic point. We assume that for $\epsilon > 0$ sufficiently small the curve γ divides $B(p, \epsilon)$ in two disjoint, open, connected, noncharacteristic components $B^{\pm}$. This assumption is satisfied if, for instance, the hypotheses of Theorem 4.31 hold.

Lemma 4.32. *The limits*

$$\nu_H(p^{\pm}) = \lim_{q \to p, q \in B^{\pm}} \nu_H(q)$$

exist and $\nu_H(p^+) = -\nu_H(p^-)$.

Proof. Let $\gamma = (\gamma_1, \gamma_2)$ be parametrized by arclength with $\gamma(0) = 0$. Without loss of generality we may assume (by rotating and translating the plane) that the point $p = (0,0)$ and that the x_1 axis is transversal to γ: $d/ds\gamma_2(0) \neq 0$. By the Implicit Function Theorem, γ can be written as a graph $(f(x_2), x_2)$ in a small neighborhood of $(0,0)$. Applying (4.31) with γ in place of γ_a, we immediately see that either $\partial^2_{x_1} u(0,0) \neq 0$ or $\partial_{x_1, x_2} u(0,0) + \frac{1}{2} \neq 0$.

Assume first that $\partial^2_{x_1} u(0,0) \neq 0$. For $z = (x_1, x_2) \notin S_u$ and for some $\bar{x}_1, \tilde{x}_1$ between x_1 and $f(x_2)$ the mean value theorem implies that

$$\begin{aligned}
\frac{\partial_{x_2} u(x_1, x_2) + \frac{x_1}{2}}{\partial_{x_1} u(x_1, x_2) - \frac{x_2}{2}} &= \frac{\partial_{x_2} u(x_1, x_2) + \frac{x_1}{2} - \partial_{x_2} u(f(x_2), x_2) - \frac{f(x_2)}{2}}{\partial_{x_1} u(x_1, x_2) - \frac{x_2}{2} - \partial_{x_1} u(f(x_2), x_2) + \frac{x_2}{2}} \\
&= \frac{(x_1 - f(x_2))\big(\partial_{x_1, x_2} u(\bar{x}_1, x_2) + \frac{1}{2}\big)}{(x_1 - f(x_2))\partial^2_{x_1} u(\tilde{x}_1, x_2)}.
\end{aligned}$$

Hence we obtain the existence and equality of the limits

$$\begin{aligned}
\lim_{\substack{(x_1,x_2) \to (0,0) \\ (x_1,x_2) \in B^+}} \frac{\partial_{x_2} u(x_1, x_2) + \frac{x_1}{2}}{\partial_{x_1} u(x_1, x_2) - \frac{x_2}{2}} &= \lim_{\substack{(x_1,x_2) \to (0,0) \\ (x_1,x_2) \in B^-}} \frac{\partial_{x_2} u(x_1, x_2) + \frac{x_1}{2}}{\partial_{x_1} u(x_1, x_2) - \frac{x_2}{2}} \\
&= \frac{\partial_{x_1}\partial_{x_2} u(0,0) + \frac{1}{2}}{\partial^2_{x_1} u(0,0)}.
\end{aligned} \tag{4.36}$$

The latter implies the existence of $\nu_H(p^{\pm})$ and that these two values may differ at most by their sign. To establish that the sign difference is -1, we need only observe that for $|x_1|$ sufficiently small, and for some $\bar{x}_1$ between x_1 and 0 one has

$$\partial_{x_1} u(x_1, 0) - x_2/2 = \partial^2_{x_1} u(\bar{x}_1, 0) x_1.$$

Hence $\partial_{x_1} u(x_1, 0) - x_2/2$ and $\partial^2_{x_1} u(\bar{x}_1, 0)$ have the same sign in one component (say B^+) and opposite sign in the other.

In case $\partial_{x_1}\partial_{x_2} u(o) + \frac{1}{2} \neq 0$ we argue as above, using the limit

$$\lim_{\substack{(x_1,x_2) \to (0,0) \\ (x_1,x_2) \in B^+}} \frac{\partial_{x_1} u(x_1, x_2) - \frac{x_2}{2}}{\partial_{x_2} u(x_1, x_2) + \frac{x_1}{2}} = \frac{\partial^2_{x_1} u(0,0)}{\partial_{x_1}\partial_{x_2} u(0,0) + \frac{1}{2}}$$

as starting point and arrive at the same conclusion. □

Remark 4.33. Using l'Hopital's rule in the second variable we obtain that $\nu_H(p^+)$ is parallel to $(\partial^2_{x_1,x_2}u - \frac{1}{2}, \partial^2_{x_2}u)$ provided the latter vector is non-zero.

Lemma 4.32 has an important corollary.

Proposition 4.34. *In the hypothesis and notation of Lemma* 4.32 *there exists a unique C^1 curve $\tilde{\gamma}$ in a neighborhood of p, passing through p and transversal to γ such that $\tilde{\gamma} \cap B^\pm$ are projections of curves in the Legendrian foliation of $\mathcal{G}_u$, i.e., characteristic curves.*

The curve $\tilde{\gamma}$ in the previous statement can be decomposed as

$$\tilde{\gamma} = \{p\} \cup \tilde{\gamma}^+ \cup \tilde{\gamma}^-,$$

where $\tilde{\gamma}^+ := \tilde{\gamma} \cap B^+$ and $\tilde{\gamma}^- := \tilde{\gamma} \cap B^-$. We will postpone the proof of Proposition 4.34 and start by observing that the uniqueness of $\tilde{\gamma}^\pm$ and their transversality to γ follow from the next lemma.

Lemma 4.35. *In the hypothesis and the notation of Lemma* 4.32*,*

(i) $\nu_H^\perp(p^+) D_0^2 U(p,\cdot) = 0$, *and*
(ii) $z(D_0^2 U(p,\cdot))^T = 0$ *for all non-zero vectors z tangent to γ at p.*

Proof. Assume $\gamma(0) = p$ and $\gamma'(0) = z$. Since $\nabla_0 U = 0$ along γ, differentiating in s and evaluating at $s = 0$ yields (ii). On the other hand, from the proof of Lemma 4.32, in particular (4.36), we see $\nu_H(p^+)$ is orthogonal to both $(\partial^2_{x_1}u, \partial^2_{x_1,x_2}u + \frac{1}{2})$ and $(\partial^2_{x_1,x_2}u - \frac{1}{2}, \partial^2_{x_2}u)$ provided these are not zero. Since these vectors form the columns of $D_0^2U(p,\cdot)$, (i) follows immediately from the previous observation. $\square$

Observe that $\nabla_0 U(p,\cdot) = (\partial_{x_1}u + \frac{x_2}{2}, \partial_{x_2}u - \frac{x_1}{2}) = 0$ for $p \in S_u$, and hence

$$\begin{aligned} \partial_{x_1}u + \frac{x_2}{2}\Big|_q &= \partial^2_{x_1}u(p)\Delta x + \Big(\partial^2_{x_1,x_2}u(p) + \frac{1}{2}\Big)\Delta y + o(\Delta s), \\ \partial_{x_2}u - \frac{x_1}{2}\Big|_q &= \Big(\partial^2_{x_1,x_2}u(p) - \frac{1}{2}\Big)\Delta x + \partial^2_{x_2}u(p)\Delta y + o(\Delta s), \end{aligned} \tag{4.37}$$

for $q \notin S_u$ near p, where we have let $q - p = (\Delta x, \Delta y)$ and $\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Let $A = D_0^2 U(p,\cdot)$ as before and observe that

$$A - A^T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We are now ready to present the proof of Proposition 4.34.

Proof. First we show that $\nu_H^\perp(p^+)$ is not tangent to γ at p. If so then Lemma 4.35 would imply $\nu_H = \nu_H^\perp(p^+)(A - A^T) = 0$ which is a contradiction. In fact, this argument leads to the identity

$$\left|\nu_H^\perp(p^+) A^T\right| = 1. \tag{4.38}$$

Next we prove uniqueness of the characteristic curve $\tilde{\gamma}^+$. Assume we have two distinct characteristic curves $\tilde{\gamma}_1^+$ and $\tilde{\gamma}_2^+$ not intersecting outside of p (which belongs to their closure) and with common tangent $\nu_H^\perp(p^+)$ at p. With $r = \Delta s$ consider the region Ω_r surrounded by $\tilde{\gamma}_1^+$, $\tilde{\gamma}_2^+$ and by ∂B_r. Such a region is contained in fan-shaped regions centered at p with aperture $\theta_r \to 0$. Letting $\Gamma_r = \partial\Omega_r \cap \partial B_r$ we note that the arc length $|\Gamma_r|$ satisfies

$$|\Gamma_r| \leq r\theta_r. \tag{4.39}$$

Following (4.37) we observe that

$$\nabla_0 U = (\Delta x, \Delta y)A^T + o(r),$$

while

$$(\Delta x, \Delta y) = r\nu_H^\perp(p^+) + o(r).$$

If we denote by $\vec{n}_r$ the outer normal to $\partial\Omega_r$, then

$$\lim_{r\to 0} \vec{n}_r = \lim_{r\to 0} r^{-1}(\Delta x, \Delta y) = \nu_H^\perp(p^+) \quad \text{along} \quad \Gamma_r.$$

Observe that $\vec{n}_r \perp \nu_H^\perp$ along γ_1 and γ_2. In view of (4.38),

$$\begin{aligned} g(r) &:= \int_{\partial\Omega_r} (\nabla_0 U)^\perp \cdot \vec{n}_r \, ds \qquad (4.40)\\ &= \int_{\Gamma_r} (\nabla_0 U)^\perp \cdot \vec{n}_r \, ds = -\int_{\Gamma_r} |\nabla_0 U| \nu_H^\perp(p^+) \cdot \vec{n}_r \, ds = (-r + o(r))|\Gamma_r|. \end{aligned}$$

On the other hand, the divergence theorem implies

$$g(r) = \int_{\Omega_r} \operatorname{div}(\nabla_0 U)^\perp \, dx_1 dx_2 = -\int_0^r |\Gamma_s| \, ds. \tag{4.41}$$

From (4.40) and (4.41) we obtain the ODE $(g'/g)(r) = (1/r) + o(1/r)$ whose solution $g(r) = cr^2 + o(r^2)$ contradicts (4.39). Thus $\tilde{\gamma}_1^+ = \tilde{\gamma}_2^+$. A similar argument yields the uniqueness and non-transversality of $\tilde{\gamma}^-$. □

4.4.2 The Legendrian foliation near isolated points of the characteristic locus

Next, we turn our attention to isolated characteristic points of t-graphs. We will show that every characteristic curve intersecting a small neighborhood of the projection of such a point p will reach p in finite time. Moreover, to every tangent direction $a \in T_{(p,u(p))}\mathcal{G}_u$ there corresponds one and only one curve of the Legendrian foliation tangent to a at p.

For $u \in C^2(\Omega)$ we define the vector field

$$\mathcal{T}(x_1, x_2) = \left(\partial_{x_2} u - \frac{x_1}{2}, -\partial_{x_1} u - \frac{x_1}{2}\right) = (\nabla_0 U)^\perp(x_1, x_2, \cdot).$$

Note that $\mathcal{T} = 0$ on S_u and that the projections of curves in the Legendrian foliation are tangent to $\mathcal{T}$ in $\Omega \setminus S_u$. We will also consider the differential of $\mathcal{T}$:

$$d_p\mathcal{T} = \begin{pmatrix} \partial^2_{x_1,x_2}u(p) - \frac{1}{2} & \partial^2_{x_2}u(p) \\ -\partial^2_{x_1}u(p) & -\partial^2_{x_1,x_2}u(p) - \frac{1}{2} \end{pmatrix}. \tag{4.42}$$

Lemma 4.36. *Assume $p \in S_u$ is an isolated (projection of a) characteristic point, and $|\mathcal{H}_0(z, u(z))| = o(\mathrm{dist}(p, z)^{-1})$. Then*

(i) $d_p\mathcal{T} = \begin{pmatrix} -1/2 & 0 \\ 0 & -1/2 \end{pmatrix}$, *and*

(ii) $\mathrm{Index}_p\,\mathcal{T} = 1$.

Proof. For $q \in \Omega \setminus \{p\}$ set $q - p = (\Delta x, \Delta y)$ and $\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ as before, and observe that (4.37) holds. Set $a = \partial^2_{x_1,x_2}u(p) - \frac{1}{2}$, $b = \partial^2_{x_2}u(p)$, $c = \partial^2_{x_1}u(p)$ and $d = \partial^2_{x_1,x_2}u(p) + \frac{1}{2}$. Since p is isolated, in view of Remark 4.30, we have $bc - ad \neq 0$. Next, we recall from (4.12) that

$$\begin{aligned} \mathcal{H}_0 &= \frac{1}{|\nabla_0 U|}\left(\mathcal{L}U - \frac{\mathcal{L}_\infty U}{|\nabla_0 U|^2}\right) \qquad (4.43) \\ &= \frac{(\partial^2_{x_2}u)(\partial_{x_1}u + \frac{x_2}{2})^2 + (\partial^2_{x_1}u)(\partial_{x_2}u - \frac{x_1}{2})^2 - 2(\partial^2_{x_1,x_2}u)(\partial_{x_1}u + \frac{x_2}{2})(\partial_{x_2}u - \frac{x_1}{2})}{|\nabla_0 U|^3}. \end{aligned}$$

Substituting (4.37) in (4.43) we obtain

$$\mathcal{H}_0 = \frac{(bc - ad)\left((\Delta x, \Delta y)\begin{pmatrix} c & a \\ d & b \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}\right) + o((\Delta s)^2)}{\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}\right|^3 + o((\Delta s)^3)}. \tag{4.44}$$

First, we will show that $c = 0$. If not, then setting $\Delta y = 0$ in (4.44) gives

$$\mathcal{H}_0 = \frac{(bc - ad)c(\Delta x)^2 + o((\Delta s)^2)}{(\Delta x)^3(a^2 + c^2)^{3/2} + o((\Delta s)^3)} \approx \frac{(bc - ad)c}{(a^2 + c^2)^{3/2}}(\Delta x)^{-1},$$

which contradicts the hypothesis $|\mathcal{H}_0| = o(\mathrm{dist}(p, x)^{-1})$. Setting $\Delta x =$ and arguing in the same fashion gives $b = 0$. Next, set $b = c = 0$ in (4.44) to obtain

$$\mathcal{H}_0 = -ad\frac{(a + d)\Delta x\Delta y + o((\Delta s)^2)}{(a^2(\Delta x)^2 + d^2(\Delta y)^2)^{3/2} + o((\Delta s)^3)},$$

from which we deduce that $2\partial^2_{x_1,x_2}u(p) = a + d = 0$. Thus $a = -d$ and $b = c = 0$ which gives (i). To conclude the proof we observe that $\det d_p\mathcal{T} = 1/4 > 0$, whence $\mathrm{Index}_p\,\mathcal{T} = 1$. $\square$

Lemma 4.37. *With the hypotheses of Lemma 4.36 in force, there exists a neighborhood U_p of p such that every characteristic curve γ which intersects $U_p \setminus \{p\}$ reaches p in finite time.*

Proof. Let $q \in \Omega$ and use polar coordinates to represent $q - p = (\Delta x, \Delta y) = (\Delta s)e^{\mathbf{i}\,\phi}$. Let $\alpha, \beta \in \mathbb{R}$ satisfy

$$\nu_H^\perp = \alpha e^{\mathbf{i}\,\phi} + \beta\,\mathbf{i}\,e^{\mathbf{i}\,\phi}.$$

In view of (4.37) and Lemma 4.36(i),

$$\begin{aligned} \partial_{x_1} u + \frac{1}{2} x_2 &= \frac{1}{2}\Delta y + o(\Delta s), \\ \partial_{x_2} u - \frac{1}{2} x_1 &= -\frac{1}{2}\Delta x + o(\Delta s), \\ \nu_H^\perp &= -\Big(\frac{\Delta x}{\Delta s} + o(1), \frac{\Delta x}{\Delta s} + o(1)\Big) \end{aligned} \tag{4.45}$$

as $\Delta s \to 0$. Consequently, $\alpha = \nu_H^\perp \cdot e^{\mathbf{i}\,\phi} = -1 + o(1)$ and $\beta = o(1)$.

Next, consider a curve γ as in the statement of the lemma, and represent it in polar coordinates: $\gamma(t) = \Delta s(t) e^{\mathbf{i}\,\phi(t)}$. Recalling from Section 4.3.2 the identity $\gamma' = -\nu_H^\perp$, we obtain

$$(\Delta s)' e^{\mathbf{i}\,\phi} + (\Delta s)\phi'\,\mathbf{i}\,e^{\mathbf{i}\,\phi} = -\nu_H^\perp = -(1 + o(1))e^{\mathbf{i}\,\phi} + o(1)\,\mathbf{i}\,e^{\mathbf{i}\,\phi}.$$

Thus

$$(\Delta s)' = -1 + o(1) \tag{4.46}$$

so, in sufficiently small neighborhoods of p, $\Delta s(t)$ reaches zero at a finite time $t = T$. □

We now combine all of the preceding work to deduce a structure theorem for the projection of isolated points in the characteristic locus.

Theorem 4.38 (Cheng–Hwang–Malchiodi–Yang). *Assume:*

(H1) *$p \in S_u$ is an isolated (projection of a) characteristic point;*

(H2) *the bound $|\mathcal{H}_0(z, u(z))| = o(\mathrm{dist}(p, z)^{-1})$ holds for $z \notin S_u$ near p;*

(H3) *for some $r_0 > 0$,*

$$\int_0^{r_0} \sup_{z \in \partial B(p,r)} |\mathcal{H}_0(z, u(z))|\, dr < \infty.$$

Then for all $a \in \mathbb{S}^1$ there exists a unique C^1 curve γ_a such that

(i) *γ_a is characteristic, i.e., the projection of a curve in the Legendrian foliation of $\mathcal{G}_u$,*

(ii) *p lies in the closure of γ_a,*

(iii) *$\lim_{q \in \gamma_a, q \to p} \nu_H(q)$ exists and is orthogonal to a.*

Moreover, as a ranges over all of $\mathbb{S}^1$, such curves γ_a cover $U_p \setminus \{p\}$ for some neighborhood U_p of p.

Proof. Let U_p be as in Lemma 4.37. Choose $\delta > 0$ sufficiently small so that $B(p,\delta) \subset U_p$. For each $q \in \partial B(p,\delta)$ denote by γ the unique characteristic curve through q. In view of Lemma 4.37, the curve γ will reach p at a finite time T. Consider a sequence of parameter values $t_j \nearrow T$ and set $q_j = \gamma(t_j)$ so $\lim_{j\to\infty} q_j = p$. Recall from Remark 4.25 that if $\nu_H(\gamma(t)) = \exp(\mathbf{i}\,\theta(t))$ then

$$\mathcal{H}_0(\gamma(t), u(\gamma(t))) = -\theta'(t). \tag{4.47}$$

Denote by θ_j the angle corresponding to q_j, i.e., $\nu_H(q_j, u(q_j)) = (\cos\theta_j, \sin\theta_j)$. Using hypothesis (H3) we will show that $\{\theta_j\}$ is a Cauchy sequence. First, observe that

$$\theta_j - \theta_k = \int_{t_k}^{t_j} \theta'(t)\,dt = \int_{t_k}^{t_j} \mathcal{H}_0\,dt. \tag{4.48}$$

Recall from (4.46) that $(\Delta s)'(t) = -1 + o(1)$ for t near T. Consequently, we can express the parameter t in terms of $r = \Delta s$ and estimate $t'(r) \approx 1$ in a neighborhood of p. Using this observation, letting $r_j = r(t_j)$, $r_k = r(t_k)$ and in view of (4.48) we obtain

$$|\theta_j - \theta_k| \le \int_{r_k}^{r_j} \sup_{z\in\partial B(p,r)} |\mathcal{H}_0(z,u(z))|\,|t'(r)|\,dr \to 0 \qquad \text{as } j,k\to\infty, \tag{4.49}$$

thus proving that (θ_j) is Cauchy. Let us denote by $\theta_{(p,q)}$ its limit as $j \to \infty$. Now we can define a map $\psi : \partial B(p,\delta) \to \mathbb{S}^1$ as follows:

$$\psi(q) = e^{\mathbf{i}\,\theta_{(p,q)}}.$$

To conclude the proof of the theorem it suffices to show that ψ is a homeomorphism.

Step 1 (ψ *is continuous*): Essentially we need to prove a result of C^1 continuity of solutions of a certain ODE with respect to initial data. Consider a sequence of points $q_j \in \partial B(p,\delta)$ converging to $q \in \partial B(p,\delta)$. Denote by $\theta_j = \theta_{(p,q_j)}$ (resp. $\hat\theta = \theta_{(p,q)}$) and by γ_j (resp. $\hat\gamma$) the corresponding characteristic curves joining q_j to p (resp. q to p). We must have that $\theta_j \to \theta_{(p,q)}$. Let ϕ_j be the angle in the polar coordinate representation of $q_j - p$. Without loss of generality we may assume that ϕ_j is strictly decreasing. Since two curves in the Legendrian foliation cannot cross in $B(p,\delta)\setminus\{p\}$, we also have $\theta_j \ge \theta_{j+1}$ for all j. As a monotone and bounded from below sequence, (θ_j) has a limit θ.

We argue by contradiction. If $\theta \ne \theta_{(p,q)}$ then necessarily $\theta > \theta_{(p,q)}$. In this case we find two rays emanating from p and forming an angle less than $\theta - \theta_{(p,q)}$ such that for j sufficiently large, both γ_j and γ avoid a *"fan-shaped"* region $\tilde\Omega$ surrounded by these two rays and $\partial B(p,R)$ for some $R > 0$.

For any point $\tilde p \in \tilde\Omega$ we consider the unique characteristic curve $\tilde\gamma$ joining $\tilde p$ to p. This curve will intersect $\partial B(p,\delta)$ at a point $\tilde q$. Since $\tilde\gamma \cap \tilde\Omega$ does not intersect any γ_j, $\theta_j > \tilde\theta_{(p,\tilde q)}$ and $\tilde q$ must lie in the arc between q_j and q for all j. Thus $\tilde q = q$ and $\tilde\gamma = \hat\gamma$ which is a contradiction.

Step 2 (ψ *is surjective*): If ψ is not surjective then there exists $\tilde\theta$ such that $e^{\mathbf{i}\tilde\theta} \in \mathbb{S}^1 \setminus \psi(\partial B(p,\delta))$. Since the latter set is open, we can find a neighborhood $I_{\tilde\theta}$ of $\tilde\theta$ such that $e^{\mathbf{i}I_{\tilde\theta}}$ is disjoint from the range of ψ. Arguing as in the proof of the continuity of ψ, we deduce the existence of a fan-shaped region $\tilde\Omega$, surrounded by two rays and a portion of a circle $\partial B(p,R)$, which avoids all characteristic curves connecting points $q \in \partial B(p,\delta)$ to p. For any $\tilde p \in \tilde\Omega$ we consider the unique characteristic curve $\tilde\gamma$ through $\tilde p$ and let $\tilde q$ be its intersection with $\partial B(p,\delta)$. Then clearly we must have $e^{\mathbf{i}\theta_{(p,\tilde q)}} \in e^{\mathbf{i}I_{\tilde\theta}} \subset \mathbb{S}^1 \setminus \psi(\partial B(p,\delta))$ which is a contradiction.

Step 3 (ψ *is injective*): Consider $q_1, q_2 \in \partial B_\delta$ distinct such that $\theta_{(p,q_1)} = \theta_{(p,q_2)}$. Denote by γ_i a characteristic curve joining p to q_i. Then the angle between the tangent vectors of γ_1 and γ_2 at p is zero. Consider regions Ω_r surrounded by γ_1, γ_2 and $\partial B(p,r)$. Clearly Ω_r is contained in a fan-shaped region with vertex p and aperture $\theta_r \to 0$. Set $\Gamma_r = \partial\Omega_r \cap \partial B(p,r)$, then

$$|\Gamma_r| \le r\theta_r \tag{4.50}$$

as before. Let $\vec n_r$ denote the outer normal to $\partial\Omega_r$ and observe that $\vec n_r \perp \nu_H^\perp$ along γ_1 and γ_2. Consequently

$$\begin{aligned} g(r) &:= \int_{\partial\Omega_r} (\nabla_0 U)^\perp \cdot \vec n_r \, ds = \int_{\Gamma_r} (\nabla_0 U)^\perp \cdot \vec n_r \, ds \\ &= -\int_{\Gamma_r} (q(s) - p + o(r)) \cdot \frac{q(s)-p}{r} \, ds = (-r + o(r))|\Gamma_r|. \end{aligned}$$

On the other hand, the divergence theorem implies

$$g(r) = \int_{\Omega_r} \operatorname{div}(\nabla_0 U)^\perp \, dx_1 dx_2 = -\int_0^r |\Gamma_s| \, ds. \tag{4.51}$$

From (4.4.2) and (4.51) we obtain the ODE

$$\frac{g'}{g} = \frac{1}{r} + o\Big(\frac{1}{r}\Big),$$

which yields $g(r) = cr^2 + o(r^2)$ for some $c > 0$, in contradiction with (4.50).

Step 4 (ψ^{-1} *is continuous*): We argue by contradiction. Assume there is a sequence $(q_j) \subset \partial B(p,\delta)$ converging to $\tilde q$ so that

$$\theta_j := \theta_{(p,q_j)} \to \theta := \theta_{(p,q)} \tag{4.52}$$

with $q \ne \tilde q \in \partial B(p,\delta)$. Without loss of generality we may assume $\theta_j \ge \theta_{j+1} \ge \theta$. Since $q \ne \tilde q$ we can find $\bar q \in \partial B(p,\delta)$ such that $q \ne \bar q$, $\tilde q \ne \bar q$ and $\theta_j \ge \theta_{(p,\bar q)} \ge \theta$. But then, by (4.52) and the injectivity of ψ we must have $\bar q = q$. With this contradiction we complete the proof of the theorem. $\square$

4.5 Further results: intrinsically regular surfaces and the Rumin complex

Submanifolds in the Heisenberg group (and more general Carnot groups) have also been studied from an intrinsic point of view. This line of investigation has been developed in detail beginning with work of Franchi, Serapioni and Serra-Cassano. These results are somewhat tangential to the theory which we aim to present in this monograph and so we only give a brief summary here and refer the reader to the original papers [106], [107], [108], [109] for a more complete description. We begin with an intrinsic notion of smooth functions on the Heisenberg group.

Definition 4.39. Suppose $U \subset \mathbb{H}$ is an open set. Denote by $C^1_{\mathbb{H}}(U)$ the vector space of continuous functions $f : U \to \mathbb{R}$ so that $\nabla_0 f$ is continuous.

We note that $C^1_{\mathbb{H}}$ is a proper subclass of C^1.

Example 4.40. Let $f(x_1, x_2, x_3) = x_1 - g\left(x_2, \frac{x_1 x_2}{2} + x_3\right)$ where

$$g(a, b) = \frac{|a|^\alpha b}{a^4 + b^2}$$

for $(a, b) \neq (0, 0)$ and $g(0, 0) = 0$. Then $f \in C^1_{\mathbb{H}}$ for $3 < \alpha < 4$ but f is not locally Lipschitz continuous (with respect to the Euclidean metric on $\mathbb{R}^3$) and hence cannot be C^1.

Using this, we define the notion of *intrinsic hypersurface.*

Definition 4.41. $S \subset \mathbb{H}$ is a *codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurface* if for any $p \in S$, there exists a neighborhood $U \subset \mathbb{H}$ of p and $f \in C^1_{\mathbb{H}}(U)$ so that

1. $S \cap U = \{q \in U | f(q) = 0\}$, and
2. $|\nabla_0 f(q)| \neq 0$ for all $q \in U$.

Note that the second condition on f guarantees the absence of characteristic points on the surface.

Before we can continue our discussion of codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurfaces, we need to introduce the notion of intrinsic graph. In order to do so, assume that the Lie algebra of $\mathbb{H}$ has been split as a direct sum: $\mathfrak{h} = \mathfrak{a} \oplus \mathfrak{b}$. Let $A = \exp(\mathfrak{a})$ and $B = \exp(\mathfrak{b})$. We use these sets to foliate $\mathbb{H}$ by setting $P_{\mathfrak{a}}(p) = p \cdot A$ and $P_{\mathfrak{b}}(p) = p \cdot B$ for each $p \in \mathbb{H}$. We can then define an intrinsic notion of graphs in $\mathbb{H}$.

Definition 4.42. A set S in $\mathbb{H}$ is an *intrinsic graph over A along B* if, for every $p \in A$, $S \cap P_{\mathfrak{a}}(p)$ contains at most one point. Equivalently, if there exists $\phi : U \subset B \to A$ so that

$$S = \{p \cdot \phi(p) | p \in U\},$$

we say that S is the (*intrinsic*) *graph* of ϕ.

This notion appears in the following extension of the implicit function theorem to this setting.

Theorem 4.43 (Franchi–Serapioni–Serra-Cassano). *Let $U \subset \mathbb{H}$ be an open set, $f \in C^1_{\mathbb{H}}(U)$, and let*

$$S = \{p \in \mathbb{H} : f(p) = 0\}.$$

Suppose that $f(0) = 0$ and $X_1 f(0) > 0$. Then there exists an open neighborhood $U_0 \subset U$ and open sets $A_0 \subset A = \exp(\mathfrak{a})$, $B_0 \subset B = \exp(\mathfrak{b})$, where $\mathfrak{a} = \operatorname{span}\{X_2, X_3\}$ and $\mathfrak{b} = \operatorname{span}\{X_1\}$, so that $S \cap U_0$ is an intrinsic graph over A_0 along B_0. Moreover, the defining function ϕ of the graph is continuous and unique.

We now return to our discussion of codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurfaces. Together with C^1 horizontal curves and Legendrian submanifolds, such hypersurfaces constitute one of the most notable examples of intrinsic regular submanifolds, as introduced in [105]. We explicitly remark that here, 'intrinsic' refers to properties defined only in terms of the group structure of $\mathbb{H}$, or, to be more precise, of its Lie algebra $\mathfrak{h}$. Roughly speaking, a subset $S \subset \mathbb{H}$ is to be considered an intrinsic regular submanifold if:

(i) S has, at each point, a tangent 'plane' and a transversal 'plane',
(ii) the tangent planes depend continuously on the point, and the notion of 'plane' is intrinsic to $\mathbb{H}$,
(iii) the tangent and transversal planes are subgroups (or better, cosets of subgroups) of $\mathbb{H}$, and $\mathbb{H}$ is their direct product,
(iv) the tangent plane to S at a point is a suitable limit of group dilations of S centered at that point.

In addition to this list, codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurfaces enjoy the following important properties:

Theorem 4.44. *Any codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurface is locally an intrinsic graph.*

Theorem 4.45. *Any codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurface has locally finite intrinsic Hausdorff $(Q-1)$-dimensional measure.*

Codimension one intrinsic $C^1_{\mathbb{H}}$-regular hypersurfaces can be very irregular objects from a Euclidean point of view. Indeed, these surfaces are in general not Euclidean C^1 submanifolds (not even locally), and in fact can be fractal. (See the notes to the chapter for more discussion.)

For further insight on this phenomenon, we recall in brief Rumin's construction of a complex of differential forms in $\mathbb{H}$ which plays the role of the De Rham complex for Euclidean spaces.

Let us denote by $\bigwedge^k \mathfrak{h}$ the vector space of k-forms over $\mathfrak{h}$ and let $\omega = dx_3 - \frac{1}{2}(x_1 dx_2 - x_2 dx_1) \in \bigwedge^1 \mathfrak{h}$ denote the contact form in $\mathbb{H}$ (see (2.5)). Define $\mathcal{I}^k \subset \bigwedge^k \mathfrak{h}$ as the differential ideal generated by ω:

$$\mathcal{I}^k = \left\{ \eta \in \bigwedge^k \mathfrak{h} : \eta = \omega \wedge \alpha + d\omega \wedge \beta \right\},$$

and let

$$\mathcal{J}^k = \left\{ \eta \in \bigwedge^k \mathfrak{h} : \eta \wedge \omega = 0, \eta \wedge d\omega = 0 \right\}.$$

We also introduce, for an open set $U \subset \mathbb{H}$, the set $\mathcal{D}^k_{\mathbb{H}}(U)$ of *Heisenberg k-differential forms*, i.e., the space of smooth sections, compactly supported in U, of $\frac{\bigwedge^k \mathfrak{h}}{\mathcal{I}^k}$ when $k = 1$, or of $\mathcal{J}^k$ when $k = 2, 3$. These spaces are endowed with the natural topology induced by the topology on $\mathcal{D}^k(U)$, the space of k-differential forms in $\mathbb{R}^3$. With this machinery in place we assert:

Theorem 4.46 (Rumin). *There exists a linear second order differential operator $D : \mathcal{D}^1_{\mathbb{H}}(U) \to \mathcal{D}^2_{\mathbb{H}}(U)$ so that the sequence*

$$0 \to C^\infty_0(U) \xrightarrow{d} \mathcal{D}^1_{\mathbb{H}}(U) \xrightarrow{D} \mathcal{D}^2_{\mathbb{H}}(U) \xrightarrow{d} \mathcal{D}^3_{\mathbb{H}}(U) \to 0 \tag{4.53}$$

is locally exact and has the same cohomology as the De Rham complex on U. Here d is the operator induced by the external differentiation from $\mathcal{D}^k_{\mathbb{H}}(U)$ to $\mathcal{D}^{k+1}_{\mathbb{H}}(U)$ when $k \neq 1$.

The objects in the Rumin complex (4.53) in dimension $k = 1$ are quotients of the usual space of 1-forms, so that their duals are contained in the duals of the usual 1-forms. In dimensions $k = 2, 3$, however, the objects of Rumin's complex are subspaces of the usual spaces of k-forms, so that their duals include (in some sense) the duals of the usual k-forms. Since one can think of surfaces as duals of forms, this is consistent with the above observation that codimension 1 intrinsic $C^1_{\mathbb{H}}$-regular hypersurfaces can be very singular sets from the Euclidean point of view.

We conclude this section by noting that Rumin's theorem suggests that we define, by duality, currents of Federer–Fleming type in $\mathbb{H}$, together with boundaries and (co-)masses. For further details, we refer the interested reader to [105] (see also [213] and [224]).

4.6 Notes

Notes for Section 4.1**.** Proposition 4.1 allows us to define a *horizontal Levi-Civita connection* associated to the sub-Riemannian metric. We state the following definition in the setting of general Heisenberg groups $\mathbb{H}^n$.

Definition 4.47. Let $\Gamma(H\mathbb{H}^n)$ denote the set of horizontal vector fields in $\mathbb{H}^n$, and $\pi_H : \Gamma(T\mathbb{H}^n) \to \Gamma(H\mathbb{H}^n)$ the projection of a tangent vector field to its horizontal component. For any extension of the sub-Riemannian metric we consider its corresponding Levi-Civita connection ∇ and define

$$\nabla^H : \Gamma(H\mathbb{H}^n) \times \Gamma(H\mathbb{H}^n) \to \Gamma(H\mathbb{H}^n)$$

by letting $\nabla^H_U V := \pi_H \nabla_U V$ for all horizontal sections U and V.

In view of Proposition 4.1, ∇^H is independent of the choice of Riemannian extension. One can easily show that ∇^H satisfies the properties of a connection. We note explicitly that ∇^H is torsion free in the horizontal direction, i.e., $(\mathrm{id} - \pi_H)(\nabla^H_X Y - \nabla^H_Y X) = 0$ where $\mathrm{id} - \pi_H$ denotes projection to the vertical component of $T\mathbb{H}^n$. Indeed,

$$(\mathrm{id} - \pi_H)(\nabla^H_X Y - \nabla^H_Y X) = (\mathrm{id} - \pi_H)\pi_H(\nabla_X Y - \nabla_Y X) = 0$$

since π_H is a projection ($\pi_H^2 = \pi_H$).

With this connection at hand one can proceed to define the second fundamental form and curvatures in analogy with the Riemannian definition. This approach has been pursued in [79] and [232]; it yields the same notions which we have introduced in this chapter.

Notes for Section 4.3**.** The characteristic set $\Sigma(S)$ of a C^2 regular surface in $\mathbb{H}$ is not large. Derridj [86] showed that its surface measure is zero, while more recently Balogh [20] proved that the Hausdorff dimension (with respect to either the Euclidean or Carnot–Carathéodory metric) of the characteristic set is at most one. Balogh also investigated the situation for surfaces of weaker regularity. Among the results which he obtains in the following:

Theorem 4.48 (Balogh). *Let $u \in C^{1,1}_{\mathrm{loc}}(\Omega)$, where $\Omega = B_R \subset \mathbb{R}^2$. Then $|S_u| = 0$, where $S_u = \{z \in \Omega : \nabla_z u(z) + z^\perp/2 = 0\}$.*

Note that S_u is precisely the projection of the characteristic set of the graph of u into the z-plane. For extensions of this circle of ideas to more general groups (and higher codimension hypersurfaces) see Magnani [190], [187].

A version of Theorem 4.9 was proved by Pauls in [221]. In that paper, the method of analysis via approximating metrics was used to gain $W^{1,p}$ estimates on solutions to the minimal surface equation in $\mathbb{H}$.

Example 4.12 is due to Arcozzi and Ferrari [13]. In this paper, among other things, the authors study the distance function from a surface and compute curvatures of its level sets.

The material in Sections 4.2 and 4.3.1 is original to this survey. The computation of the second fundamental form of level sets of regular functions u in $(\mathbb{H}^n, g_L)$ and the derivation of the horizontal mean and Gauss curvatures are a particular case of the results in [56] where the general Carnot group case is studied. The horizontal second fundamental form is defined as in [51]. See [144] and [56] for more general definitions and alternative derivations. The observation at the end of Subsection 4.3.1 regarding the behavior of the horizontal mean curvature at the characteristic locus is due to Giovanna Citti.

Proposition 4.24 was first proved in [63]. Independent proofs and formulations appeared also in [68], and [14].

Notes for Subsections 4.4.1 and 4.4.2. All the results in these subsections are first proved in [63] and are due to Hwang, Cheng, Malchiodi and Yang. In [63], these techniques have been used to great effect to study minimal surfaces in $\mathbb{H}$ and more in general in pseudo-Hermitian structures.

Notes for Section 4.5. Theorem 4.43 is due to Franchi, Serapioni and Serra-Cassano (see also the independently proved *polarized coordinates* version in [65], due to Citti and Manfredini). In their papers, [106], [107], [108], [109], [105], these authors develop an extensive program aimed at exploration of notions of sub-Riemannian rectifiability. We note that while we have stated results in $\mathbb{H}$, the original papers [106] and [105] deal with arbitrary Heisenberg groups, whereas [107], [108] and [109] deal with two-step Carnot groups and extend some of this machinary to even more general groups. For additional information, see the notes to the following chapter. Example 4.40 is taken from Remark 5.9 in [106]. The regularity of parameterizations of intrinsic hypersurfaces in the Heisenberg group is the focus of the recent paper [11] of Ambrosio, Serra-Cassano and Vittone.

We have already seen that hypersurfaces in $\mathbb{H}$ given as level sets of (Euclidean) smooth functions can be quite badly behaved from a sub-Riemannian point of view, due to the presence of characteristic points. On the other hand, $C^1_{\mathbb{H}}$-regular hypersurfaces, while well-adapted to sub-Riemannian analysis, can be extremely irregular from a Euclidean point of view. For instance, Kirchheim and Serra-Cassano [162] provide an example of a $C^1_{\mathbb{H}}$-regular hypersurface whose Euclidean Hausdorff dimension is $5/2$. (By a comparison theorem for Euclidean vs. sub-Riemannian Hausdorff dimensions due to Balogh, Rickly and Serra-Cassano [25], the value $5/2$ is best possible.) In short, hypersurfaces in a sub-Riemannian space which are well behaved for sub-Riemannian analysis are not necessarily well behaved for Riemannian analysis, and vice versa.

The Rumin complex of differential forms was introduced by Rumin in [234]. Theorem 4.46 is taken from that paper. Our presentation of the Rumin complex in $\mathbb{H}$ is a special case of the more general theory developed in [234] in the context of general contact manifolds.

Chapter 5

Sobolev and BV Spaces

In this chapter we review the definitions of Sobolev spaces, BV functions and perimeter of a set relative to the sub-Riemannian structure of $\mathbb{H}$. These notions are crucial for the development of sub-Riemannian geometric measure theory. Our treatment here is brief, focusing only on those aspects most relevant for the isoperimetric problem.

5.1 Sobolev spaces, perimeter measure and total variation

Let us begin by introducing the sub-Riemannian analog of the classical first-order Sobolev spaces. For any open set $\Omega \subset \mathbb{H}$ and any $p \geq 1$ we define the Sobolev space $S^{1,p}(\Omega)$ to be the set of functions $f \in L^p(\Omega)$ such that $\nabla_0 f$ exists in the sense of distributions with $|\nabla_0 f| \in L^p(\Omega)$, and denote the corresponding norm by

$$||f||_{S^{1,p}(\Omega)} = ||f||_{L^p(\Omega)} + ||\nabla_0 f||_{L^p(\Omega)}.$$

We recall that the underlying measure in use here is the Haar measure on $\mathbb{H}$, which agrees with both the exponential of the Lebesgue measure on $\mathfrak{h}$ and the Hausdorff 4-measure associated with the Carnot–Carathéodory metric.

Using group convolution $f * g(x) = \int_{\mathbb{H}} f(xy^{-1})g(y)dy$ and following the outline of the Euclidean argument[1] one can easily verify that $S^{1,p}(\mathbb{H})$ is the closure of $C_0^\infty(\mathbb{H})$ in the norm $|| \cdot ||_{S^{1,p}}$. The local Sobolev space $S^{1,p}_{\mathrm{loc}}(\Omega)$ is defined by replacing L^p with L^p_{loc} in the above definition.

Next we recall the sub-Riemannian analog of the classical notions of variation of a function and perimeter of a set. We begin by defining the total variation of an L^1 distribution. Let $\Omega \subset \mathbb{H}$ and denote by $\mathcal{F}(\Omega)$ the class of $\mathbb{R}^2$-valued functions $\phi = (\phi_1, \phi_2) \in C_0^1(\Omega, \mathbb{R}^2)$ such that $|\phi| \leq 1$.

[1]See, for instance, Section 7.2 in [123].

Definition 5.1. The *variation* of an $L^1_{loc}(\Omega)$ function in an open set $\Omega \subset \mathbb{H}$ is

$$\mathrm{Var}_{\mathbb{H}}(f,\Omega) = \sup_{\phi \in \mathcal{F}(\Omega)} \int_\Omega f(x)(X_1\phi_1 + X_2\phi_2)(x)\,dx. \tag{5.1}$$

If f is C^1, a simple integration by parts implies $\mathrm{Var}_{\mathbb{H}}(f,\Omega) = \int_\Omega |\nabla_0 f|$.

Definition 5.2. The space $BV(\Omega)$ of functions with *bounded variation* in Ω is the space of all functions $f \in L^1(\Omega)$ such that $||f||_{BV(\Omega)} := ||f||_{L^1(\Omega)} + \mathrm{Var}_{\mathbb{H}}(f,\Omega) < \infty$.

Clearly $S^{1,1}(\Omega) \subset BV(\Omega)$.

In the next chapters we will need the following useful approximation result.

Lemma 5.3. *Let $\Omega \subset \mathbb{H}$ be an open set, and let X be one of the function spaces $S^{1,p}$, $p \geq 1$, or BV. For any $u \in X(\Omega)$ there exists a sequence $\{u_k\}_{k\in\mathbb{N}} \subset C^\infty(\Omega)$ such that (i) $u_k \to u$ in $L^1(\Omega)$ as $k \to \infty$, and (ii) $\lim_{k\to\infty} \mathrm{Var}_{\mathbb{H}}(u_k,\Omega) = \mathrm{Var}_{\mathbb{H}}(u,\Omega)$ if $X = BV$, or $\lim_{k\to\infty} \int_\Omega |\nabla_0 u_k - \nabla_0 u|^p = 0$ if $X = S^{1,p}$.*

Definition 5.4. Let $E \subset \mathbb{H}$ be a measurable set and $\Omega \subset \mathbb{H}$ be an open set. The (*horizontal*) *perimeter* of E in Ω is given by

$$P_{\mathbb{H}}(E,\Omega) = \mathrm{Var}_{\mathbb{H}}(\chi_E,\Omega),$$

where χ_E denotes the characteristic function of E. Sets with finite perimeter are called *Caccioppoli sets*. For $\Omega = \mathbb{H}$ we let $P_{\mathbb{H}}(E,\mathbb{H}) = P_{\mathbb{H}}(E)$.

The invariance of the perimeter under left translation and the homogeneity of the perimeter under dilation are easy consequences of the definition, using the behavior of the vector fields X_i under these operations. We collect the relevant facts in the following lemma.

Lemma 5.5. *For any $y \in \mathbb{H}$, $s > 0$, $\Omega \subset \mathbb{H}$ open and $E \subset \mathbb{H}$ Caccioppoli,*

$$P_{\mathbb{H}}(\delta_s E, \delta_s \Omega) = s^3 P_{\mathbb{H}}(E,\Omega) \quad \textit{and} \quad P_{\mathbb{H}}(L_y(E), L_y(\Omega)) = P_{\mathbb{H}}(E,\Omega).$$

Here $L_y(x) = yx$ denotes the operation of left translation by y.

In studying the existence of isoperimetric sets (Theorem 8.3) we will make use of the following basic result.

Proposition 5.6. *The perimeter functional $P_{\mathbb{H}}$ is lower semi-continuous with respect to L^1_{loc} convergence.*

The perimeter of smooth sets has a very explicit integral representation in terms of the underlying Euclidean geometry.

Proposition 5.7. *Let E be a C^1 set, $d\sigma$ the surface measure on ∂E, and $\vec{n}$ the outer unit normal. Then*

$$P_{\mathbb{H}}(E,\Omega) = \int_{\partial E \cap \Omega} \left(\sum_{i=1}^{2} \langle X_i, \vec{n} \rangle^2 \right)^{1/2} d\sigma(x).$$

Proof. Denote by A the 2×3 smooth matrix whose rows are the coefficients of the vector fields $X_j = \sum_{i=1}^{3} a_{ji}\partial_{x_i}$. A simple computation yields

$$\operatorname{div}(A^T\phi) = \sum_{j=1}^{2}\sum_{i=1}^{3} a_{ji}\partial_{x_i}\phi_j = X_1\phi_1 + X_2\phi_2$$

for all $\phi \in \mathcal{F}(\Omega)$. In view of Definition 5.4 we have

$$\begin{aligned}
P_{\mathbb{H}}(E,\Omega) &= \sup_{\phi\in\mathcal{F}(\Omega)} \int_{\Omega} \chi_E(x)(X_1\phi + X_2\phi)(x)\,dx \\
&= \sup_{\phi\in\mathcal{F}(\Omega)} \int_{\Omega\cap E} \operatorname{div}(A^T\phi)\,dx \\
&= \sup_{\phi\in\mathcal{F}(\Omega)} \int_{\Omega\cap\partial E} \langle A^T\phi, \vec{n}\rangle\,d\sigma \\
&= \sup_{\phi\in\mathcal{F}(\Omega)} \int_{\Omega\cap\partial E} \langle \phi, A\vec{n}\rangle\,d\sigma \\
&= \int_{\Omega\cap\partial E} |A\vec{n}|\,d\sigma = \int_{\Omega\cap\partial E} \left(\sum_{i=1}^{2}\langle X_i, \vec{n}\rangle^2\right)^{1/2} d\sigma.
\end{aligned}$$

The proof is concluded. □

Corollary 5.8. *If $S = \{u = 0\}$ is a C^1 hypersurface in $\mathbb{H}$ which bounds an open set $E = \{u < 0\}$, then*

$$P_{\mathbb{H}}(E,\Omega) = \int_{S\cap\Omega} d\mu \tag{5.2}$$

for every domain $\Omega \subset \mathbb{H}$, where

$$d\mu = \frac{|\nabla_0 u|}{|\nabla u|}\,d\sigma = |\pi_H(\nu_1)|d\sigma = |A\vec{n}|\,d\sigma.$$

Note that a reparametrization of S will not change its perimeter.

Remark 5.9. In view of Remark 4.22 we can extend $\mathcal{H}_0$ from $S \setminus \Sigma(S)$ to all of S as a function in $L^1(S, d\mu)$. In fact, $\mathcal{H}_0|\pi_H(\nu_1)|$ is a continuous function in all of S.

Remark 5.10. Proposition 5.7 and Corollary 5.8 extend to the setting of Lipschitz surfaces. Note also that the explicit integral representation in Proposition 5.7 extends to the setting of parametric surface patches S which may not bound a domain; we use the notation $P_{\mathbb{H}}(S)$ for the perimeter of such a surface patch defined as in Proposition 5.7.

Remark 5.11. The definitions of Sobolev space, variation and perimeter measure extend *mutatis mutandis* to general Carnot groups, as do the representation formulas in Proposition 5.7 and Corollary 5.8.

5.1.1 Riemannian perimeter approximation

The representation formula (5.2) can also be obtained via the Riemannian approximation scheme, with the Heisenberg perimeter being in a certain sense the limit of the Riemannian perimeters. Consider a C^1 parameterized surface $S = f(D)$ in $\mathbb{R}^3$, where

$$f = (f_1, f_2, f_3) : D \subset \mathbb{R}^2 \to \mathbb{R}^3. \tag{5.3}$$

If $g : S \to \mathbb{R}$ is a continuous function, then we know from elementary calculus that

$$\int_S g\, d\sigma = \int_D g \circ f(u,v) |\vec{n}(u,v)|\, dudv \tag{5.4}$$

where $\vec{n}(u,v) = f_u \times f_v(u,v)$ is the Euclidean normal to S determined by the parameterization. Consider the sequence of Riemannian metrics g_L on $\mathbb{R}^3$ introduced in Section 2.4. Recall that such metrics are characterized by the condition that X_1, X_2, and $\tilde{X}_3$ form an orthonormal frame. Let C_L^T be the 3×3 matrix whose rows are the coefficients of the vector fields X_1, X_2, and $\tilde{X}_3$:

$$C_L^T = \begin{pmatrix} 1 & 0 & -\frac{1}{2}x_2 \\ 0 & 1 & \frac{1}{2}x_1 \\ 0 & 0 & L^{-1/2} \end{pmatrix}.$$

Using the frame $\mathcal{F} = \{X_1, X_2, \tilde{X}_3\}$ as a coordinate basis we may express the basis of the tangent bundle of S as

$$[\partial_1 f]_{\mathcal{F}} = \left(\partial_1 f_1, \partial_1 f_2, \sqrt{L}\left[\partial_1 f_3 - \frac{(\partial_1 f_2 x_1 - \partial_1 f_1 x_2)}{2}\right]\right)$$

and

$$[\partial_2 f]_{\mathcal{F}} = \left(\partial_2 f_1, \partial_2 f_2, \sqrt{L}\left[\partial_2 f_3 - \frac{(\partial_2 f_2 x_1 - \partial_2 f_1 x_2)}{2}\right]\right).$$

Let ν_L denote the Riemannian normal to S in $(\mathbb{R}^3, g_L)$:

$$[\nu_L]_X = [\partial_1 f]_{\mathcal{F}} \times [\partial_2 f]_{\mathcal{F}}.$$

Simple computations show that $\nu_L = \sqrt{L} C_L^T \vec{n}$ and

$$||\nu_L||_L^2 = \det \mathcal{G},$$

where $\mathcal{G} = (g_{ij})$ is the 2×2 matrix with entries $g_{ij} = \langle \partial_i f, \partial_j f \rangle_L$. Thus

$$\sqrt{\det \mathcal{G}} = ||\nu_L||_L = \sqrt{L} |C_L^T \vec{n}|$$

and the change of variables formula on $(\mathbb{R}^3, g_L)$ yields

$$\int_S d\sigma_L = \int_D ||\nu_L||_L\, dudv = \int_D \sqrt{L} |C_L^T \vec{n}|\, dudv. \tag{5.5}$$

Here $d\sigma_L$ denotes the Riemannian surface area element on S induced by the metric g_L on $\mathbb{H}$.

Note that

$$\lim_{L\to\infty} C_L^T = \lim_{L\to\infty} \begin{pmatrix} 1 & 0 & -\frac{1}{2}x_2 \\ 0 & 1 & \frac{1}{2}x_1 \\ 0 & 0 & \frac{1}{\sqrt{L}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\frac{1}{2}x_2 \\ 0 & 1 & \frac{1}{2}x_1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The last matrix is simply the matrix A from the previous discussion with a row of zeros added. Thus

$$\lim_{L\to\infty} \frac{1}{\sqrt{L}} \int_S d\sigma_L = \lim_{L\to\infty} \int_D |C_L^T \vec{n}|\, dudv = \int_D |A\vec{n}|\, dudv.$$

Using (5.4) we conclude that

$$\lim_{L\to\infty} \frac{1}{\sqrt{L}} \int_S d\sigma_L = \int_D |A\vec{n}|\, dudv \ = \int_S \frac{|A\vec{n}|}{|\vec{n}|}\, d\sigma.$$

Moreover, if $S = \{u = 0\}$ is given as a level set, we have $|A\vec{n}| = |\nabla_0 u|$ and $|\vec{n}| = |\nabla u|$, whence

$$\lim_{L\to\infty} \frac{1}{\sqrt{L}} \int_S d\sigma_L = \int_S d\mu = \int_S \frac{|\nabla_0 u|}{|\nabla u|}\, d\sigma, \tag{5.6}$$

and we conclude that the Riemannian surface measures in S, computed with respect to g_L and rescaled by the factor $1/\sqrt{L}$, tend to the perimeter measure in $\mathbb{H}$ as $L \to \infty$.

Example 5.12. We compute the perimeter measure on the boundary S_ϵ of the Heisenberg ball $B_\epsilon := \{x \in \mathbb{H} : ||x||_{\mathbb{H}} = \epsilon\}$. We parameterize $S_\epsilon = f(D)$, where $f = (f_1 + \mathbf{i}\, f_2, f_3)$ is given by

$$(f_1 + \mathbf{i}\, f_2)(\varphi, \theta) = \epsilon\sqrt{\cos\varphi}\exp(\,\mathbf{i}\,\theta)$$

and

$$f_3(\varphi, \theta) = \frac{1}{4}\epsilon^2 \sin\varphi,$$

and $D = \{(\varphi, \theta) \in \mathbb{R}^2 : -\pi/2 < \varphi < \pi/2, 0 \le \theta < 2\pi\}$. An easy computation gives $|A\vec{n}| = |A(f_\varphi \times f_\theta)| = \epsilon^3\sqrt{\cos\varphi}$ and we conclude

$$\begin{aligned} P_{\mathbb{H}}(B_\epsilon, \Omega) &= \int_{S_\epsilon \cap \Omega} d\mu \\ &= \int_{\{(\varphi,\theta)\in D : (\epsilon\sqrt{\cos\varphi}\exp(\,\mathbf{i}\,\theta), \frac{1}{4}\epsilon^2 \sin\varphi) \in \Omega\}} \epsilon^3 \sqrt{\cos\varphi}\, d\varphi\, d\theta. \end{aligned} \tag{5.7}$$

(Alternatively, we could compute $P_{\mathbb{H}}(B_\epsilon, \Omega)$ via the level set formulation starting from the representation $B_\epsilon = \{u < 0\}$, $u(x) = \epsilon - ||x||_{\mathbb{H}}$.)

5.2 A sub-Riemannian Green's formula and the fundamental solution of the Heisenberg Laplacian

Let D be a bounded C^1 domain $\mathbb{R}^3$ equipped with the approximant metric g_L. Green's formula in Ω takes the form

$$\int_D (f\mathcal{L}_{g_L} g - g\mathcal{L}_{g_L} f) = \int_{\partial D} (f\nu_L(g) - g\nu_L(f)) \frac{d\sigma_L}{\sqrt{L}} \tag{5.8}$$

for $f, g \in C^1(\overline{D})$, where $\mathcal{L}_{g_L} = X_1^2 + X_2^2 + \tilde{X}_3{}^2 = X_1^2 + X_2^2 + L^{-1}X_3^2$ denotes the Laplacian in $(\mathbb{R}^3, g_L)$. Using (5.6) we conclude the sub-Riemannian Green's formula

$$\int_D (f\mathcal{L}g - g\mathcal{L}f) = \int_{\partial D} (f\nu_H(g) - g\nu_H(f))\, d\mu. \tag{5.9}$$

We now apply the preceding discussion to compute the fundamental solution for the Heisenberg Laplacian

$$\mathcal{L} = X_1^2 + X_2^2$$

in $\mathbb{H}$. Let $f \in C_0^\infty(\mathbb{H})$. As in Example 5.12 we fix the defining function $u(x) = \epsilon - ||x||_{\mathbb{H}}$ and write $D_\epsilon = \{u < 0\} \cap \operatorname{supp} f = \{x \in \operatorname{supp} f : ||x||_{\mathbb{H}} > \epsilon\}$ and $S_\epsilon = \{u = 0\} = \{x : ||x||_{\mathbb{H}} = \epsilon\}$. To simplify the notation in what follows we write $N(x) = ||x||_{\mathbb{H}}$ for the Heisenberg norm. For later purposes we record the identity $|\nabla_0 N(x)| = |z|/N(x)$, where $x = (z, x_3)$.

Choosing $g = N^{-2}$ in (5.9) and observing that

$$\nu_H(g) = \left\langle \nabla_0 g, \frac{-\nabla_0 N}{|\nabla_0 N} \right\rangle_1 = 2N^{-3}|\nabla_0 N|$$

on the inner boundary S_ϵ of D_ϵ, we find

$$\int_{D_\epsilon} f\mathcal{L}g - g\mathcal{L}f = \int_{S_\epsilon} 2fN^{-3}|\nabla_0 N| + N^{-2}\left\langle \nabla_0 f, \frac{-\nabla_0 N}{|\nabla_0 N|} \right\rangle_1 d\mu. \tag{5.10}$$

Lemma 5.13. *$\mathcal{L}g = 0$ in $\mathbb{H} \setminus \{o\}$.*

The proof of this lemma is an easy computation. We thus obtain

$$-\int_{D_\epsilon} N^{-2}\mathcal{L}f = \int_{S_\epsilon} 2fN^{-3}|\nabla_0 N| + N^{-2}\left\langle \nabla_0 f, \frac{\nabla_0 N}{|\nabla_0 N|} \right\rangle_1 d\mu. \tag{5.11}$$

Lemma 5.14. $\int_{S_\epsilon} N^{-2}\left\langle \nabla_0 f, \frac{\nabla_0 N}{|\nabla_0 N|} \right\rangle_1 d\mu = O(\epsilon)$.

Indeed, using (5.7) in the desired integral gives

$$\int_{-\pi/2}^{\pi/2}\int_0^{2\pi} \epsilon^{-2} \left\langle \nabla_0 f, \frac{\nabla_0 N}{|\nabla_0 N|} \right\rangle_1 \Bigg|_{x=(\epsilon\sqrt{\cos\varphi}\exp(\mathbf{i}\,\theta), \frac{1}{4}\epsilon^2\sin\varphi)} \epsilon^3 \sqrt{\cos\varphi}\, d\varphi\, d\theta.$$

Another use of (5.7) gives $\int_{S_\epsilon} 2N^{-3}|\nabla_0 N|\, d\mu = 8\pi$. In the first term on the right-hand side in (5.11) we add and subtract $f(o)$ and observe that the term $\int_{S_\epsilon} 2(f - f(o))N^{-3}|\nabla_0 N| = o(1)$ by uniform continuity of f in $\overline{B_\epsilon}$. In conclusion, passing to the limit as $\epsilon \to 0$ in (5.11), we deduce the identity

$$-\int_{\mathbb{H}} N^{-2}\mathcal{L}f = 8\pi f(o),$$

valid for all $f \in C_0^\infty(\mathbb{H})$. An application of the horizontal integration by parts formulas

$$\int_{\mathbb{H}} (X_i f)g = -\int_{\mathbb{H}} f(X_i g)$$

converts this into the following singular integral representation formula:

$$\begin{aligned} f(o) &= \frac{1}{8\pi}\int_{\mathbb{H}} \langle \nabla_0 f(y), \nabla_0(N^{-2})(y)\rangle_1\, dy \\ &= -\frac{1}{4\pi}\int_{\mathbb{H}} \langle \nabla_0 f(y), \nabla_0 N(y)\rangle_1 N(y)^{-3}\, dy. \end{aligned} \tag{5.12}$$

In other words,

Theorem 5.15. *The function*

$$u(x) = (8\pi)^{-1}||x||_{\mathbb{H}}^{-2}$$

is the fundamental solution for the Heisenberg Laplacian $\mathcal{L}$.

An alternate form for (5.12) is

$$\begin{aligned} f(x) &= \frac{1}{8\pi}\int_{\mathbb{H}} \langle \nabla_0 f(y), \nabla_0(N^{-2})(y^{-1}x)\rangle_1\, dy \\ &= -\frac{1}{4\pi}\int_{\mathbb{H}} \langle \nabla_0 f(y), \nabla_0 N(y^{-1}x)\rangle_1 N(y^{-1}x)^{-3}\, dy. \end{aligned} \tag{5.13}$$

(5.13) can be obtained by inserting the test function $f_x(y) = f(xy^{-1})$ in (5.12) and transforming the integrand via the isometry $y \mapsto xy^{-1}$.

5.3 Embedding theorems for the Sobolev and BV spaces

The basic embedding theorems for the Sobolev spaces $S^{1,p}$ have a form similar to those in the Euclidean case. The exponent governing the transition to the supercritical case is the homogeneous dimension of the underlying space. We present statements of the Sobolev embedding theorems for maps in $S^{1,p}(\mathbb{H})$. For later purposes we also discuss the extension of the geometric Sobolev inequality from $S^{1,1}$

to the space BV, local inequalities of Sobolev–Poincaré type, and the compactness of the embedding $BV \subset L^1$ on well-behaved domains. Best constants for the Sobolev inequalities in the Heisenberg group and more general Carnot groups are discussed in Chapter 9.

Theorem 5.16 (Sobolev embedding theorem in the Heisenberg group). $S^{1,p}(\mathbb{H}) \hookrightarrow L^{\frac{4p}{4-p}}(\mathbb{H})$ *for* $1 \le p < 4$ *and* $S^{1,p}(\mathbb{H}) \hookrightarrow C^{0,1-4/p}(\mathbb{H})$ *for* $p > 4$.

More precisely, there exist constants $C_p(\mathbb{H}) < \infty$ for each $p \neq 4$ so that

$$||f||_{4p/(4-p)} \le C_p(\mathbb{H})||\nabla_0 f||_p$$

for all $f \in S^{1,p}(\mathbb{H})$ if $1 \le p < 4$, while if $p > 4$ and $f \in S^{1,p}(\mathbb{H})$, then there exists a representative $\tilde{f}$ of f satisfying the Hölder condition

$$|\tilde{f}(x) - \tilde{f}(y)| \le C_p(\mathbb{H}) d(x,y)^{1-4/p} ||\nabla_0 f||_p$$

for all $x, y \in \mathbb{H}$. In the borderline case ($p = 4$), we have an embedding theorem for an exponentially integrable class (Moser–Trudinger inequality). See Chapter 9.

5.3.1 The geometric case (Sobolev–Gagliardo–Nirenberg inequality)

We begin by proving Theorem 5.16 in the geometric case $p = 1$. We need some preliminary background results. Let $B(o,r) \subset \mathbb{H}$ be a metric ball,[2] and let $q > 1$. Recall that the weak L^q space $L^{q,*}(B(o,r))$ consists of all measurable functions $f : B(o,r) \to \mathbb{R}$ for which the quantity

$$||f||^q_{L^{q,*}(B(o,r))} = \sup_{\lambda > 0} \lambda^q \left|\{x \in B(o,r) \mid |f(x)| > \lambda\}\right|$$

is finite. (We have denoted here by $|A|$ the Haar measure of $A \subset \mathbb{H}$.) We also need the Hardy–Littlewood maximal operator with respect to Carnot–Carathéodory metric balls,

$$Mf(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\, dy.$$

Thanks to the fact that $\mathbb{H}$, equipped with the Carnot–Carathéodory metric d, is homogeneous in the sense of [71],[3] we know that $f \mapsto Mf$ satisfies a weak-type $(1,1)$ estimate: there exists $C > 0$ such that

$$|\{x \in \mathbb{H} : |Mf(x)| > \lambda\}| \le \frac{C}{\lambda} ||f||_{L^1(\mathbb{H})} \tag{5.14}$$

for all $f \in L^1(\mathbb{H})$ and $\lambda > 0$.

We begin with the weak-type Sobolev–Gagliardo–Nirenberg inequality.

[2]By homogeneity we could simply assume $r = 1$ and then rescale the resulting estimates. Similarly, by translation invariance the same estimates hold for balls centered at any point in $\mathbb{H}$.

[3]In modern terminology, $(\mathbb{H}, d)$ is a *doubling metric space*. This follows from the Ahlfors 4-regularity of the metric measure space $(\mathbb{H}, d, \mathcal{H}^4)$. See [136], Chapters 1–3, for more details.

Proposition 5.17. *There exists a constant $C > 0$, such that*

$$\left|\{x \in B(o,r) : |f(x)| > \lambda\}\right| \leq C\lambda^{-\frac{4}{3}} ||\nabla_0 f||^{4/3}_{L^1(B(o,r))}$$

for all Lipschitz functions f compactly supported in $B(o,r)$, and for all $\lambda > 0$.

Proof. Our starting point is the representation formula (5.13). The equivalence of the Carnot–Carathéodory metric d and $d_{\mathbb{H}}$ and the estimate $|\nabla_0 N| \leq 1$ yield

$$|f(x)| \leq C \int_{\mathbb{H}} |\nabla_0 f(y)| d(x,y)^{-3}\, dy. \tag{5.15}$$

In view of (5.15), we see that in order to show the theorem we need only prove that the fractional integration operator

$$g \to I_1 g(x) = \int_{\mathbb{H}} |g(y)| d(x,y)^{-3}\, dy,$$

satisfies a weak-type $(1, 4/3)$ estimate, i.e., there exists a positive constant C such that

$$\left|\{x \in B(o,r) : |I_1 g(x)| > \lambda\}\right| \leq C\lambda^{-\frac{4}{3}} ||g||^{\frac{4}{3}}_{L^1(\mathbb{H})} \tag{5.16}$$

for all $g \in L^1(\mathbb{H})$ with compact support in $B(o,r)$. Following the argument in [54, Theorem 2.1] or [136, Chapter 3] we set $I_1 g = 0$ outside $B(o,r)$ and let $\epsilon = \epsilon(\lambda, ||g||_{L^1(\mathbb{H})}) > 0$ be sufficiently small (to be chosen later). We write

$$I_1 g = I_1^1 g + I_1^2 g,$$

where

$$I_1^1 g(x) = \int_{B(x,\epsilon)} |g(y)| d(x,y)^{-3}\, dy$$

and

$$I_1^2 g(x) = \int_{B(o,1)\setminus B(x,\epsilon)} |g(y)| d(x,y)^{-3}\, dy.$$

A simple dyadic decomposition argument yields the existence of a constant $C_1 > 0$ such that

$$\begin{aligned} I_1^1(g) &\leq \sum_{k=0}^{\infty} (2^{-k-1}\epsilon)^{-3} \int_{B(x,2^{-k}\epsilon)\setminus B(x,2^{-k-1}\epsilon)} |g(y)|\, dy \\ &\leq C_1 \sum_{k=0}^{\infty} 2^{-k}\epsilon \left(\frac{1}{|B(x,2^{-k}\epsilon)|} \int_{B(x,2^{-k}\epsilon)} |g(y)|\, dy \right) \\ &\leq 2C_1 \epsilon M g(x). \end{aligned} \tag{5.17}$$

We also have the trivial estimate

$$I_1^2 g(x) \leq \epsilon^{-3} ||g||_{L^1(\mathbb{H})}. \tag{5.18}$$

Next, using (5.14) and (5.17), we have

$$\begin{aligned}\left|\{I_1 g > \lambda\}\right| &\le \left|\{I_1^1 g > \frac{\lambda}{2}\}\right| + \left|\{I_1^2 g > \frac{\lambda}{2}\}\right| \\ &\le \left|\{Mg > \frac{\lambda}{2C_1\epsilon}\}\right| + \left|\{I_1^2 g > \frac{\lambda}{2}\}\right| \\ &\le \frac{2C_1\epsilon}{\lambda}||g||_{L^1(\mathbb{H})} + \left|\{I_1^2 g > \frac{\lambda}{2}\}\right|.\end{aligned}$$

To estimate the second term on the right-hand side it suffices to assume

$$\lambda > 2r^{-3}||g||_{L^1(\mathbb{H})}. \tag{5.19}$$

Indeed, if $\lambda \le 2r^{-3}||g||_{L^1(\mathbb{H})}$, then trivially $|B(o,r)| \le C\lambda^{-4/3}||g||^{4/3}_{L^1(\mathbb{H})}$ and we obtain the desired estimate $|\{I_1 g > \lambda\}| \le |B(o,r)| \le C\lambda^{-4/3}||g||^{4/3}_{L^1(\mathbb{H})}$. Thus assume that (5.19) holds, and choose $\epsilon = (\lambda^{-1}2||g||_{L^1(\mathbb{H})})^{1/3} < r$. By (5.18), $I_1^2 g(x) < \lambda/2$ whence $|\{I_1^2 g > \frac{\lambda}{2}\}| = 0$ and

$$\left|\{I_1 g > \lambda\}\right| \le \frac{2C_1}{\lambda}||g||_{L^1(\mathbb{H})}(\lambda^{-1}2||g||_{L^1})^{1/3} \le C\lambda^{-4/3}||g||^{4/3}_{L^1}.$$

The proof is concluded. □

An elegant truncation argument due to Maz'ya allows us to pass from the weak-type inequality to the corresponding strong-type inequality.

Proposition 5.18. *There exists a constant $C_1(\mathbb{H}) < \infty$ so that*

$$||f||_{4/3} \le C_1(\mathbb{H})||\nabla_0 f||_1 \tag{5.20}$$

for all $f \in S^{1,1}(\mathbb{H})$.

Proof. It suffices to prove the estimate for nonnegative, smooth, compactly supported functions f on $\mathbb{H}$. Let f be such a function, choose $R > 0$ so that $B(o,R)$ contains the support of f, and write

$$A_j = \{x \in B(o,R) : 2^j < u(x) \le 2^{j+1}\}, \qquad j \in \mathbb{Z}.$$

Writing

$$f_j = \max\{0, \min\{f - 2^j, 2^j\}\},$$

we observe that $\nabla_0 f_j$ is supported on A_j. By the weak-type estimate in Proposition 5.17 and Theorem 5.15,

$$\begin{aligned}|A_{j+1}| &\le |\{f_j > 2^{j-1}\}| \\ &\le |\{I_1(|\nabla_0 f_j|) > C^{-1}2^{j-1} \\ &\le C\left(2^{-j}\int_{A_j}|\nabla_0 f_j|\right)^{4/3} = C\left(2^{-j}\int_{A_j}|\nabla_0 f|\right)^{4/3}\end{aligned}$$

(note that f_j is Lipschitz). Thus

$$\begin{aligned}\int_{\mathbb{H}} |f|^{4/3} &= \sum_{j\in\mathbb{Z}} \int_{A_j} |f|^{4/3} \le \sum_{j\in\mathbb{Z}} (2^{j+1})^{4/3}|A_j| \\ &\le C \sum_{j\in\mathbb{Z}} \left(\int_{A_j} |\nabla_0 f| \right)^{4/3} \\ &\le C \left(\sum_{j\in\mathbb{Z}} \int_{A_j} |\nabla_0 f| \right)^{4/3} = C \left(\int_{\mathbb{H}} |\nabla_0 f| \right)^{4/3}.\end{aligned}$$

The proof is complete. □

The estimate in Proposition 5.18 holds more generally for elements of the space BV. Approximating $f \in BV$ by C_0^∞ functions as in Lemma 5.3 leads to

Proposition 5.19. *There exists $C > 0$ so that $||f||_{4/3} \le C \operatorname{Var}_{\mathbb{H}}(f)$ for all $f \in BV(\mathbb{H})$.*

5.3.2 The subcritical case

The subcritical case $1 \le p < 4$ of Theorem 5.16 can be derived from the geometric case $p = 1$ by an elementary trick. Again, note that by Lemma 5.3 it suffices to prove the estimate for $f \in C_0^\infty$. We apply the geometric inequality (5.20) to a suitable power of f, i.e., $g = |f|^s$, to obtain

$$\begin{aligned}\left(\int_{\mathbb{H}} |f|^{4s/3} \right)^{3/4} &\le C_1(\mathbb{H}) s \int_{\mathbb{H}} |f|^{s-1} |\nabla_0 f| \\ &\le C_1(\mathbb{H}) s \left(\int_{\mathbb{H}} |f|^{(s-1)q} \right)^{1/q} ||\nabla_0 f||_p,\end{aligned}$$

where q denotes the Hölder conjugate of p. With $s = (3p)/(4-p)$ we have

$$(s-1)q = \frac{4s}{3} = \frac{4p}{4-p}$$

and

$$||f||_{4p/(4-p)} \le sC_1(\mathbb{H}) ||\nabla_0 f||_p,$$

which proves the estimate with

$$C_p(\mathbb{H}) \le \frac{3p}{4-p} C_1(\mathbb{H}). \tag{5.21}$$

5.3.3 The supercritical case

We prove the following theorem:

Theorem 5.20. *If $p > 4$, then every element of $S^{1,p}(\mathbb{H})$ has a representative in $C^{0,1-4/p}(\mathbb{H})$.*

To prove this result we will introduce two useful spaces: A function $u \in L^1_{loc}(\mathbb{H})$ is in the *Morrey space* $L^{1,\lambda}(\mathbb{H})$ if

$$\sup_{B(x,R)} R^{-\lambda} \int_{B(x,R)} |u(y)|\, dy < \infty. \tag{5.22}$$

Similarly, a function $u \in L^1_{loc}(\mathbb{H})$ is in the *Campanato space* $\mathcal{L}^{1,\lambda}(\mathbb{H})$ if

$$[u]_{1,\lambda} = \sup_{B(x,R)} R^{-\lambda} \int_{B(x,R)} |u(y) - u_{B(x,R)}|\, dy < \infty. \tag{5.23}$$

Using the Poincaré inequality (5.38) and Hölder's inequality, we immediately have

Lemma 5.21. *If $u \in S^{1,1}_{loc}(\mathbb{H})$ and $|\nabla_0 u| \in L^{1,4-\lambda}(\mathbb{H})$, then $u \in \mathcal{L}^{1,5-\lambda}(\mathbb{H})$. If $u \in S^{1,p}_{loc}(\mathbb{H})$ with $p > 4$, then $u \in \mathcal{L}^{1,5-\frac{4}{p}}(\mathbb{H})$.*

In view of the previous lemma, Theorem 5.20 is an immediate corollary of

Proposition 5.22. *If $0 < \lambda < 1$, then every element of $\mathcal{L}^{1,5-\lambda}(\mathbb{H})$ has a representative in $C^{0,1-\lambda}(\mathbb{H})$.*

Proof. Set $\alpha = 5 - \lambda$. For $0 < r < R$ we estimate

$$\begin{aligned} r^4|u_{B(x,R)} - u_{B(x,r)}| &= C \int_{B(x,r)} |u_{B(x,R)} - u_{B(x,r)}| \\ &\leq C \left[\int_{B(x,R)} |u(y) - u_{B(x,R)}| dy + \int_{B(x,r)} |u(y) - u_{B(x,r)}| dy \right] \\ &\leq C R^\alpha [u]_{1,\alpha}, \end{aligned} \tag{5.24}$$

for some $C > 0$ not depending on λ.

Choose $R > 0$ and set $R_i = R2^{-i}$, $i \in \mathbb{N}$. From (5.24), and iterating from i to j via a telescoping sum argument, we obtain

$$|u_{B(x,R_i)} - u_{B(x,R_j)}| \leq C[u]_{1,\alpha} R_i^{\alpha-4} \text{ for } i < j. \tag{5.25}$$

In view of (5.25), $(u_{B(x,R_i)})$ is a Cauchy sequence. Hence $\tilde{u}(x) := \lim_{i\to\infty} u_{B(x,R_i)}$ exists. By Lebesgue's theorem, $\tilde{u}$ is a (Lebesgue) representative of u. Letting $j \to \infty$ in (5.25) yields

$$|u_{B(x,R_i)} - \tilde{u}(x)| \leq C[u]_{1,\alpha} R_i^{\alpha-4}. \tag{5.26}$$

This estimate implies that $u_{B(x,R_i)}$ converges uniformly to $\tilde{u}(x)$, hence $\tilde{u}$ is continuous. To prove Hölder continuity we consider $x, y \in \mathbb{H}$ and set $R = d(x,y)$. Applying (5.26) again we have

$$\begin{aligned} |\tilde{u}(x) - \tilde{u}(y)| &\le |\tilde{u}(x) - u_{B(x,2R)}| + |u_{B(x,2R)} - u_{B(y,2R)}| + |\tilde{u}(y) - u_{B(y,2R)}| \\ &\le CR^{\alpha-4} + |u_{B(x,2R)} - u_{B(y,2R)}|. \end{aligned}$$

To complete the proof, we observe that an argument similar to the one employed in the proof of (5.24) yields the desired estimate for the term $|u_{B(x,2R)} - u_{B(y,2R)}|$. □

5.3.4 Compactness of the embedding $BV \subset L^1$ on John domains

We recall the definition of a John domain. A bounded, connected open set Ω in a metric space (X, d) is a *John domain* if there exists a point $x_0 \in \Omega$ (the *center* of the domain) and a constant $\delta > 0$ such that for any point $x \in \Omega$ there exists an arc length parameterized rectifiable path $\gamma : [0, L] \to \Omega$ so that (i) $\gamma(0) = x$ and $\gamma(L) = x_0$, and (ii) $d(\gamma(t), X \setminus \Omega) \ge \delta t$.

In Euclidean spaces every Lipschitz domain is a John domain. In the Heisenberg group it is considerably more difficult to verify the John property.

Lemma 5.23. *Both Carnot–Carathéodory balls $B(x,R)$ and gauge balls $B_{\mathbb{H}}(y,R) = \{x \in \mathbb{H} \,:\, d_{\mathbb{H}}(x,y) < R\}$ are John domains.*

Theorem 5.24 (Garofalo–Nhieu). *Let $\Omega \subset \mathbb{H}$ be a John domain. The space $BV(\Omega)$ is compactly embedded in $L^1(\Omega)$.*

The John property enters into play in the proof of Theorem 5.24 in a crucial way in the following two lemmas. The first is a semi-global version of the Sobolev–Poincaré embedding of BV in $L^{4/3}$ from the previous section.

Lemma 5.25. *Let $\Omega \subset \mathbb{H}$ be a John domain. There exists a constant $C_{1,\Omega} > 0$ so that*

$$\left(\frac{1}{|\Omega|} \int_\Omega |u - u_\Omega|^{4/3} \right)^{3/4} \le C_{1,\Omega} \frac{\operatorname{diam}(\Omega)}{|\Omega|} \operatorname{Var}_{\mathbb{H}}(u, \Omega)$$

for all $u \in BV(\Omega)$. Here we have denoted by u_Ω the average of the function u over the domain Ω.

For simplicity, in the following we will use the same notation C_P to denote any constant for which the Poincaré–Sobolev and the Poincaré inequality both hold.

The second lemma is a relative isoperimetric inequality for John domains. We state it only in the case of metric balls.

Lemma 5.26. *Let $B \subset \mathbb{H}$ be a metric ball. Then there exists a constant $C > 0$ so that*

$$\min(|A|, |B \setminus A|)^{3/4} \le C P_{\mathbb{H}}(A, B)$$

for any measurable set $A \subset B$.

Proof of Theorem 5.24. In view of Lemma 5.3 it is sufficient to prove that the set $S = \{u \in BV(\Omega) \cap C^\infty(\Omega) : ||u||_{BV} \le 1\}$ is totally bounded. Let $\epsilon > 0$ and let $\Omega_\epsilon \subset \Omega$ be an open set such that

$$C_P \, \frac{\operatorname{diam}(\Omega)}{|\Omega|^{1/4}} |\Omega \setminus \Omega_\epsilon|^{1/4} + \frac{|\Omega \setminus \Omega_\epsilon|}{|\Omega|} < \frac{\epsilon}{6}, \tag{5.27}$$

where C_P is a Poincaré–Sobolev constant as in Lemma 5.25. For any $u \in S$, using Lemma 5.25, we have the estimate

$$\begin{aligned} \int_{\Omega\setminus\Omega_\epsilon} |u| &\le \int_{\Omega\setminus\Omega_\epsilon} |u - u_\Omega| + |u_\Omega||\Omega \setminus \Omega_\epsilon| \\ &\le \left(\int_\Omega |u - u_\Omega|^{4/3} \right)^{3/4} |\Omega \setminus \Omega_\epsilon|^{1/4} + \frac{|\Omega \setminus \Omega_\epsilon|}{|\Omega|} \int_\Omega |u| \\ &\le C_P \frac{\operatorname{diam}(\Omega)}{|\Omega|^{1/4}} |\Omega \setminus \Omega_\epsilon|^{1/4} \operatorname{Var}_{\mathbb{H}}(u, \Omega) + \frac{|\Omega \setminus \Omega_\epsilon|}{|\Omega|} \int_\Omega |u|. \end{aligned} \tag{5.28}$$

Since $u \in S$, (5.27) and (5.28) yield

$$\int_{\Omega\setminus\Omega_\epsilon} |u| < \frac{\epsilon}{6}. \tag{5.29}$$

Since Ω_ϵ is precompact we may use a Vitali-type covering argument. There exists M positive so that for any sufficiently small $\delta > 0$ one can construct a family of balls $\{B_j\}$, $j = 1, \ldots, N$ with the following properties: (i) the diameter of each B_j is $\delta\epsilon$, (ii) $\Omega_\epsilon \subset \cup_{j=1}^N B_j \subset \Omega$, (iii) $\sum_{j=1}^N \chi_{C_P B_j} \le M \chi_\Omega$. We choose $\delta > 0$ so that

$$2 C_P M \delta < \frac{1}{3}. \tag{5.30}$$

Next, we consider any function $v \in BV(\Omega) \cap C^\infty(\Omega)$ and estimate

$$\begin{aligned} \int_{\Omega_\epsilon} |v| &\le \sum_{j=1}^N \int_{B_j} |v| \, dx \le \sum_{j=1}^N \int_{B_j} |v - v_{B_j}| \, dx + |B_j||v_{B_j}| \\ &\le \sum_{j=1}^N \left(C_P \delta\epsilon \int_{C_P B_j} |\nabla_0 v| \, dx + \left| \int_{B_j} v \, dx \right| \right) \\ &\le C_P M \delta\epsilon \operatorname{Var}_{\mathbb{H}}(v, \Omega) + \sum_{j=1}^N \left| \int_{B_j} v \, dx \right| \\ &\le \frac{\epsilon}{6} \operatorname{Var}_{\mathbb{H}}(v, \Omega) + \sum_{j=1}^N \left| \int_{B_j} v \, dx \right|. \end{aligned} \tag{5.31}$$

At this point we define a compact linear operator $T : BV(\Omega) \to \mathbb{R}^N$ as follows:

$$T(u) = \left(\int_{B_1} u\,dx, \ldots, \int_{B_N} u\,dx \right).$$

By compactness there exist functions $u_1, \ldots, u_M \in S$ so that for any $u \in S$,

$$|Tu - Tu_j| = \sum_{k=1}^{N} \int_{B_k} |u - u_j|\,dx < \frac{\epsilon}{3} \tag{5.32}$$

for some $j \in \{1, \ldots, M\}$. Moreover,

$$||u - u_j||_{L^1(\Omega)} \le \int_{\Omega_\epsilon} |u - u_j|dx + \int_{\Omega \setminus \Omega_\epsilon} |u - u_j|dx = I + II. \tag{5.33}$$

From (5.31) and (5.32) we have

$$\begin{aligned} I &\le \frac{\epsilon}{6} \operatorname{Var}_{\mathbb{H}}(u - u_j) + \sum_{k=1}^{N} \left| \int_{B_k} (u - u_j)\,dx \right| \\ &\le \frac{\epsilon}{6}(||u||_{BV(\Omega)} + ||u_j||_{BV(\Omega)}) + \frac{\epsilon}{3} \le \frac{2\epsilon}{3}. \end{aligned} \tag{5.34}$$

For the second term in (5.33) we use (5.29) to estimate

$$II \le \int_{\Omega \setminus \Omega_\epsilon} |u|dx + \int_{\Omega \setminus \Omega_\epsilon} |u_j|dx \le \frac{\epsilon}{3}. \tag{5.35}$$

From (5.33), (5.34) and (5.35) we obtain that S is totally bounded. □

5.4 Further results: Sobolev and Sobolev–Poincaré embedding theorems and analysis in metric spaces

Sobolev embedding theorem in Carnot groups. Theorem 5.16 holds in general Carnot groups in the following form:

Theorem 5.27 (Sobolev embedding theorem for Carnot groups). *Let $\mathbb{G}$ be a Carnot group of homogeneous dimension Q. Then $S^{1,p}(\mathbb{H}) \hookrightarrow L^{\frac{Qp}{Q-p}}(\mathbb{H})$ for $1 \le p < Q$ and $S^{1,p}(\mathbb{H}) \hookrightarrow C^{0,1-Q/p}(\mathbb{H})$ for $p > Q$.*

The proof which we gave in the case of $\mathbb{H}$ works in general as well. The only difference comes at the beginning of the proof: an explicit fundamental solution for the Laplacian $\mathcal{L}$ is not known in arbitrary groups. However, Folland [99] showed the existence of such a fundamental solution Γ and the associated estimates

$$|\Gamma(x)| \le C_1 d(x, o)^{2-Q}$$

and
$$|\nabla_0\Gamma(x)| \le C_2 d(x,o)^{1-Q}$$
for appropriate constants $C_1, C_2 > 0$, where d denotes the Carnot–Carathéodory distance in $\mathbb{G}$ and Q is the homogeneous dimension. Then the rest of the proof proceeds without change starting from (5.15).

Sobolev–Poincaré inequalities and analysis on metric measure spaces. The Sobolev embedding theorem 5.16 refers to functions defined on all of $\mathbb{H}$. It is easy to see that the estimate $||f||_{4/3,B} \le C||\nabla_0 f||_{1,B}$ cannot hold for arbitrary smooth functions f defined only on a ball $B \subset \mathbb{H}$ (consider the case when f is constant). An appropriate local analog for Theorem 5.18 is the *Poincaré inequality*
$$\left(\frac{1}{|C^{-1}B|}\int_{C^{-1}B} |f(x)-f_B|^{4/3}\,dx\right)^{3/4} \le C\,\mathrm{rad}(B)\left(\frac{1}{|B|}\int_B |\nabla_0 f(x)|\,dx\right) \quad (5.36)$$
for functions $f \in S^{1,1}(B)$ defined on a ball B. Here $f_B = |B|^{-1}\int_B f$ denotes the average value of f on B, and $C^{-1}B$ denotes the ball concentric with B whose radius is C^{-1} times that of B. The constant C is independent of both f and B.

We omit the proof of (5.36) here, but observe that it also follows from the weak $(1,4/3)$ estimate for the fractional integral operator I_1, together with the following modified representation estimate:
$$|f(x)-f_B| \le C\int_B |\nabla_0 f(y)| d(x,y)^{-3}\,dy \quad (5.37)$$
valid for $x \in C^{-1}B$ and an arbitrary $f \in C^\infty(B)$. See [180] or [136] for further details.

Hölder's inequality applied to both sides of (5.36) yields the following family of Poincaré-type inequalities:
$$\frac{1}{|C^{-1}B|}\int_{C^{-1}B} |f(x)-f_B|\,dx \le C\,\mathrm{rad}(B)\left(\frac{1}{|B|}\int_B |\nabla_0 f(x)|^p\,dx\right)^{1/p}, \quad (5.38)$$
$1 \le p < \infty$. Inequality (5.38) is the so-called $(1,p)$-*Poincaré inequality* which plays a key role in the abstract development of first-order calculus on metric measure spaces. See [136] and the notes to this chapter for further discussion.

Horizontal polar coordinates in $\mathbb{H}$. The Heisenberg group $\mathbb{H}$ admits a system of horizontal polar coordinates, whose features parallel in many respects the classical polar coordinate system in Euclidean space. This family of curves can be used to give an alternate derivation of the representation formula (5.13).

Theorem 5.28. *There exists a family of curves $\Gamma = \{\gamma_y\}$ in $\mathbb{H}$, parameterized by points in the twice-punctured Korányi sphere $S := \{y \in \mathbb{H} : ||y||_{\mathbb{H}} = 1, y_3 \neq 0\}$, with the following properties:*

1. *each curve* $\gamma_y : (0, \infty) \to \mathbb{H}$ *is horizontal and satisfies* $\gamma_y(1) = y$;
2. *the curves* γ_y *are adapted to the Korányi norm:* $||\gamma_y(s)||_{\mathbb{H}} = s$;
3. $\bigcup_{y \in S} \gamma_y = \{x \in \mathbb{H} : x_3 \neq 0\}$;
4. *there exists a Borel measure* $d\lambda$ *on* S *so that for all* $f \in L^1(\mathbb{H}) \cap C^0(\mathbb{H})$, *the following integration formula holds true:*

$$\int_{\mathbb{H}} f(x)\, dx = \int_S \int_0^\infty f(\gamma_y(s))\, s^3\, ds\, d\lambda(y). \tag{5.39}$$

Remark 5.29. If we parameterize S as in Example 5.12, we can write the measure $d\lambda$ in the explicit form $d\lambda = d\varphi d\theta$. (Note that this differs from the perimeter measure $P_{\mathbb{H}}$ which we computed in Example 5.12.)

Note that the simpler polar coordinate integration formula

$$\int_{\mathbb{H}} f(x)\, dx = \int_S \int_0^\infty f(\delta_s(y))\, s^3\, ds\, d\lambda(y) \tag{5.40}$$

also holds for $f \in L^1(\mathbb{H}) \cap C^0(\mathbb{H})$. Indeed, the change of variables from $\{s, \varphi, \theta\}$ to $\{x_1, x_2, x_3\}$ has Jacobian determinant equal to s^3. However, (5.40) cannot be used to establish the representation formula (5.12) since the radial curves $s \mapsto \tilde\gamma_s(y) = \delta_s(y)$ are not horizontal (note that $\frac{\partial}{\partial s}\tilde\gamma_s(y) = y_1X_1 + y_2X_2 + 2sy_3X_3$).

The curves γ_y which figure in Theorem 5.28 were already identified in equation (4.25) as the Legendrian foliation of the Heisenberg cones P_α. For the reader's convenience, we restate the definition here:

$$\gamma_y(s) = \left(swe^{4\,\mathbf{i}\,\frac{y_3}{|w|^2}\log s}, \frac{1}{4}s^2y_3 \right), \qquad y = (w, y_3) \in S. \tag{5.41}$$

Part 1 of Theorem 5.28 is a restatement of the fact that these curves are the Legendrian foliation of P_α, part 2 is an easy exercise, and part 3 follows from the fact that $\mathbb{H} \setminus \{y_3 = 0\} = \cup_{\alpha \in \mathbb{R}} P_\alpha$. Part 4 follows from a computation of the Jacobian determinant of the change of variables

$$\Phi : \begin{cases} x_1 = s\sqrt{\cos\varphi}\cos(\theta + 4\tan\varphi\log s), \\ x_2 = s\sqrt{\cos\varphi}\sin(\theta + 4\tan\varphi\log s), \\ x_3 = s^2\sin\varphi \end{cases}$$

from $\{s, \varphi, \theta\}$ to $\{x_1, x_2, x_3\}$. Indeed, $J_\Phi \equiv s^3$. An easy computation gives

$$\begin{aligned} \frac{\partial}{\partial s}\gamma_y(s)\bigg|_{\gamma_s(y)=x} &= \frac{1}{s}(x_1 - 4\alpha x_2)X_1 + \frac{1}{s}(x_2 + 4\alpha x_1)X_2 \\ &= \frac{1}{s^3\cos\varphi}\left((x_1|z|^2 - 4x_2x_3)X_1 + (x_2|z|^2 + 4x_1x_3)X_2\right), \end{aligned}$$

where we used again the fact that $|z| = s\sqrt{\cos\varphi}$. For each $y \in S$, since $\cos\varphi = (1+\alpha^2)^{-1/2}$ is constant along γ_y, we find

$$\begin{aligned}
-\cos\varphi f(o) &= \int_0^\infty \frac{d}{ds} f(\gamma_y(s)) \cos\varphi \, ds \\
&= \int_0^\infty \langle \nabla_0 f(\gamma_y(s)), \frac{\partial}{\partial s}\gamma_y(s)\rangle_1 \cos\varphi \, ds \\
&= \int_0^\infty (X_1 f)(\gamma_y(s)) \frac{x_1|z|^2 - 4x_2x_3}{s^3} + (X_2 f)(\gamma_y(s)) \frac{x_2|z|^2 + 4x_1x_3}{s^3} \, ds.
\end{aligned}$$

Integrating over $-\pi/2 < \varphi < \pi/2$ and $0 \le \theta \le 2\pi$ and using (5.39) gives

$$-4\pi f(o) = \int_{\mathbb{H}} X_1 f(x) \frac{x_1|z|^2 - 4x_2x_3}{||x||_{\mathbb{H}}^6} + X_2 f(x) \frac{x_2|z|^2 + 4x_1x_3}{||x||_{\mathbb{H}}^6} \, dx. \tag{5.42}$$

Note that the vector field

$$\frac{x_1|z|^2 - 4x_2x_3}{||x||_{\mathbb{H}}^3} X_1 + \frac{x_2|z|^2 + 4x_1x_3}{||x||_{\mathbb{H}}^3} X_2$$

is precisely the horizontal gradient of the Korányi norm. The identity (5.42) may thus be rewritten in the form

$$-4\pi f(o) = \int_{\mathbb{H}} \langle \nabla_0 f, \nabla_0 N \rangle_1 N^{-3},$$

coinciding with (5.13).

5.5 Notes

Notes for Section 5.1. The classical notion of perimeter was introduced by De Giorgi [84], see also Giusti [125]. A related notion in the sub-Riemannian setting of the first Heisenberg group was first introduced by Pansu in [217]. The definitions of perimeter (a la De Giorgi) and BV spaces presented here have been introduced independently by Capogna, Danielli and Garofalo [53], by Franchi, Gallot and Wheeden [104] and by Biroli and Mosco [35]. The definitions are meaningful in the setting of Carnot–Carathéodory spaces associated to families of Hörmander vector fields and even in more general spaces.

The approximation of horizontal perimeter via perimeters on the Riemannian approximants in Section 5.1.1 is based on the work of Korányi and Reimann [170].

Franchi, Serapioni and Serra-Cassano [106], [107], [108], [109] have used their notion of $C^1_{\mathbb{H}}$-regularity to develop a theory of sub-Riemannian rectifiability for

Cacciopoli sets. According to the point of view taken in these papers, a set $E \subset \mathbb{H}$ is a *codimension* 1 *rectifiable set* if there exists a sequence of $C^1_{\mathbb{H}}$-regular hypersurfaces $\{S_j\}$ (see Section 4.5 for the definition) so that

$$\mathcal{H}^3_{\mathbb{H}}(E \setminus \bigcup_j S_j) = 0.$$

The notion of horizontal normal may be extended in a generalized sense to any Caccіopoli set $E \subset \mathbb{H}$: we call ν_H a *generalized outer normal* to E if

$$\int_E X_1\phi_1 + X_2\phi_2 = \int_{\partial E} \langle \nu_H, \phi \rangle_1 \, d\mu$$

for any $\phi = (\phi_1, \phi_2) \in C_0^\infty(\mathbb{H}, \mathbb{R}^2)$. The existence of such ν_H follows in a standard way from the Riesz representation theorem. Transposing the Euclidean notion to the Heisenberg context, one next defines the *reduced boundary* $\partial^* E$ as the collection of points $x \in \partial E$ for which $\mu(B(x,r)) > 0$ for all $r > 0$ and

$$\lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} \nu_H \, d\mu$$

exists and has norm 1. Using the Implicit Function Theorem discussed in Section 4.5 and a version of the Whitney extension theorem, Franchi, Serapioni and Serra-Cassano show the following analog of the celebrated structure theorem of de Giorgi for sets of finite perimeter:

Theorem 5.30. *If $E \subset \mathbb{H}$ is a Caccioppoli set, then the reduced boundary $\partial^* E$ is rectifiable.*

In addition, they provide a wealth of additional geometric information concerning the reduced boundary. For instance, a key ingredient in the proof is a blow-up argument which identifies suitable generalized tangent spaces at points of the reduced boundary with vertical half-planes. Quite recently, Theorem 5.30 has been used by Cheeger and Kleiner to prove the bi-Lipschitz nonembeddability of $\mathbb{H}$ in L^1 [61].

In [109], the authors adapt their machinery in $\mathbb{H}^n$ to deal with intrinsic submanifolds of any dimension and codimension in general Carnot groups and create an analog of Federer–Fleming currents in this setting. Recently, V. Magnani [188] has provided a different approach to this problem.

Notes for Section 5.2. Theorem 5.15 is due to Folland [98], and extends to higher-dimensional Heisenberg groups and more general groups of Heisenberg type. (See the notes to Chapter 2.) The proof which we give via the horizontal Green's formula is due to Korányi and Reimann [170]. The alternate derivation which we present in Section 5.4, via the horizontal polar coordinate system in $\mathbb{H}$, is in the spirit of modern analysis on metric spaces, where representation formulas and estimates, Poincaré inequalities, etc. are derived abstractly from the existence of rich families of rectifiable curves in metric spaces. For more information, see [136] or [240]. The horizontal polar coordinate system in $\mathbb{H}$ was introduced by Korányi and Reimann [170]; the general theory of Carnot groups supporting such a polar coordinate system was studied by Balogh and Tyson [26].

In general Carnot groups no explicit formula for the corresponding fundamental solution of the Laplacian is known, however, Folland [99] established the existence and basic estimates. For explicit integral representations of the fundamental solution to the Laplacian on groups of step two, see Beals, Gaveau and Greiner [29]. For additional information, see the notes to Chapter 2.

Notes for Section 5.3. The Sobolev embedding theorems (Section 5.3) are standard tools in PDE and analysis. See [136], [132], [259] for various proofs and discussion. In the context of vector fields satisfying the Hörmander condition, such embedding theorems were obtained in the $p > 1$ case independently by Lu in [180] and [181] and Capogna-Danielli-Garofalo [52]. Our treatment is taken from [53], but see also Hajłasz and Koskela [132], which contains a much more complete and detailed list of references for work in this area. The subcritical Sobolev embedding theorem for $S^{1,p}$, $1 < p < 4$, can also be derived as a corollary of a suitable boundedness theorem for the fractional integration operator I_1 on L^p. Indeed, I_1 satisfies a strong-type $(p, 4p/(4-p))$ inequality; the proof is similar to that presented here in the geometric case. This line of argument originates in work of Hedberg; see [3]. The trick to derive the strong-type Sobolev inequality from the corresponding weak type estimate goes back to work of Maz'ya. See also [136, Chapter 3]. An extensive theory of representation formulas and embedding theorems in sub-Riemannian spaces has been developed by Lu and Wheeden; a partial list of references includes [179, 182–186].

The Poincaré inequality for vector fields satisfying the Hörmander condition is due to Jerison [155]. Varopoulos [254] has given an elegant, purely elementary, proof in the context of Carnot groups, see also Proposition 11.17 in [132]. Sobolev–Poincaré inequalities in this setting were further explored in the papers [54] and [53]. For a new proof of the Poincaré inequality for Hörmander vector fields, see Lanconelli and Morbidelli [174], and for a striking generalization to vector fields with minimal regularity assumptions, see Montanari and Morbidelli [198]. Axiomatic formulations of the Poincaré inequality (5.38) are the starting point for a rich theory of first-order analysis and Sobolev spaces on metric measure spaces. This rapidly expanding area dates back to the 1998 paper of Heinonen and Koskela [141]. See [143], [136] or the forthcoming monograph [142].

The Campanato spaces introduced in the proof of the supercritical case and the proof of the subelliptic analogue of Campanato's theorem (5.22) follow closely the Euclidean counterpart [49, Theorem 5.22] (see also [121, Chapter 3]). With minimal modifications, all the proofs in this section extend to spaces associated to systems of smooth vector fields satisfying Hörmander's finite rank condition.

Notes for Subsection 5.3.4. Theorem 5.24 is due to Garofalo and Nhieu [116] and is a cornerstone of the existence theory for isoperimetric profiles which we describe in Chapter 8.

The geometry of the base domain plays a crucial role in the development of Sobolev–Poincaré type estimates. In [155], Jerison showed that CC balls satisfy a chaining property known as the *Boman chain condition*, which arose in unpublished work of Boman and echoes a Whitney decomposition argument due to Kohn [164]. After Jerison the Boman chain condition was used by numerous authors to prove successively more general versions of the Sobolev–Poincaré inequality, to the extent that domains satisfying this condition are sometimes referred to as PS (Poincaré–Sobolev) domains. For a detailed account of the developments in this field, an excellent reference is the monograph of Hajłasz and Koskela [132]. Despite the obvious relevance of this geometric property, until recently, even in the simple setting of the Heisenberg group, the only known examples of sub-Riemannian PS domains were CC balls and finite unions of CC balls. The equivalence between John and PS domains was shown in [116] in the Carnot–Carathéodory context, and in [46] in the general setting of doubling metric spaces. The importance of this result lies in the fact that it allows the invocation of results from geometric function theory to construct large classes of John domains. For instance, since any NTA (non-tangentially accessible), uniform or (ϵ, δ) domain is John, the results from [55], [208], [209] and [57] lead to several examples of John domains. The most general result in this direction is due to Monti and Morbidelli [208] who proved that every $C^{1,1}$ domain in any Carnot group of step two is a John domain.

Chapter 6

Geometric Measure Theory and Geometric Function Theory

In this chapter we introduce some basic notions which are crucial for the development of sub-Riemannian geometric measure theory. Our treatment here is brief, focusing only on those aspects most relevant for the isoperimetric problem. We review and discuss Pansu's formulation of the Rademacher differentiation theorem for Lipschitz functions on the Heisenberg group, and the basic area and co-area formulas. As an application of the former we sketch the equivalence of horizontal perimeter and Minkowski 3-content in $\mathbb{H}$. In Section 6.4 we present two derivations of first variation formulas for the horizontal perimeter: first, away from the characteristic locus, and second, across the characteristic locus. In the final section, we give a rough outline of Mostow's rigidity theorem for cocompact lattices in the complex hyperbolic space $H^2_{\mathbb{C}}$, emphasizing the appearing of sub-Riemannian geometric function theory in the asymptotic analysis of boundary maps on the sphere at infinity.

Sub-Riemannian geometric measure theory is a rapidly expanding discipline. In this chapter, we have omitted any substantive discussion of numerous important recent developments, such as the investigations of rectifiability by Ambrosio–Kirchheim and Franchi–Serapioni–Serra-Cassano. Our focus throughout is on those topics and results most relevant for submanifold geometry and the sub-Riemannian isoperimetric problem.

6.1 Area and co-area formulas

The area and co-area formulas are fundamental tools in classical geometric measure theory. These integral formulas have sub-Riemannian analogs and indeed can be generalized to much more general metric spaces. To begin, we remind the reader of the classical area formula (see [95] Section 3.2.1):

Theorem 6.1 (Euclidean area formula). *Let $A \subset \mathbb{R}^n$ be a measurable set, let $f : A \to \mathbb{R}^m$ be a Lipschitz map, and let $m \geq n$. Then,*

$$\int_A J(f)(x)\, d\mathcal{H}^n(x) = \int_{\mathbb{R}^m} N(f, A, y)\, d\mathcal{H}^m(y)$$

where $J(f)$ is the Jacobian of f, $\mathcal{H}^n$ is the n-dimensional Hausdorff measure, and

$$N(f, A, y) = \operatorname{card}\{x \in A : f(x) = y\}.$$

To generalize Theorem 6.1 to the setting of Carnot groups, we require a notion of horizontal Jacobian for Lipschitz maps between Carnot groups. We note that while the co-area formula holds for arbitrary Carnot groups, the reader focused on understanding the Heisenberg group may wish to assume that $\mathbb{G}_i \in \{\mathbb{H}, \mathbb{R}^n\}$ in what follows. We begin with a definition of a metric notion of the Jacobian.

Definition 6.2. Let $\mathbb{G}_1, \mathbb{G}_2$ be Carnot groups equipped with sub-Riemannian metrics d_1, d_2 respectively. Let $A \subset \mathbb{G}_1$ be an $\mathcal{H}^Q$ measurable set, where Q is the homogeneous dimension of $\mathbb{G}_1$, and let $f : A \to \mathbb{G}_2$ be Lipschitz with respect to the sub-Riemannian metrics on $\mathbb{G}_1$ and $\mathbb{G}_2$. The *metric Jacobian* of f at $x \in A$ is defined to be

$$J_f^m(x) = \liminf_{r \to 0} \frac{\mathcal{H}^Q_{d_2}(f(B_A(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))}$$

where $B_A(x,r) = \{y \in A | d_1(x,y) \leq r\}$.

With this notion we easily recover the following standard integration formula:

$$\int_A J_f^m(x) d\mathcal{H}^Q_{d_1}(x) = \mathcal{H}^Q_{d_2}(f(A)). \tag{6.1}$$

We note that (6.1) trivially yields a property analogous to the Sard theorem, namely, if $J_f^m = 0$ a.e. in A, then $\mathcal{H}^Q_{d_2}(f(A)) = 0$.

A key ingredient of the proof of the sub-Riemannian area formula is an analysis of the set $A_0 = \{x \in A : J_f^m(x) = 0\}$. To better understand this set, we introduce Pansu's sub-Riemannian analog for the differential.

Definition 6.3. Let $(\mathbb{G}_1, d_1)$ and $(\mathbb{G}_2, d_2)$ be Carnot groups with homogeneous structures determined by the dilations δ_s^i, $i = 1, 2$. Suppose that $f : \mathbb{G}_1 \to \mathbb{G}_2$ is a measurable map. Then the *Pansu differential* of f at x is the map

$$D_0 f(x) : \mathbb{G}_1 \to \mathbb{G}_2$$

defined by

$$D_0 f(x)(y) = \lim_{s \to 0} \delta^2_{1/s} f(x)^{-1} f(x \delta^1_s y)$$

whenever the limit exists.

For a Carnot group $\mathbb{G}$ with sub-Riemannian metric $d_{\mathbb{G}}$, we may recognize the metric tangent space $T_x\mathbb{G}$ at a point $x \in \mathbb{G}$ as the pointed Gromov–Hausdorff limit of the sequence of metric spaces $(x^{-1}\mathbb{G}, \lambda d_{\mathbb{G}})$ as $\lambda \to \infty$. As $\mathbb{G}$ is equipped with a family of homotheties (δ_s),

$$(x^{-1}\mathbb{G}, sd_{\mathbb{G}}) = (x^{-1}\mathbb{G}, d_{\mathbb{G}} \circ (\delta_s \times \delta_s)),$$

whence $(x^{-1}\mathbb{G}, d_{\mathbb{G}})$ and $(x^{-1}\mathbb{G}, sd_{\mathbb{G}})$ are isometric via the map δ_s. Thus the sequence above is constant and $T_x\mathbb{G}$ is isometric to $\mathbb{G}$. From this point of view, we can think of $D_0 f$ as a map between tangent spaces. Alternatively, by conjugating with the exponential map, we can view $D_0 f$ as a map between the Lie algebras: $(D_0 f)_* : \mathfrak{g}_1 \to \mathfrak{g}_2$.

Pansu's extension of the Rademacher differentiability theorem to Carnot groups reads as follows:

Theorem 6.4 (Pansu–Rademacher differentiation theorem). *Let $\mathbb{G}_1, \mathbb{G}_2$ be Carnot groups and let $A \subset \mathbb{G}_1$ be a measurable set. Let $f : A \subset \mathbb{G}_1 \to \mathbb{G}_2$ be Lipschitz with respect to the sub-Riemannian metrics on $\mathbb{G}_1$ and $\mathbb{G}_2$. Then, for a.e. $x \in A$, $D_0 f(x)$ exists and is a horizontal linear map (i.e., a graded homogeneous group homomorphism, see Definition* 2.1*) between $\mathbb{G}_1$ and $\mathbb{G}_2$.*

In Section 6.2, we prove a special case of this theorem when $\mathbb{G}_1 = \mathbb{H}$ and $\mathbb{G}_2 = \mathbb{R}$.

Example 6.5. Let $\mathbb{G}_1 = \mathbb{G}_2 = \mathbb{H}$. The Pansu differential $D_0 f$ of a Lipschitz map $f = (f_1, f_2, f_3) : \mathbb{H} \to \mathbb{H}$, acting on the Lie algebra $\mathfrak{h}$ and expressed in terms of the standard basis X_1, X_2, X_3, takes the form

$$\begin{pmatrix} X_1 f_1 & X_1 f_2 & 0 \\ X_2 f_1 & X_2 f_2 & 0 \\ 0 & 0 & X_1 f_1 X_2 f_2 - X_1 f_2 X_2 f_1 \end{pmatrix}.$$

The next lemma establishes a link between $D_0 f$ and J_f^m.

Lemma 6.6. *Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be Carnot groups with homogeneous dimensions $Q = Q_1$ and $Q' = Q_2$ respectively, with $Q' \geq Q$, and let $f : A \subset \mathbb{G}_1 \to \mathbb{G}_2$ be Lipschitz. If $D_0 f(x)$ is not injective for some $x \in A$, then $J_f^m(x) = 0$.*

Proof. We sketch the main ideas in the proof in the case $A = \mathbb{G}_1$. Let $P = D_0 f(x)(\mathbb{G}_1)$ be the image of the entire group under the differential mapping. Since $D_0 f$ is not injective at x, the Hausdorff dimension of P is less than or equal to $Q - 1$. Moreover, as the Pansu differential exists, we have

$$d_2(D_0 f(x)(y), \delta^2_{1/s} f(x)^{-1} f(x\delta^1_s y)) = o(1)$$

as $s \to 0$. Using left invariance of the metric and the dilation property (2.20), we find

$$d_2(D_0 f(x)(y), \delta^2_{1/s} f(x)^{-1} f(x\delta^1_s y)) = \frac{1}{s} d_2(f(x)\delta^2_s D_0 f(x)(y), f(x\delta^1_s y))$$

so

$$d_2(f(x)\delta_s^2 D_0 f(x)(y), f(x\delta_s^1 y)) = o(s).$$

Since $D_0 f(x)(\delta_s^1 y) = \delta_s^2 D_0 f(x)(y)$, we may rewrite this as

$$d_2(f(x)D_0 f(x)(y), f(xy)) = o(d_1(x, y)).$$

Consequently, if $B(x, r) \subset \mathbb{G}_1$ is the metric ball of radius r centered at x, then its image under f lies in an $o(r)$ neighborhood N of $f(x)D_0 f(x)(B(o, r))$ where $o \in \mathbb{G}_1$ is the identity element. Since f is Lipschitz,

$$\mathcal{H}^Q_{d_2}(f(A)) \le (\operatorname{Lip} f)^Q \mathcal{H}^Q_{d_1}(A) \tag{6.2}$$

for all measurable $A \subset \mathbb{G}_1$, where

$$\operatorname{Lip} f(x) = \limsup_{r \to 0} \frac{\sup\{d_2(f(x), f(y)) : d_1(x, y) \le r\}}{r}$$

denotes the *pointwise Lipschitz constant* of f. Combining (6.2) with the observation that $f(x)D_0 f(x)(B(o, r))$ has Hausdorff dimension less than or equal to $Q-1$, it follows from a covering argument that

$$\mathcal{H}^Q_{d_2}(f(B(x, r))) \le \mathcal{H}^Q_{d_2}(N) = o(r^Q) = o(\mathcal{H}^Q_{d_1}(B(x, r))).$$

The result follows from the definition of the metric Jacobian. $\square$

We omit the proof of the following lemma, which is an adaptation of a classical argument (see 3.2.2 in [95]). For further details, see the notes to this chapter.

Lemma 6.7. *Suppose that $f : A \subset \mathbb{G}_1 \to \mathbb{G}_2$ is a Lipschitz map and let $\lambda > 1$. Let $\tilde{A}$ be the set of points of density of A where $D_0 f$ exists and is injective. Then there exist Borel sets $\{E_i\}$ partitioning $\tilde{A}$ so that for each i,*

- *$f|_{E_i}$ is injective,*
- *there exist injective horizontal linear maps L_i so that*

$$\frac{1}{\lambda} d_2(L_i(z), o) \le d_2(D_0 f(x)(z), o) \le \lambda d_2(L_i(z), o) \tag{6.3}$$

  for $x \in E_i$ and $z \in \mathbb{G}_1$, where $o \in \mathbb{G}_2$ is the identity, and
- *we have*

$$\operatorname{Lip}(f|_{E_i} \circ L_i|_{C_i}^{-1}) \le \lambda, \quad \operatorname{Lip}(L_i|_{E_i} \circ f|_{E_i}^{-1}) \le \lambda. \tag{6.4}$$

Using this setup and notion of metric Jacobian, we can prove a version of the area formula for Carnot groups.

Theorem 6.8 (Sub-Riemannian area formula). *Let $(\mathbb{G}_1, d_1)$ and $(\mathbb{G}_2, d_2)$ be Carnot groups and let $A \subset \mathbb{G}_1$ be an $\mathcal{H}^Q_{d_1}$ measurable set. Let $Q = Q_1$ and $Q' = Q_2$ be the Hausdorff dimensions of $\mathbb{G}_i$, $i = 1, 2$, and assume that $Q' \geq Q$. If $f : A \to \mathbb{G}_2$ is Lipschitz, then*

$$\int_A J^m_f(x)\, d\mathcal{H}^Q_{d_1}(x) = \int_{\mathbb{G}_2} N(f, A, y)\, d\mathcal{H}^Q_{d_2}(y)$$

where $N(f, A, y) = \operatorname{card}\{x \in A : f(x) = y\}$.

Proof. Let A_0 be the set of points of density of A at which either $J^m_f = 0$ or $D_0 f(x)$ does not exist. Since $\mathcal{H}^Q_{d_2}(f(A_0)) = 0$ we may disregard the set A_0 in the proof of the area formula as it makes no contribution to either side of the equation. By Lemma 6.6, $D_0 f(x)$ is injective for $x \in A \setminus A_0$. Using Lemma 6.7, we find sets $\{C_i\}$ partitioning $A \setminus A_0$, with $f_i = f|_{C_i}$ injective. Let $J^m_i(x) := J^m_{f_i}(x)$. Since the closed balls in a Carnot group form a Vitali relation, we know that

$$\lim_{r \to 0} \frac{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))}{\mathcal{H}^Q_{d_1}(B_A(x, r))} = 1 \tag{6.5}$$

for all i and all points of density $x \in C_i$. A quick calculation using (6.5) shows that $J^m_f = J^m_i$ at such points:

$$\begin{aligned} J^m_i(x) &= \liminf_{r \to 0} \frac{\mathcal{H}^Q_{d_2}(f_i(B_{A \cap C_i}(x, r)))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} \\ &\leq \liminf_{r \to 0} \frac{\mathcal{H}^Q_{d_2}(f(B_A(x, r)))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} \\ &= \liminf_{r \to 0} \frac{\mathcal{H}^Q_{d_2}(f(B_A(x, r)))}{\mathcal{H}^Q_{d_1}(B_A(x, r))} = J^m_f(x). \end{aligned}$$

Similarly,

$$\begin{aligned} J^m_f(x) &\leq \liminf_{r \to 0} \left(\frac{\mathcal{H}^Q_{d_2}(f(B_{A \setminus C_i}(x, r)))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} + \frac{\mathcal{H}^Q_{d_2}(f(B_{A \cap C_i}(x, r)))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} \right) \\ &\leq \liminf_{r \to 0} \left((\operatorname{Lip} f)^Q \frac{\mathcal{H}^Q_{d_1}(B_{A \setminus C_i}(x, r))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} + \frac{\mathcal{H}^Q_{d_2}(f(B_{A \cap C_i}(x, r)))}{\mathcal{H}^Q_{d_1}(B_{A \cap C_i}(x, r))} \right) \\ &= J^m_i(x). \end{aligned}$$

Using (6.1) and the fact that the f_i are injective, we have

$$\int_{C_i} J^m_f(x)\, d\mathcal{H}^Q_{d_1}(x) = \int_{C_i} J^m_i(x)\, d\mathcal{H}^Q_{d_1}(x) = \int_{\mathbb{G}_2} \chi_{f(C_i)}\, d\mathcal{H}^Q_{d_2}. \tag{6.6}$$

Summing over all i yields

$$\int_A J_f^m(x)\, d\mathcal{H}_{d_1}^Q(x) = \int_{A\setminus A_0} J_f^m(x)\, d\mathcal{H}_{d_1}^Q(x) = \int_{\mathbb{G}_2} N(f,C,y)\, d\mathcal{H}_{d_2}^Q(y). \qquad \square$$

To complete our discussion of the area formula, we note that a more geometric notion of the Jacobian is equivalent to the metric Jacobian at points of Pansu differentiability. Recall the notion of a horizontal linear map between Carnot groups (Definition 2.1).

Definition 6.9. The *horizontal Jacobian* of a horizontal linear map $\phi : \mathbb{G}_1 \to G_2$ is

$$J_H(\phi) = \frac{\mathcal{H}_{d_2}^Q(\phi(A))}{\mathcal{H}_{d_1}^Q(A)},$$

where A is any measurable subset of $\mathbb{G}_1$ of positive finite measure.

We note that a simple covering argument, combined with the homogeneity and left invariance of the Hausdorff measure, shows that the value of $J_H(\phi)$ is independent of the choice of A.

Proposition 6.10. *Let $f : A \subset \mathbb{G}_1 \to \mathbb{G}_2$ be a Lipschitz map. Then $J_f^m = J_H(D_0 f)$ at point of Pansu differentiability in A.*

Proof. Let x be a point of Pansu differentiability for f in A. By Lemma 6.6, we may assume that $D_0 f(x)$ is injective; else both sides are zero. Using Lemma 6.7 with a sequence $\lambda_n \to 1$, we find sets E_n containing x as points of density and horizontal linear maps L_n satisfying (6.3) and (6.4) with $\lambda = \lambda_n$. Thus

$$\begin{aligned}\lambda_n^{-Q} J_H(L_n) &= \lim_{r\to 0} \lambda_n^{-Q} \frac{\mathcal{H}_{d_2}^Q(L_n(B_A(x,r)))}{\mathcal{H}_{d_1}^Q(B_A(x,r))} \\ &= \lim_{r\to 0} \lambda_n^{-Q} \frac{\mathcal{H}_{d_2}^Q(L_n(B_{A\cap E_n}(x,r)))}{\mathcal{H}_{d_1}^Q(B_A(x,r))},\end{aligned}$$

since x is a point of density of E_n and L_n is a horizontal linear map

$$\le \limsup_{r\to 0} \frac{\mathcal{H}_{d_2}^Q(f(B_{A\cap E_n}(x,r)))}{\mathcal{H}_{d_1}^Q(B_A(x,r))},$$

by (6.3)

$$\le \limsup_{r\to 0} \frac{\mathcal{H}_{d_2}^Q(f(B_A(x,r)))}{\mathcal{H}_{d_1}^Q(B_A(x,r))} = J_f^m(x).$$

Similarly,

$$\begin{aligned}
J_f^m(x) &= \liminf_{r\to 0} \frac{\mathcal{H}^Q_{d_2}(f(B_A(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))} \\
&\le \liminf_{r\to 0} \left(\frac{\mathcal{H}^Q_{d_2}(f(B_{A\setminus E_n}(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))} + \frac{\mathcal{H}^Q_{d_2}(f(B_{A\cap E_n}(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))} \right) \\
&\le \liminf_{r\to 0} \left((\operatorname{Lip} f)^Q \frac{\mathcal{H}^Q_{d_1}(B_{A\setminus E_n}(x,r))}{\mathcal{H}^Q_{d_1}(B_A(x,r))} + \frac{\mathcal{H}^Q_{d_2}(f(B_{A\cap E_n}(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))} \right),
\end{aligned}$$

since x is a point of density for E_n

$$\le \lambda_n^Q \lim_{r\to 0} \frac{\mathcal{H}^Q_{d_2}(L_n(B_{A\cap E_n}(x,r)))}{\mathcal{H}^Q_{d_1}(B_A(x,r))},$$

by (6.4) $\qquad \le \lambda_n^Q J_H(L_n).$

After possibly passing to a subsequence, we have $J_H(L_n) \to J_H(D_0 f(x))$ and $\lambda_n \to 1$, yielding the claim. □

While the proof is beyond the scope of this survey, there are also sub-Riemannian analogs of the co-area formula. We present one such result in the setting of the Heisenberg group.

Theorem 6.11 (Co-area formula). *For all $f \in BV(\Omega)$, $\Omega \subset \mathbb{H}$, we have*

$$||f||_{BV(\Omega)} = \int_{\mathbb{R}} P_{\mathbb{H}}(E_t, \Omega)\, dt, \tag{6.7}$$

where $E_t = \{x \in \Omega : f(x) > t\}$.

6.2 Pansu–Rademacher theorem

In this section, we prove Pansu's Rademacher theorem, Theorem 6.4, for real-valued Lipschitz functions on the Heisenberg group. Our starting point is the following proposition.

Proposition 6.12. *Let $\Omega \subset \mathbb{H}$ be a domain and let $f : \Omega \to \mathbb{R}$ be a Lipschitz function. Then $\nabla_0 f(x)$ exists a.e. in Ω.*

Proof. Consider the case $\Omega = \mathbb{H}$ (the general case is similar). We may fiber $\mathbb{H}$ by integral curves for the vector field X_1. Denote by $L_p = \{pe^{tX_1}\}$ the integral curve passing through an arbitrary point $p \in \{(x_1, x_2, x_3) \in \mathbb{H} : x_1 = 0\}$. Since L_p is a horizontal line, the map $f_p : \mathbb{R} \to \mathbb{R}$ given by $f_p(t) = f(pe^{tX_1})$ is Lipschitz, and hence is differentiable a.e. Note that $f_p'(t) = X_1 f(pe^{tX_1})$. Thus $X_1 f$ exists $\mathcal{H}^1$-a.e. on L_p. By Fubini's theorem, $X_1 f$ exists a.e. in $\mathbb{H}$. In a similar fashion, we obtain that $X_2 f$ exists a.e. □

At points $x \in \Omega$ satisfying the conclusion in Proposition 6.12, we define $D_0 f(x) : \mathbb{H} \to \mathbb{R}$ by the formula

$$D_0 f(x)(y) = X_1 f(x) y_1 + X_2 f(x) y_2. \tag{6.8}$$

The fact that $D_0 f$ is the Pansu differential of f is expressed in the following theorem.

Theorem 6.13. *Let $f : \Omega \to \mathbb{R}$ be Lipschitz. Then*

$$\lim_{y \to 0} \frac{f(xy) - f(x) - D_0 f(x)(y)}{d(0, y)} = 0$$

for a.e. $x \in \Omega$.

To begin the proof of Theorem 6.13, we observe that $\nabla_0 f \in L^\infty \subset L^p_{\mathrm{loc}}$ for all $p \geq 1$. Fix an index $p > 4$. Since (Ω, d) is a homogeneous space, we may apply the Lebesgue differentiation theorem to conclude

$$\lim_{r \to 0} \frac{1}{|B(o,r)|} \int_{B(o,r)} |\nabla_0 f(xy) - \nabla_0 f(x)|^p \, dy = 0 \tag{6.9}$$

for almost every $x \in \Omega$. Fix $x \in \Omega$ so that $|\nabla_0 f(x)| < \infty$ and (6.9) holds true. Define

$$g(y) = f(xy) - D_0 f(x)(y) = f(xy) - \langle \nabla_0 f(x), \pi(y) \rangle$$

and observe that

$$\nabla_0 g(y) = \nabla_0 f(xy) - \nabla_0 f(x) \tag{6.10}$$

since the horizontal gradient is left invariant.

For each $y \in \mathbb{H}$ such that $xy \in \Omega$, we have the inequality

$$\begin{aligned} |f(xy) - f(x) - D_0 f(x)(y)| &= |g(y) - g(0)| \\ &\leq C r^{1-4/p} ||\nabla_0 g||_{p, B(o,r)} \\ &\leq C r \left(\frac{1}{|B(o,r)|} \int_{B(o,r)} |\nabla_0 g|^p \right)^{1/p} \end{aligned}$$

for $r = 2d(0, y)$ by the supercritical case of the Sobolev embedding theorem. Thus

$$\frac{|f(xy) - f(x) - D_0 f(x)(y)|}{d(0, y)} \leq C \left(\frac{1}{|B(o,r)|} \int_{B(o,r)} |\nabla_0 g|^p \right)^{1/p}$$

which tends to zero by (6.9) and (6.10). This finishes the proof of Theorem 6.13.

Lemma 6.14. *$D_0 f(x)$ is a group homomorphism from $\mathbb{H}$ to $\mathbb{R}$ which commutes with the dilations:*

$$D_0 f(x)(\delta_s y) = s D_0 f(x)(y).$$

These facts follow immediately from the definition (6.8) and the form for the group law in the horizontal layer. As a consequence of Theorem 6.13 and Lemma 6.14, we have that

$$\lim_{s \to 0} \frac{f(x\delta_s y) - f(x)}{s} \tag{6.11}$$

exists and equals $D_0 f(x)(y)$ for all $y \in \mathbb{H}$ and a.e. $x \in \Omega$. Thus $D_0 f(x)$ agrees with the Pansu differential as defined in Definition 6.3. Observe that $||\nabla_0 f||_\infty \le L$ if f is L-Lipschitz. Indeed, (6.11) implies that

$$|X_1 f(x) y_1 + X_2 f(x) y_2| = |D_0 f(x)(y)| \le L d(0, y)$$

for all y; choosing $y = (\nabla_0 f(x)/|\nabla_0 f(x)|, 0)$ gives $|\nabla_0 f(x)| \le L d(0, y) = L$.

Example 6.15. For a closed set $F \subset \mathbb{H}$, let $\rho(x) = \operatorname{dist}(x, F) = \inf\{d(x, y) : y \in F\}$. Then $\rho : \mathbb{H} \to \mathbb{R}$ is a 1-Lipschitz map. We verify the *eikonal equation*

$$|\nabla_0 \rho| = 1 \qquad \text{a.e. in } \mathbb{H} \setminus F. \tag{6.12}$$

The upper bound $|\nabla_0 \rho(x)| \le 1$ holds by Proposition 6.12. Moreover,

$$\lim_{y \to 0} \frac{\rho(xy) - \rho(x) - D_0 \rho(x)(y)}{d(0, y)} = 0 \tag{6.13}$$

for a.e. x by Theorem 6.13. Fix such an x in the complement of F. Since F is closed, there exists $p \in G$ which realizes the distance from F to x: $d(x, p) = \operatorname{dist}(x, F) = \rho(x)$. Let $\gamma : [0, T] \to \mathbb{H}$ be an arc length parameterized geodesic joining 0 to $x^{-1}p$. Then $x\gamma$ is a geodesic from x to p, and

$$\rho(x\gamma(t)) = T - t$$

for all $0 \le t \le T$. Hence

$$\frac{\rho(x\gamma(t)) - \rho(x)}{d(x, x\gamma(t))} = \frac{(T - t) - T}{t} = -1$$

for all such t. Applying (6.13) with $y = \gamma(t)$ gives

$$\rho(x\gamma(t)) - \rho(x) = D_0 \rho(x)(\gamma(t)) + o(t)$$

or

$$D_0 \rho(x)(\gamma(t)) = -t + o(t).$$

Let $\pi : \mathbb{H} \to \mathbb{C}$ denote the projection $\pi(y) = y_1 + \mathbf{i}\, y_2$. Since $\pi(\gamma(t)) \ne 0$ and $|\pi(\gamma(t))| \le d(0, \gamma(t)) = t$ for $0 < t < T$ we find

$$|\nabla_0 \rho(x)| \ge \frac{|D_0 \rho(x)(\gamma(t))|}{|\pi(\gamma(t))|} \ge \frac{t - o(t)}{t} = 1 - o(1)$$

as $t \to 0$. This finishes the proof of the eikonal equation (6.12).

6.3 Equivalence of perimeter and Minkowski content

An alternate notion of surface measure in the Heisenberg group is provided by the Minkowski (3-)content. Conveniently, this notion agrees with the horizontal perimeter measure, at least for sufficiently smooth sets. We begin with the definition.

Definition 6.16. Let $\operatorname{dist}(x,E) := \inf\{d(x,y) : y \in E\}$ be the distance to a bounded set $E \subset \mathbb{H}$ and let $E_\epsilon = \{x \in \mathbb{H} : \operatorname{dist}(x,E) < \epsilon\}$ be the ϵ-*neighborhood* of E. The *Minkowski* (3-)*content* of ∂E is

$$\mathcal{M}_3(\partial E) := \lim_{\epsilon \to 0} \frac{|E_\epsilon \setminus E|}{\epsilon},$$

provided the limit exists.

Proposition 6.17. *$P_{\mathbb{H}}(E) = \mathcal{M}_3(\partial E)$ for all bounded open sets $E \subset \mathbb{H}$ with C^2 boundary.*

Proof. We denote by $\mathcal{M}_3^+(\partial E)$, resp. $\mathcal{M}_3^-(\partial E)$, the quantities

$$\limsup_{\epsilon \to 0} \frac{|E_\epsilon \setminus E|}{\epsilon} \qquad \text{and} \qquad \liminf_{\epsilon \to 0} \frac{|E_\epsilon \setminus E|}{\epsilon},$$

and prove the estimates

$$P_{\mathbb{H}}(E) \le \mathcal{M}_3^-(\partial E) \tag{6.14}$$

and

$$P_{\mathbb{H}}(E) \ge \mathcal{M}_3^+(\partial E) \tag{6.15}$$

which give the desired conclusion.

The estimate in (6.14) follows directly from the lower semi-continuity of the perimeter, as we now show. We introduce the signed distance function

$$\rho(x) = \begin{cases} \operatorname{dist}(x, \partial E), & \text{if } x \in E, \\ -\operatorname{dist}(x, \partial E), & \text{if } x \in \mathbb{H} \setminus E, \end{cases}$$

and consider the sequence of normalized Lipschitz truncations

$$\phi_\epsilon(x) = \begin{cases} \frac{1+\rho(x)}{2\epsilon}, & \text{if } -\epsilon \le \rho(x) \le \epsilon, \\ 1, & \text{if } \rho(x) \ge \epsilon, \\ 0, & \text{if } \rho(x) \le -\epsilon. \end{cases}$$

A quick computation gives

$$||\phi_\epsilon||_{BV} = \frac{1}{2\epsilon} \int_{|\rho| \le \epsilon} |\nabla_0 \rho| \le \frac{|\{x \in \mathbb{H} : -\epsilon < \rho(x) < \epsilon\}|}{2\epsilon}$$

whence

$$P_{\mathbb{H}}(E) \le \liminf_{\epsilon \to 0} ||\phi_\epsilon||_{BV} \le \mathcal{M}_3^-(\partial E)$$

by Proposition 5.6.

The proof of (6.15) uses the Riemannian approximants $(\mathbb{R}^3, g_L)$. We denote by d_L, resp. ρ_L, resp. $\mathcal{M}^{\pm}_{3,L}(\partial E)$, the associated metric on $\mathbb{R}^3$, signed distance function to ∂E, and upper and lower Minkowski 3-contents of ∂E. By an application of the Riemannian co-area formula and arguing as in the Euclidean case, we deduce the identity

$$\mathcal{M}^{+}_{3,L}(\partial E) = \frac{1}{\sqrt{L}} \int_{\partial E} d\sigma_L,$$

where $d\sigma_L$ denotes the Riemannian surface area element for the metric g_L. Since $\rho \geq \rho_L$ for each L we find

$$\mathcal{M}^{+}_{3}(\partial E) \leq \lim_{L\to\infty} \mathcal{M}^{+}_{3,L}(\partial E) = \lim_{L\to\infty} \frac{1}{\sqrt{L}} \int_{\partial E} d\sigma_L = P_{\mathbb{H}}(E)$$

by (5.6). □

The boundary of a Carnot–Carathéodory ball in $\mathbb{H}$ (centered at the origin, say) is not of class C^2 due to the presence of singularities on the x_3-axis. However, the equality of the previous proposition continues to hold for such sets.

Proposition 6.18. $P_{\mathbb{H}}(B) = \mathcal{M}_3(\partial B)$ *for every CC-ball* $B \subset \mathbb{H}$.

Proof. We may assume that $B = B(o,1)$. Note that $B_\epsilon = B(o, 1+\epsilon)$. Combining the sub-Riemannian co-area formula (6.7) and the eikonal equation (6.12), and using the scaling properties of the horizontal perimeter, we obtain

$$\begin{aligned} |B_\epsilon \setminus B| &= \int_{B_\epsilon \setminus B} |\nabla_0 \operatorname{dist}(\cdot, \partial B)| = \int_0^\epsilon P_{\mathbb{H}}(B(o, 1+t))\, dt \\ &= P_{\mathbb{H}}(B) \int_0^\epsilon (1+t)^3\, dt = P_{\mathbb{H}}(B) \cdot \frac{(1+\epsilon)^4 - 1}{4} \end{aligned}$$

whence

$$\mathcal{M}_3(\partial B) = \lim_{\epsilon\to 0} \frac{|B_\epsilon \setminus B|}{\epsilon} = P_{\mathbb{H}}(B). \qquad \square$$

6.4 First variation of the perimeter

In this section we present two derivations of the first variation formula for the perimeter of surfaces in $\mathbb{H}$. First we consider parametric representations of the surface and variations which vanish in a neighborhood of the characteristic set. Next, we present an argument in which variations over the full surface are allowed.

We start by introducing some useful notation: Let $\Omega \subset \mathbb{H}$ be a bounded region enclosed by a surface S, and consider variations Ω_t and S_t along a given vector field $Z \in C^1(\mathbb{H}, \mathbb{R}^3)$. We say that Ω is *perimeter stationary* (or area stationary) if $d/dt P_{\mathbb{H}}(\Omega_t)\,|_{t=0} = 0$ for all choices Z. We say that Ω is *volume-preserving perimeter stationary* if $d/dt P_{\mathbb{H}}(\Omega_t)\,|_{t=0} = 0$ for all Z such that $d/dt|\Omega_t|\,|_{t=0} = 0$. Note that if Z is tangent to S at every point, then variations along Z do not change $S = S_t$, hence tangential variations always correspond to $d/dt P_{\mathbb{H}}(\Omega_t)\,|_{t=0} = 0$.

6.4.1 Parametric surfaces and noncharacteristic variations

We compute the first variation of the perimeter with respect to the Euclidean normal $\vec{n}$ and then with respect to a horizontal frame. Throughout the section the emphasis is on explicit computations using the underlying Euclidean structure of $\mathbb{R}^3$. Let $\Omega \subset \mathbb{R}^2$ be a domain, let $\epsilon > 0$, and let $\Xi : \Omega \times (-\epsilon, \epsilon) \to \mathbb{H}$. We view Ξ as representing a flow $S_t = \Xi(\Omega, t)$ of noncharacteristic surface patches in $\mathbb{H}$, with $\Xi(\Omega, 0) = S$. Denote by A the 2×3 matrix of coefficients of the horizontal frame X_1, X_2. Recall that we have agreed to write $P_{\mathbb{H}}(S_t)$ for the horizontal perimeter of the surface patch S_t. Our goal is to evaluate

$$\begin{aligned}\left.\frac{d}{dt}P_{\mathbb{H}}(S_t)\right|_{t=0} &= \left.\frac{d}{dt}\int_{S_t}|A\vec{n}|\,d\sigma\right|_{t=0}\\ &= \left.\frac{d}{dt}\int_{\Omega}|A(\Xi_u \times \Xi_v)|\,dudv\right|_{t=0}\\ &= \int_{\Omega}\frac{1}{|AV|}\left.\left(\left\langle AV, A\frac{dV}{dt}\right\rangle + \left\langle AV, \left(\frac{dA}{dt}\right)V\right\rangle\right)\right|_{t=0},\end{aligned} \tag{6.16}$$

where we have let $V = \Xi_u \times \Xi_v$ and $\vec{n} = V/|V|$. We want to stress once more that we are denoting by $\langle\cdot,\cdot\rangle$ and $|\cdot|$ the Euclidean inner product and the corresponding norm.

To simplify the notation we will always ignore the higher order terms in the Taylor expansion of Ξ, writing

$$\Xi(u,v,t) = \Xi(u,v,0) + t\frac{d\Xi}{dt}(u,v,0) =: X(u,v) + t\lambda\vec{n}(u,v,0).$$

A simple computation yields $V = X_u \times X_v + t(X_u \times (\lambda\vec{n})_v + (\lambda\vec{n})_u \times X_v)$, and consequently

$$\left.\frac{dV}{dt}\right|_{t=0} = \lambda(X_u \times \vec{n}_v + \vec{n}_u \times X_v) + \lambda_v X_u \times \vec{n} + \lambda_u \vec{n} \times X_v.$$

Recall that $\vec{n}_\alpha = -b_\alpha^\beta X_\beta$, with $b_\alpha^\beta = b_{\alpha,\gamma}g^{\beta,\gamma}$, where $b_{\alpha,\gamma}$ are the coefficients of the second fundamental form and $\{g^{\beta,\gamma}\}$ is the inverse of the Riemannian metric on S induced by the Euclidean metric in $\mathbb{C} \times \mathbb{R}$. Since $X_u \times \vec{n}_v = -b_v^v X_u \times X_v$, and $\vec{n}_u \times X_v = -b_u^u X_u \times X_v$,

$$\left.\frac{dV}{dt}\right|_{t=0} = \lambda\mathcal{H}X_u \times X_v + \lambda_v X_u \times \vec{n} + \lambda_u \vec{n} \times X_v, \tag{6.17}$$

where $\mathcal{H} = b_u^u + b_v^v$ denotes the mean curvature of S.

Since

$$A|_\Xi = \begin{pmatrix} 1 & 0 & -\frac{1}{2}\Xi_2 \\ 0 & 1 & \frac{1}{2}\Xi_1 \end{pmatrix}$$

we find
$$\left.\frac{dA}{dt}\right|_{\Xi} = \begin{pmatrix} 0 & 0 & -\frac{1}{2}\lambda n_2 \\ 0 & 0 & \frac{1}{2}\lambda n_1 \end{pmatrix}, \qquad \vec{n} = (n_1, n_2, n_3).$$
Thus
$$\frac{dA}{dt}V = \frac{1}{2}\lambda V_3(-n_2, n_1), \qquad V = (V_1, V_2, V_3). \tag{6.18}$$
By virtue of (6.16), (6.17) and (6.18) we obtain

$$\begin{aligned}\left.\frac{d}{dt}P_{\mathbb{H}}(S_t)\right|_{t=0} = \int_\Omega \frac{1}{|AV|}\Big\langle AV, \lambda\mathcal{H}A(X_u \times X_v) \\ +\lambda_v A(X_u \times \vec{n}) + \lambda_u A(\vec{n} \times X_v) + \frac{1}{2}\lambda V_3(-n_2, n_1)\Big\rangle\Big|_{t=0}\, dudv. \end{aligned}\tag{6.19}$$

Integrating by parts in u and v gives the following expression for the first variation of the perimeter:

$$\begin{aligned}\int_\Omega \lambda\Bigg[\mathcal{H}|AV| - \left(\frac{\langle AV, A(X_u \times \vec{n})\rangle}{|AV|}\right)_v + \left(\frac{\langle AV, A(X_v \times \vec{n})\rangle}{|AV|}\right)_u \\ + \frac{\frac{1}{2}V_3}{|AV|}\langle AV, (-n_2, n_1)\rangle\Bigg]\, dudv.\end{aligned}\tag{6.20}$$

In essence, (6.20) is the derivative of the perimeter functional in the direction $Y = \lambda\vec{n}$. Since variations along purely tangential directions are zero, then normal variations represent the complete "gradient" of the perimeter functional.

We now want to generalize the preceding to the case of general perturbations of the original surface S. Thus consider a variation of the form

$$\Xi = \mathcal{X} + t(aX_1 + bX_2 + cT) + o(t),$$

where $\mathcal{X} = (x_1, x_2, x_3)$ is a parameterization of S, $X_1 = (1, 0, -\frac{1}{2}x_2)$ and $X_2 = (0, 1, \frac{1}{2}x_1)$ are the standard left invariant basis of the horizontal bundle, and $T = (\frac{1}{2}x_2, -\frac{1}{2}x_1, 1)$ (note that T differs from the standard left invariant vector field $X_3 = (0, 0, 1)$ in the Heisenberg group). In order to compute $\Xi_u \times \Xi_v$ we need to calculate

$$\begin{aligned}
\mathcal{X}_\alpha \times X_1 &= \left(-\frac{1}{2}x_2x_{2,\alpha}, \frac{1}{2}x_2x_{1,\alpha} + x_{3,\alpha}, -x_{2,\alpha}\right),\\
\mathcal{X}_\alpha \times X_2 &= \left(-x_{3,\alpha} + \frac{1}{2}x_1x_{2,\alpha},\ -\frac{1}{2}x_1x_{1,\alpha},\ x_{1,\alpha}\right),\\
X_{1,\alpha} \times \mathcal{X}_\beta &= \left(\frac{1}{2}x_{2,\alpha}x_{2,\beta},\ -\frac{1}{2}x_{2\alpha}x_{1,\beta},\ 0\right),\\
X_{2,\alpha} \times \mathcal{X}_\beta &= \left(-\frac{1}{2}x_{1,\alpha}x_{2,\beta},\ \frac{1}{2}x_{1,\alpha}x_{1,\beta},\ 0\right),
\end{aligned}$$

$$
\begin{aligned}
\mathcal{X}_\alpha \times T &= \left(x_{2,\alpha} + \frac{1}{2} x_1 x_{3,\alpha},\ \frac{1}{2} x_2 x_{3,\alpha} - x_{1,\alpha},\ -\frac{1}{2} x_1 x_{1,\alpha} - \frac{1}{2} x_2 x_{2,\alpha} \right), \\
\mathcal{X}_\beta \times T_\alpha &= \left(-\frac{1}{2} x_{1,\alpha} x_{3,\beta},\ \frac{1}{2} x_{2,\alpha} x_{3,\beta},\ -\frac{1}{2} x_{1,\alpha} x_{1,\beta} - \frac{1}{2} x_{2,\alpha} x_{2,\beta} \right).
\end{aligned}
$$

Here $\alpha, \beta \in \{u, v\}$ and we have written $x_{i,\alpha} = \partial x_i / \partial \alpha$, etc. These in turn yield

$$
\begin{aligned}
A(\mathcal{X}_\alpha \times X_1) &= (0, \omega(\mathcal{X}_\alpha)), \\
A(\mathcal{X}_\alpha \times X_2) &= (-\omega(\mathcal{X}_\alpha), 0), \\
A(X_{1,\alpha} \times \mathcal{X}_\beta) &= \frac{1}{2} x_{2,\alpha}\, \mathbf{i}\, z_\beta, \\
A(X_{2,\alpha} \times \mathcal{X}_\beta) &= -\frac{1}{2} x_{1,\alpha}\, \mathbf{i}\, z_\beta, \\
A(\mathcal{X}_\alpha \times X_3) &= \mathbf{i}\, z_\alpha, \\
A(\mathcal{X}_\beta \times X_{3,\alpha}) &= 2\omega(\mathcal{X}_\beta) z_\alpha + \frac{1}{2} (|z|^2)_\alpha\, \mathbf{i}\, z_\beta,
\end{aligned} \tag{6.21}
$$

where we have let $z = x_1 + \mathbf{i}\, x_2$ and $\omega(\mathcal{X}_\alpha) = x_{3,\alpha} + \frac{1}{2}(x_2 x_{1,\alpha} - x_1 x_{2,\alpha})$.

Thanks to (6.21) we have

$$
\begin{aligned}
A \frac{d}{dt} (\Xi_u \times \Xi_v) &= a_v (0, \omega(\mathcal{X}_u)) - a_u (0, \omega(\mathcal{X}_v)) + \frac{1}{2} a\, \mathbf{i}\, (x_{2u} z_u - x_{2v} z_v) \\
&\quad + b_v(-\omega(\mathcal{X}_u), 0) - b_u(-\omega(\mathcal{X}_v), 0) + \frac{1}{2} b\, \mathbf{i}\, (x_{1v} z_u - x_{1u} z_v) \\
&\quad + c \left[\frac{1}{2} \omega(\mathcal{X}_u) z_v - \frac{1}{2} \omega(\mathcal{X}_v) z_u + \frac{1}{2} (|z|^2)_v\, \mathbf{i}\, z_u - \frac{1}{2} (|z|^2)_u\, \mathbf{i}\, z_v \right] \\
&\quad + c_v \left[(1 + \frac{1}{4}|z|^2)\, \mathbf{i}\, z_u + \frac{1}{2} \omega(\mathcal{X}_u) z \right] - c_u \left[(1 + \frac{1}{4}|z|^2)\, \mathbf{i}\, z_v + \frac{1}{2} \omega(\mathcal{X}_v) z \right].
\end{aligned} \tag{6.22}
$$

Set $\mathcal{F} = \{X_1, X_2\}$ the horizontal, orthonormal frame and express the horizontal normal ν_H in this frame as the coordinate vector

$$
[\nu_H]_{\mathcal{F}} = \frac{A(\mathcal{X}_u \times \mathcal{X}_v)}{|A(\mathcal{X}_u \times \mathcal{X}_v)|} = (\nu_H^1, \nu_H^2). \tag{6.23}
$$

Note that

$$
A(\mathcal{X}_u \times \mathcal{X}_v) = \mathbf{i}\, [\omega(\mathcal{X}_u) z_v - \omega(\mathcal{X}_v) z_u]. \tag{6.24}
$$

Using the latter and (6.22) we have (after several cancellations)

$$
\begin{aligned}
\left\langle [\nu_H]_{\mathcal{F}}, A \frac{d}{ds} (\Xi_u \times \Xi_v) \right\rangle &= a \Big(\langle [\nu_H]_{\mathcal{F}}, (0, \omega(\mathcal{X}_v)) \rangle_u - \langle [\nu_H]_{\mathcal{F}}, (0, \omega(\mathcal{X}_u)) \rangle_v \Big) \\
&\quad + b \Big(\langle [\nu_H]_{\mathcal{F}}, (\omega(\mathcal{X}_u), 0) \rangle_v - \langle [\nu_H]_{\mathcal{F}}, (\omega(\mathcal{X}_v), 0) \rangle_u \Big) \\
&\quad + c \Big(\langle \partial_u [\nu_H]_{\mathcal{F}}, (1 + \frac{1}{4}|z|^2) \mathbf{i} z_v + \frac{1}{2} \omega(\mathcal{X}_v) z \rangle - \langle \partial_v [\nu_H]_{\mathcal{F}}, (1 + \frac{1}{4}|z|^2) \mathbf{i} z_u + \frac{1}{2} \omega(\mathcal{X}_u) z \rangle \Big).
\end{aligned} \tag{6.25}
$$

On the other hand,

$$\left(\frac{dA}{dt}\right)\mathcal{X}_u \times \mathcal{X}_v = \frac{1}{2}V_3(-Y_2, Y_1),$$

where we have set $V = \mathcal{X}_u \times \mathcal{X}_v$ and $Y = (Y_1, Y_2, Y_3) \in \mathbb{R}^3$. We have

$$(-Y_2, Y_1) = a(0,1) + b(-1,0) + \frac{1}{2}cz.$$

Combining the latter with (6.16) and (6.25) we finally obtain the first variation of the perimeter:

Variation along X_1 ($a = 1$, $b = c = 0$):

$$[\nu_H^2 \omega(\mathcal{X}_v)]_u - [\nu_H^2 \omega(\mathcal{X}_u)]_v + \frac{1}{2}V_3 \nu_H^2. \tag{6.26}$$

Variation along X_2 ($b = 1$, $a = c = 0$):

$$[\nu_H^1 \omega(\mathcal{X}_u)]_v - [\nu_H^1 \omega(\mathcal{X}_v)]_u - \frac{1}{2}V_3 \nu_H^1. \tag{6.27}$$

Variation along X_3 ($c = 1$, $a = b = 0$):

$$\begin{aligned} &\left\langle \partial_u[\nu_H]_{\mathcal{F}}, \left(1 + \frac{1}{4}|z|^2\right) \mathbf{i}\, z_v + \frac{1}{2}\omega(\mathcal{X}_v)z \right\rangle \\ &\qquad - \left\langle \partial_v[\nu_H]_{\mathcal{F}}, \left(1 + \frac{1}{4}|z|^2\right) \mathbf{i}\, z_u + \frac{1}{2}\omega(\mathcal{X}_u)z \right\rangle + \frac{1}{4}V_3 \langle [\nu_H]_{\mathcal{F}}, z\rangle. \end{aligned} \tag{6.28}$$

At this point we restrict our attention to horizontal variations by setting $c = 0$. The maximum variation, among horizontal variations, is the one corresponding to a flow in the direction $aX_1 + bX_2$, where the vector (a, b) is chosen according to (6.26) and (6.27), i.e., is given by

$$\left([\nu_H^2\omega(\mathcal{X}_v)]_u - [\nu_H^2\omega(\mathcal{X}_u)]_v + \frac{1}{2}V_3\nu_H^2 \;;\; [\nu_H^1\omega(\mathcal{X}_u)]_v - [\nu_H^1\omega(\mathcal{X}_v)]_u - \frac{1}{2}V_3\nu_H^1\right). \tag{6.29}$$

Inserting these in (6.26) and (6.27) leads to the following

Proposition 6.19. *The maximum variation of the perimeter among horizontal variations and outside the characteristic locus is obtained along the direction $aX_1 + bX_2$, where (a, b) is chosen as in* (6.29)*, and is equal to $I + II$, where*

$$I = \left(\omega(\mathcal{X}_u)_v - \omega(\mathcal{X}_v)_u - \frac{1}{2}V_3\right) \mathbf{i}\,[\nu_H]_{\mathcal{F}}$$

and

$$II = \omega(\mathcal{X}_u)\partial_v(\mathbf{i}\,[\nu_H]_{\mathcal{F}}) - \omega(\mathcal{X}_v)\partial_u(\mathbf{i}\,[\nu_H]_{\mathcal{F}}), \tag{6.30}$$

here $[\nu_H]_{\mathcal{F}}$ is as in (6.23)*. Moreover, the component corresponding to I is tangent to the surface.*

Proof. We are only left with the proof of the last statement. Since $[\nu_H]_{\mathcal{F}}$ is a unit vector, the factors I and II are orthogonal. From

$$-\nu_H^2 X_1 + \nu_H^1 X_2 = A^T\,\mathbf{i}\,A(\mathcal{X}_u \times \mathcal{X}_v)/|A(\mathcal{X}_u \times \mathcal{X}_v)|$$

we obtain

$$\left\langle \vec{n}, -\nu_H^2 X_1 + \nu_H^1 X_2 \right\rangle = \left\langle \frac{(\mathcal{X}_u \times \mathcal{X}_v)}{|(\mathcal{X}_u \times \mathcal{X}_v)|}, \frac{A^T\,\mathbf{i}\,A(\mathcal{X}_u \times \mathcal{X}_v)}{|A(\mathcal{X}_u \times \mathcal{X}_v)|} \right\rangle = 0. \qquad \square$$

Remark 6.20. Observe that if we can choose a parametrization of the surface so that z_u and z_v are orthonormal and use the fact that $V_3 = \langle \mathbf{i}\, z_u, z_v \rangle$, we can rewrite I more explicitly in the following form:

$$I = \left[\frac{1}{4}\langle \mathbf{i}\, z_u, z_v \rangle - \frac{1}{2} V_3\right] \; \mathbf{i}\,[\nu_H]_{\mathcal{F}} = -\frac{1}{4} V_3\, \mathbf{i}\,[\nu_H]_{\mathcal{F}}.$$

Since we just proved that the component of the variation corresponding to I is horizontal and *tangent* to the surface, it thus corresponds only to a reparametrization of the surface. In the following we ignore it and focus on the component II.

Proposition 6.21. *The non-tangential component of the maximal horizontal variation of the perimeter $P_{\mathbb{H}}$ occurs along the vector $Z = aX_1 + bX_2$, where*

$$(a, b) = 2\mathcal{H}_0 |A(\mathcal{X}_u \times \mathcal{X}_v)| \; [\nu_H]_{\mathcal{F}},$$

i.e., $Z = 2\mathcal{H}_0 \nu_H |\pi_H(\nu_1)|_1$ with π_H denoting the (Euclidean) orthogonal projection on the horizontal bundle.

Proof. To identify II, we consider the Legendrian foliation of the surface. Recall that this foliation is composed of horizontal curves $\tilde{\gamma}$ lying on the surface which are flow lines of the horizontal vector field

$$\mathbf{i}\,[\nu_H]_{\mathcal{F}} \cdot \nabla_0 := -\nu_H^2 X_1 + \nu_H^1 X_2.$$

Since we are using a parametric representation of the surface S, we consider a curve $\gamma = (u, v) : [0, L] \to \mathbb{R}^2$, such that $\tilde{\gamma}(s) = \mathcal{X}(\gamma(s)) = \mathcal{X}(u(s), v(s))$.

Note that

$$\frac{d}{ds}\pi_z \tilde{\gamma}(s) = z_u u' + z_v v'.$$

On the other hand, by (6.24), and by the definition of the Legendrian foliation we also have

$$\frac{d}{ds}\pi_z \tilde{\gamma}(s) = \mathbf{i}\,[\nu_H]_{\mathcal{F}} = \frac{\omega(\mathcal{X}_v) z_u - \omega(\mathcal{X}_u) z_v}{|\omega(\mathcal{X}_u) z_v - \omega(\mathcal{X}_v) z_u|}.$$

As a consequence,

$$u' = \frac{w(\mathcal{X}_v)}{|\omega(\mathcal{X}_u) z_v - \omega(\mathcal{X}_v) z_u|}, \qquad \text{and} \qquad v' = -\frac{w(\mathcal{X}_u)}{|\omega(\mathcal{X}_u) z_v - \omega(\mathcal{X}_v) z_u|}.$$

Now, by virtue of Lemma 4.24 we have that the curvature of $\pi_z\tilde{\gamma}$ is given by $-2\mathcal{H}_0[\nu_H]_{\mathcal{F}}$, hence

$$\begin{aligned}
-2\mathcal{H}_0[\nu_H]_{\mathcal{F}} = \frac{d^2}{ds^2}\pi_z\tilde{\gamma} &= \frac{d}{ds}(\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})(u(s),v(s)) \\
&= (\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})_u u' + (\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})_v v' \\
&= \frac{(\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})_u \omega(\mathcal{X}_v) - (\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})_v \omega(\mathcal{X}_u)}{|A(\mathcal{X}_u \times \mathcal{X}_v)|} \qquad (6.31)\\
&= \frac{-II}{|A(\mathcal{X}_u \times \mathcal{X}_v)|}. \qquad \square
\end{aligned}$$

Formula (6.31) provides an explicit representation of the horizontal mean curvature for noncharacteristic parametric surface.

Proposition 6.22. *If $S \subset \mathbb{H}$ is a C^2 surface parametrized by the map $\mathcal{X} : \Omega \to \mathbb{H}$, then outside the characteristic set $\Sigma(S)$ one has*

$$\mathcal{H}_0[\nu_H]_{\mathcal{F}} = \frac{\omega(\mathcal{X}_u)\partial_v(\,\mathbf{i}\,[\nu_H]_{\mathcal{F}}) - \omega(\mathcal{X}_v)\partial_u(\,\mathbf{i}\,[\nu_H]_{\mathcal{F}})}{|A(\mathcal{X}_u \times \mathcal{X}_v)|}, \tag{6.32}$$

where $[\nu_H]_{\mathcal{F}}$ is given by (6.23) *and ω is the contact form.*

Remark 6.23. Since tangential variations result in reparametrizations of the surface, which do not modify the perimeter, it is important that we consider only the normal component of the maximal horizontal variation identified above, i.e.,

$$\begin{aligned}
&\left\langle A^T\left(2\mathcal{H}_0|A(\mathcal{X}_u \times \mathcal{X}_v)|\; [\nu_H]_{\mathcal{F}} + \frac{1}{2}\langle\,\mathbf{i}\, z_u, z_v\rangle\;\, \mathbf{i}\,[\nu_H]_{\mathcal{F}}\right),\; \vec{n}\right\rangle \\
&\qquad = 2\mathcal{H}_0\frac{|A(\mathcal{X}_u \times \mathcal{X}_v)|}{|\mathcal{X}_u \times \mathcal{X}_v|} = 2\mathcal{H}_0|A\vec{n}| = 2\mathcal{H}_0|\pi_H(\nu_1)|_1.
\end{aligned}$$

6.4.2 General variations

We present an extension of the first variation formula which allows for variations across the characteristic locus.

Proposition 6.24. *Let $S \subset \mathbb{H}$ be an oriented C^2 immersed surface with g_1-Riemannian normal ν_1 and horizontal normal ν_H. Suppose that U is a C^1 vector field with compact support on S, let $\phi_t(p) = \exp_p(tU)$ and let S_t be the surface $\phi_t(S)$. Then*

$$\frac{d}{dt}P_{\mathbb{H}}(S_t)\bigg|_{t=0} = \int_{S\setminus\Sigma(S)} \left[u(\operatorname{div}_S \nu_H) - \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}})\right] d\sigma, \tag{6.33}$$

where $u = \langle U, \nu_1\rangle_1$. Moreover, if $\operatorname{div}_S \nu_H \in L^1(S, d\sigma)$, then

$$\frac{d}{dt}P_{\mathbb{H}}(S_t)\bigg|_{t=0} = \int_S u(\operatorname{div}_S \nu_H)\, d\sigma - \int_S \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}})\, d\sigma, \tag{6.34}$$

where $u = \langle U, \nu_1\rangle_1$.

Here v_{tang} denotes the tangential component of v and $d\sigma$ denotes the Riemannian surface area element on S. We will denote by N_H the g_1 projection of ν_1 on the horizontal distribution, $N_H = \sum_{i=1}^2 \langle \nu_1, X_i\rangle_1 X_i$. Recall that in view of Corollary 5.8, $|N_H| = \langle \nu_H, \nu_1\rangle_1$ is the density of the perimeter measure.

Proof. Let $d\sigma_t$ be the Riemannian surface area element on the surface S_t. By abuse of notation, we will also denote by ν_1 an extension of the normal to S to all of $\mathbb{H}$, so that $\nu_1|_{S_t}$ is the Riemannian normal to S_t. Then

$$\begin{aligned} P_{\mathbb{H}}(S_t) &= \int_{S_t} |N_H|\, d\sigma_t \\ &= \int_S (|N_H| \circ \phi_t)|J_{\phi_t}|\, d\sigma \\ &= \int_{S\setminus\Sigma(S)} (|N_H| \circ \phi_t)|J_{\phi_t}|\, d\sigma \end{aligned}$$

by the Riemannian co-area formula. The last equality uses the result of Derridj [86] (later extended by [20] and [189, 190]) according to which the Riemannian surface measure of $\Sigma(S)$ is zero.

In the next computation, we use the divergence identity

$$\frac{d}{dt}|J_{\phi_t}|\bigg|_{t=0} = \operatorname{div}_S U \tag{6.35}$$

and the standard formula

$$\operatorname{div}_S(fV) = f \operatorname{div}_S V + V_{\text{tang}}(f) \tag{6.36}$$

for $f : S \to \mathbb{R}$ and a C^1 vector field V on S. Here, as before, V_{tang} denotes the tangential component of V. We will use the notation V_{norm} to denote the (Riemannian) normal component of V. Differentiating with respect to t and using (6.35) and (6.36) gives

$$\begin{aligned} \frac{d}{dt}P_{\mathbb{H}}(S_t)\bigg|_{t=0} &= \int_{S\setminus\Sigma(S)} (U(|N_H|) + |N_H| \operatorname{div}_s U)\, d\sigma \\ = &\int_{S\setminus\Sigma(S)} (U_{\text{norm}}(|N_H|) + |N_H| \operatorname{div}_S(U_{\text{norm}}) + \operatorname{div}_S(|N_H| U_{\text{tang}}))\, d\sigma \\ = &\int_{S\setminus\Sigma(S)} (U_{\text{norm}}(|N_H|) + |N_H| \operatorname{div}_S(U_{\text{norm}}))\, d\sigma. \end{aligned} \tag{6.37}$$

In the last line, we use the Riemannian divergence formula to conclude that the integral of the divergence of the compactly supported Lipschitz vector field $|N_H| U_{\text{tang}}$ is zero.

Noting that $N_H = \nu_1 - \langle \nu_1, X_3\rangle_1 X_3$ one has

$$\begin{aligned} V(|N_H|) =& \langle D_V N_H, \nu_H\rangle_1 \\ =& \langle D_V N_H, \nu_H\rangle_1 + \langle X_3, \nu_H\rangle_1 V(\langle \nu_1, X_3\rangle_1) \\ =& \langle D_V \nu_1, \nu_H\rangle_1 - \langle \nu_1, X_3\rangle_1 \langle D_V X_3, \nu_H\rangle_1. \end{aligned}$$

Using $D_{U_{\mathrm{norm}}}\nu_1 = -\nabla_S u$, and $\langle D_{U_{\mathrm{norm}}} X_3, \nu_H\rangle_1 = 0$, we have

$$U_{\mathrm{norm}}(|N_H|) = \langle D_{U_{\mathrm{norm}}}\nu_1, \nu_H\rangle_1 - \langle \nu_1, X_3\rangle_1 \langle D_{U_{\mathrm{norm}}} X_3, \nu_H\rangle_1 = -\langle \nabla_S u, \nu_H\rangle_1.$$

Thus the integrand in the last line of (6.37) is

$$\begin{aligned} &U_{\mathrm{norm}}(|N_H|) + |N_H| \operatorname{div}_S(U_{\mathrm{norm}}) \\ &\quad = \langle -\nabla_S u, \nu_H\rangle_1 + |N_H| \operatorname{div}_S U_{\mathrm{norm}} \\ &\quad = -(\nu_H)_{\mathrm{tang}}(u) + u|N_H| \operatorname{div}_S \nu_1 \\ &\quad = -\operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) + u \operatorname{div}_S((\nu_H)_{\mathrm{tang}}) + u \operatorname{div}_S(|N_H|\nu_1) \\ &\quad = -\operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) + u \operatorname{div}_S \nu_H. \end{aligned} \tag{6.38}$$

Noting that

$$U_{\mathrm{norm}}(|N_H|) + |N_H| \operatorname{div}_S(U_{\mathrm{norm}}))$$

is compactly supported and bounded on S (and hence in $L^1(S, d\sigma)$), we obtain (6.33). Under the additional assumption $\operatorname{div}_S \nu_H \in L^1(S, d\sigma)$, we conclude that $\operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) \in L^1(S, d\sigma)$ as well. Substituting (6.38) into (6.37), we obtain (6.34), thus concluding the proof. □

Remark 6.25. Through an approximation by Lipschitz vector fields, the hypotheses can be slightly weakened: we may consider compactly supported, bounded vector fields U on S such that $|N_H| \operatorname{div}_S(U_{\mathrm{norm}}) \in L^1(S)$, $|N_H| U_{\mathrm{tang}}$ is Lipschitz continuous, and $\operatorname{div}_S \nu_H \in L^1(S, u d\sigma)$. This observation is important as it allows us to consider horizontal variations $U = a\nu_H$, $a \in C^1_0(S)$ which correspond to $u d\sigma = a d\mu$; in this case the assumptions are essentially reduced to the requirement $\mathcal{H}_0 \in L^1(S, d\mu)$. In view of (4.21), the latter condition is always true for C^2 surfaces.

Remark 6.26. If U is compactly supported in $S \setminus \Sigma(S)$, then the (6.33) reduces to

$$\frac{d}{dt} P_{\mathbb{H}}(S_t)\bigg|_{t=0} = \int_S u(\operatorname{div}_S \nu_H)\, d\sigma$$

with $u = \langle U, \nu_1\rangle_1$.

6.5 Mostow's rigidity theorem for $H^2_{\mathbb{C}}$

The relevance of quasiconformal mappings on the Heisenberg group (and other Carnot groups) stems in large part from their role in the proof of Mostow's celebrated rigidity theorem for lattices in the complex hyperbolic space $H^2_{\mathbb{C}}$ (and other rank 1 symmetric spaces). In this section, we give a brief summary and the main lines of the proof of Mostow's theorem in this case, and use this as motivation to summarize some aspects of the rich theory of quasiconformal maps in $\mathbb{H}$ as developed by Korányi, Reimann, Pansu and others.

As this section is somewhat tangential to the main direction of the text, we are deliberately more expository and make no attempt to include full details of all of the proofs. Consequently, we break from our standard convention here and include references to the literature within the text as well as in the notes.

Recall that $SU(1,2)$ acts[1] on $H^2_{\mathbb{C}}$ with isotropy subgroup $U(2)$; thus $H^2_{\mathbb{C}}$ is naturally realized as the symmetric space $SU(1,2)/U(2)$. Let Γ be a lattice in $H^2_{\mathbb{C}}$, e.g., a discrete subgroup of the isometry group $\operatorname{Isom} H^2_{\mathbb{C}}$ such that the quotient manifold $M_\Gamma := H^2_{\mathbb{C}}/\Gamma$ has finite volume. Mostow's rigidity theorem asserts, in essence, that the algebraic structure of Γ determines the metric structure of M_Γ in a strong sense:

Theorem 6.27 (Mostow). *Let Γ, Γ' be two lattices in $H^2_{\mathbb{C}}$ with corresponding quotients $M_\Gamma = H^2_{\mathbb{C}}/\Gamma$ and $M_{\Gamma'} = H^2_{\mathbb{C}}/\Gamma'$. If Γ and Γ' are isomorphic, then M_Γ and $M_{\Gamma'}$ are isometric, in particular, they are conformally equivalent.*

As a point of comparison, note that the corresponding theorem for the *real* hyperbolic space $H^2_{\mathbb{R}}$ is false; indeed, the study of Teichmüller space (the moduli space of marked conformal structures on a Riemann surface $S = H^2_{\mathbb{R}}/\Gamma$) is, according to one point of view, the study of the failure of Mostow rigidity in this setting.

To simplify matters, we will only discuss the proof of Theorem 6.27 in the case of *cocompact* lattices, e.g., when M_Γ is compact. The proof breaks into two main parts:

I. large scale (e.g., coarse/Gromov hyperbolic) geometry of $H^2_{\mathbb{C}}$,
II. quasiconformal analysis on $\partial_\infty H^2_{\mathbb{C}}$.

Part I. Choose an isomorphism $f : \Gamma \to \Gamma'$. Since M_Γ and M'_Γ are compact, we may find precompact fundamental domains $U_\Gamma, U_{\Gamma'} \subset H^2_{\mathbb{C}}$ for the action of Γ and Γ', respectively. We work with the ball model for $H^2_{\mathbb{C}}$, using the metric in (3.18). The map f induces an equivariant map $F : H^2_{\mathbb{C}} \to H^2_{\mathbb{C}}$ which is a quasi-isometry, that is,

$$A^{-1} d(p,q) - B \le d(F(p), F(q)) \le A d(p,q) + B \tag{6.39}$$

for all $p, q \in H^2_{\mathbb{C}}$ and some (fixed) $A, B > 0$. The map F can be defined, for example, by sending an arbitrary point $p \in H^2_{\mathbb{C}}$ to $f(\gamma) \cdot 0$, where $\gamma \in \Gamma$ is chosen

[1] This action can be seen using the projective model $H^2_{\mathbb{C}} = \{[y] \in \mathbb{C}P^2 : Q(y,y) < 0\}$, $Q(x,y) = \overline{x_1} y_1 + \overline{x_2} y_2 - \overline{x_3} y_3$, and nonhomogeneous coordinates $w_i = y_i/y_3$, $i = 1,2$; the action of $g = (g_{ij}) \in SU(1,2)$ on $(w_1, w_2) \in \{w \in \mathbb{C}^2 : |w| < 1\}$ is given by

$$w'_j = \frac{g_{j1} w_1 + g_{j2} w_2 + g_{j3}}{g_{31} w_1 + g_{32} w_2 + g_{33}}, \qquad j = 1, 2.$$

so that $\gamma^{-1} \cdot p \in U_\Gamma$. Note that such a map F need not even be continuous, but is "bi-Lipschitz on large scales", e.g., for points $p, q \in H^2_{\mathbb{C}}$ with $d(p, q) \gg 1$. In particular, geodesic rays in $H^2_{\mathbb{C}}$ (emanating from the origin, say) are taken to quasi-geodesics (images $\gamma = F([0, +\infty))$ by maps $F : [0, +\infty) \to H^2_{\mathbb{C}}$ satisfying (6.39) for some A, B). We now state without proof a fundamental feature of Gromov hyperbolic spaces.

Lemma 6.28 (Stability of quasi-geodesics). *Let γ be an (A, B)-quasi-geodesic ray in a Gromov δ-hyperbolic space X. Then there exists $M < \infty$ dependending only on δ, A and B so that γ is contained in the M-neighborhood of a true geodesic ray γ' in X. In particular, γ determines a unique point in $\partial_\infty X$, namely, the point $[\gamma']$.*

See, for example, Theorem III.H.1.7 in [43]. Thus $\partial_\infty X$ can alternately be described as the space of equivalence classes $[\gamma]$ of quasi-geodesic rays $\gamma : [0, \infty) \to X$, with the same equivalence relation as before.

Via Lemma 6.28, we see that F induces a boundary map $F_\infty : \partial_\infty H^2_{\mathbb{C}} \to \partial_\infty H^2_{\mathbb{C}}$, as follows:

$$F_\infty([\xi]) = [F(\xi)].$$

We equip $\partial_\infty H^2_{\mathbb{C}}$ with the visual metric ρ_1 as in Subsection 3.4.5.

Lemma 6.29. *F_∞ is a homeomorphism of $(\partial_\infty H^2_{\mathbb{C}}, \rho_1)$.*

In fact, more is true. A homeomorphism $f : X \to Y$ of metric spaces is called *quasiconformal* (according to the *metric definition*) if there exists $H < \infty$ so that

$$\limsup_{r \to 0} \frac{\sup_{x' : d(x, x') \leq r} d(f(x), f(x'))}{\inf_{x'' : d(x, x'') \geq r} d(f(x), f(x''))} \leq H \tag{6.40}$$

for all $x \in X$. Here we denoted the metric in both X and Y by d.

Lemma 6.30. *F_∞ is a quasiconformal homeomorphism of $(\partial_\infty H^2_{\mathbb{C}}, \rho_1)$.*

This is a general result about quasi-isometries of δ-hyperbolic spaces which admits several proofs. Margulis [194] gave an elegant proof using the notion of *geodesic shadows*; see also Bourdon [40]. The key point is a comparison of geodesic shadows in the target with images of geodesic shadows from the source; this comparison is mediated via quasi-geodesic shadows in the target through the Stability Lemma. Observe that the equivariance of F with respect to the actions of Γ, Γ' on $H^2_{\mathbb{C}}$ implies a similar property for F_∞ with respect to the actions of Γ, Γ' on the Gromov boundary.

Part II. In the second half of the proof, regularity and rigidity phenomena for quasiconformal maps, together with an ergodic theoretic argument, are used to deduce the following result:

Proposition 6.31. *Every (Γ, Γ')-equivariant quasiconformal map $F_\infty : \partial_\infty H^2_{\mathbb{C}} \to \partial_\infty H^2_{\mathbb{C}}$ is the boundary map associated with the action of an isometry on $H^2_{\mathbb{C}}$.*

The first step is to verify higher regularity for quasiconformal maps, and confirm the equivalence of the metric definition (6.40) with an analytic formulation.

Proposition 6.32. *A homeomorphism $G : \partial_\infty H^2_{\mathbb{C}} \to \partial_\infty H^2_{\mathbb{C}}$ is quasiconformal (in the sense of (6.40)) if and only if it is in the local horizontal Sobolev class $S^{1,4}_{\mathrm{loc}}$, is almost everywhere Pansu differentiable, and has Pansu differential D_0G satisfying the distortion inequality*

$$||(D_0G)(\xi)_*||^4 \le K \det(D_0G)(\xi)_* \qquad \textit{for a.e. } \xi \tag{6.41}$$

for some absolute constant $K < \infty$. The constants K and H depend quantitatively on each other.

Here we have extended the notions of the horizontal Sobolev class $S^{1,p}$ and of Pansu differentiability from the setting of the Heisenberg group to the Gromov boundary of $H^2_{\mathbb{C}}$; this is easy to accomplish using the generalized stereographic projection. For example, membership of G in the local horizontal Sobolev class $S^{1,4}_{\mathrm{loc}}$ refers to membership of the coordinate functions of the conjugated maps $\pi \circ G \circ \pi^{-1} : \mathbb{H} \to \mathbb{H}$ in $S^{1,4}_{\mathrm{loc}}$ for all generalized stereographic projections $\pi : \partial_\infty H^2_{\mathbb{C}} \to \mathbb{H}$.

The condition in Proposition 6.32 is known as the *analytic definition* for quasiconformality. Observe that, after projecting to $\mathbb{H}$ and writing in terms of the standard basis X_1, X_2, X_3, the Pansu differential of G takes the form

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & ad - bc \end{pmatrix};$$

see Example 6.5. The norm of $(D_0G)(\xi)_*$ in (6.41) refers to the maximum induced stretch in horizontal directions:

$$||(D_0G)(\xi)_*|| = \sup_{V \in H_\xi \partial_\infty H^2_{\mathbb{C}} : |V|=1} |(D_0G)(\xi)_*(V)|.$$

In our setting, F_∞ becomes quasiconformal according to the analytic definition, and

$$\frac{||(D_0F_\infty)_*||^4}{\det(D_0F_\infty)_*}$$

is essentially bounded on $\partial_\infty H^2_{\mathbb{C}}$. The equivariance of F_∞ combined with an ergodicity theorem of Mautner yields 1-*quasiconformality* of F_∞. In a similar vein as in the Euclidean case, Liouville's rigidity theorem holds: every 1-quasiconformal map of $\partial_\infty H^2_{\mathbb{C}}$ is in fact conformal, and induced as the boundary map of an isometry of $H^2_{\mathbb{C}}$. (See also Theorem 6.33.) This finishes the proof of Proposition 6.31, and also finishes the proof of Mostow's Theorem 6.27, once we observe that the resulting isometry of $H^2_{\mathbb{C}}$ is equivariant and therefore descends to an isometry between the original manifolds M_Γ and M'_Γ.

6.5.1 Quasiconformal mappings on $\mathbb{H}$

Quasiconformal maps on $\mathbb{H}$ play a crucial role in the above argument. Classically, quasiconformal maps may be defined in a variety of ways: metrically, geometrically or analytically. Let us review some of the basic definitions. Let $f : U \to U'$ be a homeomorphism between domains in $\mathbb{H}$.

(i) We say that f is *metric quasiconformal* if (6.40) holds for some constant $H < \infty$. This is an infinitesimal condition of uniformly bounded relative distortion of metric spheres.

(ii) We say that f is (*locally*) *quasisymmetric* if there exists an increasing homeomorphism $\eta : [0, \infty) \to [0, \infty)$ so that for each Whitney ball $B \subset U$,[2]

$$\frac{d(f(x), f(y))}{d(f(x), f(z))} \le \eta\left(\frac{d(x, y)}{d(x, z)}\right) \tag{6.42}$$

for all $x, y, z \in B$, $x \neq z$. This is a local (but not infinitesimal) condition of uniformly bounded relative distortion of metric triples, the validity of (6.42) for some η clearly implies the validity of (6.40) with $H = \eta(1)$.

(iii) We say that f is *analytic quasiconformal* if its coordinate functions belong to the local horizontal Sobolev class $S^{1,4}_{\mathrm{loc}}$, f is Pansu differentiable (in particular, a contact map), and the Pansu differential $(D_0 f)_* : \mathfrak{h} \to \mathfrak{h}$ verifies the pointwise distortion inequality

$$||(D_0 f)(x)_*||^4 \le K \det(D_0 G)(x)_*$$

for a.e. $x \in U$.

Theorem 6.33 (Korányi–Reimann). *Conditions* (i), (ii) *and* (iii) *are quantitatively equivalent, for homeomorphisms of domains in $\mathbb{H}$. Moreover, condition* (i) *with $H = 1$ implies condition* (iii) *with $K = 1$, and any such* conformal *map is a composition of maps of the following types:*

- *left translations,*
- *dilations,*
- *rotations about the x_3-axis,*
- *the* Korányi inversion

$$j_{\mathbb{H}}(z, x_3) = \left(\frac{-z}{|z|^2 + 4\,\mathbf{i}\,x_3}, \frac{-x_3}{|z|^4 + 16x_3^2}\right)$$

(*see* (2.14)).

All such conformal maps correspond, under generalized stereographic projection, with the boundary maps associated with the action of the isometry group $SU(1, 2)$ on $H^2_{\mathbb{C}}$.

---

[2] B is a *Whitney ball* in U if $2B \subset U$.

Korányi and Reimann also conclusively demonstrated the existence of an extensive supply of nontrivial (e.g., nonconformal) quasiconformal mappings of $\mathbb{H}$ by characterizing the infinitesimal generators of one-parameter flows of smooth quasiconformal maps.

Theorem 6.34 (Korányi–Reimann). *Let $p \in C^\infty(\mathbb{H})$ and assume that $X_1X_1p - X_2X_2p, X_1X_2p + X_2X_1p \in L^\infty(\mathbb{H})$. Then the vector field*

$$V = pX_3 + (X_2p)X_1 - (X_1p)X_2$$

generates a one-parameter family of smooth quasiconformal maps $f_s : \mathbb{H} \to \mathbb{H}$, $s > 0$, as solutions to the Cauchy problem

$$\frac{d}{ds} f_s = V(f_s), \qquad f_0 = \mathrm{id}\,.$$

6.6 Notes

The canonical treatment of geometric measure theory remains the comprehensive tome of Federer [95]; much contemporary work in sub-Riemannian and general metric space-valued geometric measure theory involves the detailed development and elaboration of ideas and programs stemming from [95].

Notes for Section 6.1. In this section we closely follow the development of the area formula by Magnani [190] (but see also [189], and [222] for an equivalent but different treatment). Magnani's proof is a simplified version of the proof of an area formula in a much more general class of metric spaces (see, for example, [161] or [9]). One can also give a proof of the area formula in Carnot groups analogous to the classical proof. For the details see [190, pp. 99–100]. Extensive work on the sub-Riemannian co-area formula has been done by Magnani [190–192]. The sub-Riemannian analog for the differential introduced in Definition 6.3 bears the name of Pierre Pansu, who introduced it and proved Theorem 6.4 in his important work [220]. The Pansu differential and Pansu–Rademacher differentiation theorem are foundational tools in the modern theory of geometric analysis in the Carnot–Carathéodory environment.

Notes for Section 6.2. The proof of Theorem 6.13 which we give is due to Calderón; see also Heinonen [136, Chapter 6]. Pansu's original paper on the a.e. differentiability of Lipschitz functions on Carnot groups is [220]. For a recent far-reaching generalization to metric spaces, see Cheeger [60]. Arcozzi and Morbidelli [15], [16] have given an analytic characterization for bi-Lipschitz self-maps of the Heisenberg group using the Pansu derivative, as well as a related stability theorem for Heisenberg isometries in the spirit of F. John [157]

The eikonal equation for the CC metric was proved by Monti in $\mathbb{H}$ [204], and by Monti and Serra-Cassano in a wide class of CC spaces [211]. Example 6.15 is also taken from [211].

Notes for Section 6.3. Proposition is taken from Monti and Serra-Cassano [211], who work in a much more general setting of Carnot–Carathéodory spaces. For an even more general perspective, see [7]. Proposition 6.18 is due to Monti [204].

Notes for Section 6.4. Our derivation of the noncharacteristic first variation of the perimeter is taken from Bonk–Capogna [37]. Alternative derivations of first variation formulas can be found in [63], [78], [221], [37], [199], [232] and [239]. The general form of the first variation is a result of Ritoré and Rosales [231]. An essential novelty of our discussion, however, is the formulation of an explicit first variation formula for parameteric surfaces. Such formulas can be used, for instance, to compute the mean curvature of surfaces represented as graphs over spheres or other closed manifolds.

Notes for Section 6.5. Mostow's rigidity theorem inaugurated the study of quasiconformal maps in Carnot groups, general Carnot–Carathéodory spaces, and most recently in metric measure spaces. While we have stated Mostow's theorem only in the case of $H^2_{\mathbb{C}}$, where the boundary quasiconformal analysis resides on the (one-point compactification of) the Heisenberg group, the original result [212] was formulated for general rank 1 symmetric spaces. Recall that a complete list of noncompact, negatively curved, rank 1 symmetric spaces consists of the real, complex and quaternionic hyperbolic spaces of dimension at least 2:

$$H^n_{\mathbb{R}}, \quad H^n_{\mathbb{C}}, \quad H^n_{\mathbb{K}}, \qquad n \geq 2$$

($\mathbb{K}$ denotes the division algebra of quaternions), and the Cayley hyperbolic plane

$$H^2_{\mathbb{O}}.$$

Mostow's theorem 6.27 holds in all of these cases except for $H^2_{\mathbb{R}}$; the essential obstruction in the proof involves the absolute continuity in measure of the boundary quasiconformal (more properly, quasisymmetric) maps, which fails in the case of maps of $\mathbb{S}^1 = \partial_\infty H^2_{\mathbb{R}}$. Pansu [220] obtained a stronger rigidity statement in the quaternionic and Cayley situations. See also [135], especially Section 6.

Korányi and Reimann developed the full theory of quasiconformal maps on the Heisenberg groups $\mathbb{H}^n$ in [168], [171]. The case $K = 1$ in Theorem 6.33 follows from their work through a regularity theorem for nonlinear subelliptic PDE proved in [246] and [50]. Alternative methods to construct quasiconformal maps in the Heisenberg groups can be found in [57]. The existence of a rich theory of quasiconformal maps in this specific non-Riemannian setting motivated further study of quasiconformal function theory in the setting of general metric measure spaces. The seminal work in this arena is due to Heinonen and Koskela [140], [141], who further studied the equivalence of definitions of quasiconformality in Theorem 6.33 (in more general Carnot groups and abstract metric measure spaces) and extended much of the ensuing Euclidean theory to this setting. Further work on the equivalence of definitions of quasiconformality, including the *geometric definition* of quasiconformality (which we have not touched on here) was done by

Tyson [251], [252]. For the most recent summary of these developments, we refer to Heinonen et al. [143]. Note that the concept of quasisymmetry, for mappings of metric spaces, was already introduced by Tukia and Väisälä [250] in 1980.

Quasiregular maps are a generalization of quasiconformal maps where the assumption of injectivity is relaxed. Heinonen and Holopainen [138] developed nonlinear potential theory and quasiregular maps on Carnot groups.

Chapter 7

The Isoperimetric Inequality in $\mathbb{H}$

The isoperimetric inequality in $\mathbb{H}$ with respect to the horizontal perimeter was first proved by Pansu. We first state it in the setting of C^1 sets.

Theorem 7.1 (Pansu's isoperimetric theorem in $\mathbb{H}$). *There exists a constant $C > 0$ so that*

$$|E|^{3/4} \le C P_{\mathbb{H}}(E) \tag{7.1}$$

for any bounded open set $E \subset \mathbb{H}$ with C^1 boundary.

In this chapter, we present two very different proofs for this theorem. The first proof is based on the geometric Sobolev embedding $S^{1,1} \subset L^{4/3}$ from Chapter 6. The second proof follows Pansu's original approach and rests on Santaló's formula from integral geometry as used by Croke; it gives an explicit (nonsharp) value for C.

7.1 Equivalence of the isoperimetric and geometric Sobolev inequalities

We give a quick sketch of the argument which shows that the isoperimetric inequality (7.1) is equivalent with the geometric Sobolev inequality (5.20). Suppose that $E \subset \mathbb{H}$ is a bounded, open, C^1 set, and let $R > 0$ such that $E \subset B(o, R)$. Choose $\delta > 0$ such that $2\delta < \operatorname{dist}(\bar{E}, \partial B(o, R))$, where $\operatorname{dist}(\cdot, \bar{E})$ represents the Euclidean distance from $\bar{E}$. Define the (Euclidean) Lipschitz function

$$f_\delta(x) = \left(1 - \frac{\operatorname{dist}(x, \bar{E})}{\delta}\right)^+.$$

Applying Proposition 5.17 to f_δ we obtain

$$\begin{aligned} |E|^{\frac{3}{4}} &\leq \left|\{x \in B(o,R) : f_\delta(x) > t\}\right|^{\frac{3}{4}} \\ &\leq \frac{C}{t} \int_{B(o,R)} |\nabla_0 f_\delta(y)|\, dy \end{aligned} \tag{7.2}$$

for each $t < 1$. Let A_δ be the intersection of $B(o,R)$ with a tubular neighborhood of E of radius δ. From (7.2) and the co-area formula (see for instance [95, Theorem 3.2.3]), we obtain

$$|E|^{3/4} \leq \frac{C}{t\delta} \int_{A_\delta} |\nabla_0 \operatorname{dist}(y, \bar{E})|\, dy;$$

letting $t \to 1$ yields

$$|E|^{3/4} \leq \frac{C}{\delta} \int_0^\delta \int_{\{y \in B(o,R): \operatorname{dist}(y,\bar{E})=s\}} \frac{|\nabla_0 \operatorname{dist}(\cdot, \bar{E})|}{|\nabla \operatorname{dist}(\cdot, \bar{E})|}\, d\mathcal{H}^{n-1}\, ds,$$

where $d\mathcal{H}^{n-1}$ denotes the $n-1$-dimensional Hausdorff measure with respect to the background Euclidean metric. The proof of (7.1) is concluded once we let $\delta \to 0$ and recall Corollary 5.8.

7.2 Isoperimetric inequalities in Hadamard manifolds

Pansu's proof of Theorem 7.1 is based on ideas of Croke [74]. In this section, we sketch the main lines of Croke's proof of the inequality

$$\operatorname{Vol}(\partial\Omega) \geq C \operatorname{Vol}(\Omega)^{(n-1)/n}, \tag{7.3}$$

for any domain Ω in a simply connected n-manifold M with non-positive sectional curvature. Here C denotes a positive constant depending only on n.

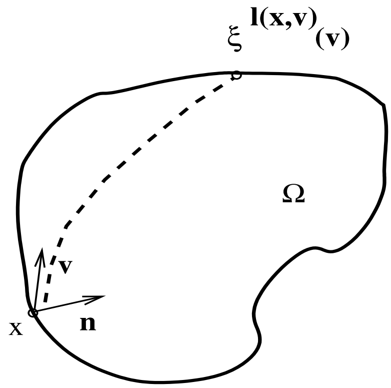
We begin by fixing some notation. We denote by $UM|_{\partial\Omega}$ the unit tangent bundle of M restricted to $\partial\Omega$. We denote vectors in UM as $w = (x,v)$ where $x \in M$ and $v \in T_xM$. For $x \in \partial\Omega$, we denote by $\vec{n}$ the inward pointing unit normal to $\partial\Omega$ at x. Given a point $x \in \Omega$ and a vector $v \in U_x\Omega$, let $r \mapsto \gamma_{(x,v)}(r)$ be the unit speed geodesic starting at x with initial tangent vector v. Finally, let $\xi_x^r(v) = (\gamma_{(x,v)}(r), \gamma'_{(x,v)}(r))$ be the geodesic flow.

Now, for $(x,v) \in U\Omega$, we set

$$l(x,v) = \sup\{t \,|\, \gamma_{(x,v)}(t) \in \Omega\}.$$

In other words, $l(x,v)$ is the first time when $\gamma_{(x,v)}(t)$ exits Ω (see Figure 7.1 for an illustration). Next, let

$$\tilde{l}(x,v) = \sup\{t \leq l(x,v) \,|\, \gamma_{(x,v)} \text{ minimizes up to } t\}.$$



An illustration of $l(x,v)$

Figure 7.1: Illustration of Pansu's approach.

If we denote by $\mathrm{Cut}(x,v)$ the distance from x to the cut locus of x along $\gamma_{(x,v)}$, then $\tilde{l}(x,v) \leq \mathrm{Cut}(x,v)$. We next define well-behaved subsets of the unit tangent bundle

$$\tilde{U}M = \{(x,v) \in UM | \tilde{l}(x,v) = l(x,v)\}$$

and

$$U^{+}\partial\Omega = \{(x,v) \in UM|_{\partial\Omega} | \langle\!\langle v, \vec{n} \rangle\!\rangle_x \geq 0\}$$

where $\vec{n}$ denotes the unit outer normal to $\partial\Omega$ in M and $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ is the Riemannian metric on M.

The proof of (7.3) is based on Santaló's formula: if $f : UM \to \mathbb{R}$ is an integrable function, then

$$\int_{\tilde{U}M|_{\Omega}} f(x,v)\, dx\, dv = \int_{U^{+}\partial\Omega} \int_{0}^{\tilde{l}(x,v)} f(\xi_x^r(v)) dr \langle\!\langle v, \vec{n} \rangle\!\rangle(x)\, dx\, dv. \tag{7.4}$$

In particular, for $f = 1$, we have

$$\omega_{n-1} \operatorname{Vol}(\Omega) \geq \operatorname{Vol}(\tilde{U}M|_{\Omega}) = \int_{U^{+}\partial\Omega} \tilde{l}(x,v) \langle\!\langle v, \vec{n} \rangle\!\rangle\, dv\, dx \tag{7.5}$$

where $(x,v) \in UM|_{\partial\Omega}$ and ω_{n-1} denotes the surface area of $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. If we define the density $F(w,r)$ for $w \in UM$ and $0 \leq r < \mathrm{Cut}(w)$ by the formula

$$\operatorname{Vol}(M) = \int_{UM(x)} \int_{0}^{\mathrm{Cut}(w)} F(w,r)\, dr\, dw, \tag{7.6}$$

then

$$\int_{0}^{l} \int_{0}^{l-\tau} F(\xi_x^{\tau}(v), z)\, dz\, d\tau \geq C(n) \frac{l^{n+1}}{\pi^{n+1}} \tag{7.7}$$

for any $x \in \partial\Omega$, $v \in U_xM$ and $l \leq Cut((x,v))$, where $C(n) = \frac{\pi\alpha(n)}{2\alpha(n-1)}$ and $\alpha(n)$ is the volume of the unit n-sphere. The validity of (7.3) is a consequence of (7.4) and (7.7), and is sketched in the following chain of inequalities. For any $x \in \partial\Omega$,

$$\begin{aligned}
\mathrm{Vol}(\Omega)^2 &= \int_\Omega \int_{UM(x)} \int_0^{\mathrm{Cut}((x,v))} F((x,v),r)\, dr\, dv\, dx \\
&\geq \int_\Omega \int_{\tilde{U}_x M} \int_0^{\mathrm{Cut}((x,v))} F((x,v),r)\, dr\, dv\, dx \\
&\geq \int_{\tilde{U}M|_\Omega} \int_0^{\tilde{l}(x,v)} F((x,v),t)\, dt\, dx\, dv,
\end{aligned}$$

by the definition of $\tilde{l}$

$$= \int_{U^+\partial\Omega} \int_0^{\tilde{l}(x,v)} \int_0^{\tilde{l}(\xi_x^s(v))} F(\xi_x^s(v),t) \langle\langle v, \vec{n}\rangle\rangle\, dt\, ds\, dx\, dv,$$

from Santaló's formula (7.4)

$$\geq \int_{U^+\partial\Omega} \left[\int_0^{\tilde{l}(x,v)} \int_0^{\tilde{l}(x,v)-s} F(\xi_x^s(v),t)\, dt\, ds \right] \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv,$$

because $\tilde{l}(\xi_x^s(v)) \geq \tilde{l}(x,v) - s$

$$\geq C \int_{U^+\partial\Omega} \tilde{l}(x,v)^{n+1} \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv,$$

from (7.7)

$$\geq C \frac{\left(\int_{U^+\partial\Omega} \tilde{l}(x,v) \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv \right)^{n+1}}{\left(\int_{U^+\partial\Omega} \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv \right)^{n}},$$

by Hölder's inequality. From this calculation and (7.5) we deduce

$$\begin{aligned}
\mathrm{Vol}(\Omega)^2 \left(\int_{U^+\partial\Omega} \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv \right)^n &\geq C \left(\int_{\tilde{U}M|_{\partial\Omega}} \tilde{l}(x,v) \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv \right)^{n+1} \\
&= C\, \mathrm{Vol}(\tilde{U}M|_\Omega)^{n+1} \geq C\, \mathrm{Vol}(\Omega)^{n+1}.
\end{aligned} \tag{7.8}$$

In addition,

$$\int_{U^+\partial\Omega} \langle\langle v, \vec{n}\rangle\rangle\, dx\, dv \leq \int_{U^+\partial\Omega} dx\, dv = \frac{1}{2}\omega_{n-1}\ \mathrm{Vol}(\partial\Omega). \tag{7.9}$$

The isoperimetric inequality (7.3) is a direct consequence of (7.8) and (7.9).

7.3 Pansu's proof of the isoperimetric inequality in $\mathbb{H}$

In this section, we illustrate how Pansu adapts Croke's argument to the Heisenberg group. Let $\Omega \subset \mathbb{H}$ be a smooth open set and denote by ν_H the *horizontal normal* to $\partial\Omega$ as defined in (4.9). For all $x \in \partial\Omega$ and $\theta \in [0, 2\pi)$ denote by $e^{\mathbf{i}\,\theta}$ the horizontal vector field $\cos\theta X_1(x) + \sin\theta X_2(x)$, and define the *first exit time*

$$r(x, \theta) = \sup\{r > 0 | \exp(te^{\mathbf{i}\,\theta}) \in \Omega\}.$$

The first step in Pansu's argument consists in proving the following Santaló-type formula:

$$2\pi|\Omega| = \int_{\partial\Omega}\int_0^{2\pi} r(x,\theta)\langle e^{\mathbf{i}\,\theta}, \nu_H\rangle_1 d\theta d\mu(x), \tag{7.10}$$

where μ is the perimeter measure defined in Corollary 5.8.

For any point $x \in \mathbb{H}$ we define its *canonical section* Σ_x to be the set of horizontal lines through x. In other words, if $s_x : \mathbb{R}^2 \to \mathbb{H}$ denotes the section

$$s_x(a, b) = x \exp(aX_1 + bY_1),$$

then $\Sigma_x = s_x(\mathbb{R}^2)$. In our presentation of the Heisenberg group, Σ_x is a plane with (Euclidean) normal vector in the direction $(-a/2, b/2, -1)$ where $x = (a, b, c)$. Choose polar coordinates (r, θ) in $\mathbb{R}^2$ centered at the point z, the projection of x onto $\mathbb{R}^2$. Then we can pull back the perimeter measure $d\mu$ through s_x to a measure in $\mathbb{R}^2$ given by

$$s_x^*(d\mu) = \frac{1}{2}r^2\, dr \wedge d\theta. \tag{7.11}$$

Proposition 7.2. *Let $D \subset \mathbb{R}^2$ be a smooth, bounded open set and let $V \subset \mathbb{H}$ be a smooth surface with boundary $s_o(\partial D)$. Then*

$$\mu(V) \geq \mu(s_o(D)).$$

Proof. The set $s_o(D)$ is foliated by horizontal lines, hence has mean curvature zero at all points. Consequently it is a sub-Riemannian minimal surface (see Section 4.3 for more details) and hence is a critical point for the perimeter variation. A calibration argument shows that $s_o(D)$ is a true minimizer. $\square$

Define the open subset of Σ_x which is *visible* from the point x:

$$P_x = \{x \exp(\rho e^{\mathbf{i}\,\theta}) \,|\, x\exp(te^{\mathbf{i}\,\theta}) \in \Omega \text{ for all } 0 < t < \rho\ \}.$$

We remark that Σ_x divides $\partial\Omega$ in two connected components whose boundaries coincide with ∂P_x.

We call these components V_1 and V_2, so that

$$\mu(\partial\Omega) = \mu(V_1) + \mu(V_2). \tag{7.12}$$

Lemma 7.3. *In the notation established above,*

$$\int_{\partial\Omega}\int_0^{2\pi} r(x,\theta)\,d\theta\,d\mu(x) = 2\int_{\partial\Omega}\int_{P_x} \rho^{-2}\,d\mu\,d\mu(x).$$

Proof. Denote by (ρ,ϕ) the coordinates in $\Sigma_x \subset \mathbb{H}$ induced by the polar coordinates $(r,\theta) \in \mathbb{R}^2$. From the definition of P_x and (7.11) we have

$$\begin{aligned}\int_0^{2\pi} r(x,\theta)\,d\theta &= \int_{\partial P_x} \rho d\phi = \int_{P_x} d(\rho d\phi)\\ &= \int_{P_x} d\rho\wedge d\phi = 2\int_{P_x}\frac{1}{\rho^2}\,d\mu.\end{aligned}$$

An integration over $\partial\Omega$ with respect to μ completes the proof. $\square$

Lemma 7.4. *In the notation established above,*

$$\int_{P_x} \rho^{-2}\,d\mu \leq (3\pi^2)^{1/3}\mu(P_x)^{1/3}. \tag{7.13}$$

Proof. To simplify the derivation, we pull back to the plane via the horizontal section map $s_x : \mathbb{R}^2 \to \Sigma_x$. Then (7.13) reads

$$\frac{1}{2}\int_D \frac{1}{|x|}\,dx \leq (3\pi^2)^{1/3}\left(\frac{1}{2}\int_D |x|\,dx\right)^{1/3},$$

where $D = s_x^{-1}(P_x)$, i.e.,

$$\int_D \frac{1}{|x|}\,dx \leq (12\pi^2)^{1/3}\left(\int_D |x|\,dx\right)^{1/3}. \tag{7.14}$$

Let $B(o,R)$ be a disc with area $|D|$. To obtain (7.14), we estimate

$$\begin{aligned}\int_D \frac{1}{|x|}\,dx &\leq \int_{B(o,R)} \frac{1}{|x|}\,dx = 2\pi R\\ &= (12\pi^2)^{1/3}\left(\int_{B(o,R)} |x|\,dx\right)^{1/3}\\ &\leq (12\pi^2)^{1/3}\left(\int_D |x|\,dx\right)^{1/3}\end{aligned}$$

by two applications of Lemma 7.5. This completes the proof of Lemma 7.4. $\square$

Lemma 7.5. *If $B(o,R)$ is a ball in $\mathbb{R}^n$ with volume $|D|$, and $h : [0,\infty) \to [0,\infty)$ is increasing, then*

$$\int_D h(|x|)\,dx \geq \int_{B(o,R)} h(|x|)\,dx.$$

If h is decreasing then the inequality is reversed.

Proof. Since $|B(o,R) \setminus D| = |D \setminus B(o,R)|$, we have

$$\begin{aligned}\int_{D\setminus B(o,R)} h(|x|)\,dx &\geq h(R)|D \setminus B(o,R)| \\ &= h(R)|B(o,R) \setminus D| \\ &\geq \int_{B(o,R)\setminus D} h(|x|)\,dx.\end{aligned}$$

Now add $\int_{D\cap B(o,R)} h(|x|)\,dx$ to both sides. □

Lemma 7.6. *In the notation established above,*

$$\int_{\partial\Omega} \mu(P_x)^{1/3} d\mu \leq 2^{-1/3}\mu(\partial\Omega)^{4/3}.$$

Proof. From Proposition 7.2 and from (7.12) we see that

$$2\mu(P_x) \leq \mu(V_1) + \mu(V_2) = \mu(\partial\Omega).$$

An integration over $\partial\Omega$ with respect to μ completes the proof. □

A proof of the isoperimetric inequality in $\mathbb{H}$ can now be obtained from (7.10) and Lemmata 7.3–7.6 via the following string of estimates:

$$\begin{aligned}2\pi|\Omega| &= \int_{\partial\Omega}\int_0^{2\pi} r(x,\theta)\langle e^{\mathbf{i}\,\theta}, \nu_H\rangle_1\,d\theta\,d\mu(x) \\ &\leq \int_{\partial\Omega}\int_0^{2\pi} r(x,\theta)\,d\theta\,d\mu(x) \\ &= 2\int_{\partial\Omega}\int_{P_x} \rho^{-2} d\mu d\mu \\ &\leq 2(3\pi^2)^{1/3}\int_{\partial\Omega} \mu(P_x)^{1/3}\,d\mu \\ &\leq (12\pi^2)^{1/3}\mu(\partial\Omega)^{4/3}\end{aligned}$$

or

$$\frac{|\Omega|^{3/4}}{P_{\mathbb{H}}(\Omega)} \leq \left(\frac{3}{2\pi}\right)^{1/4}. \tag{7.15}$$

7.4 Notes

The approach to the isoperimetric inequality via the geometric Sobolev inequality goes back to the work of Fleming and Rishel [97], see also Maz'ya [197]. The sub-Riemannian case was proved independently in [53], [104] for Hörmander vector fields and more general structures, and [35] in the setting of Dirichlet forms. Pansu's proof of the isoperimetric inequality appeared in his 1982 paper [217], see also [219]. For the argument of Croke, see [74]; Santaló's formula (7.4) can be found in his textbook on integral geometry [237]. An extension of Santaló's formula (7.10) to all Carnot groups can be found in [200]. A different sharp isoperimetric inequality in Hadamard manifolds (comparing with model space forms) has been given by Kleiner [163].

Chapter 8

The Isoperimetric Profile of $\mathbb{H}$

This chapter is the core of this survey. We recall the definition of isoperimetric profile of $\mathbb{H}$ and Pansu's 1982 conjecture. Next we present a proof of the existence of an isoperimetric profile and describe some of the existing literature on the isoperimetric problem. Our aim is to reveal the main ideas and outlines of the proofs of various partial results and sketch some further techniques and methods which may lead to a solution, in order to guide the reader through the literature and to give a sense of the larger ideas that are in play.

8.1 Pansu's conjecture

The isoperimetric constant of the Heisenberg group is the best constant $C_{\mathrm{iso}}(\mathbb{H})$ for which the isoperimetric inequality

$$\min\{|\Omega|^{3/4}, |\mathbb{H}\setminus\Omega|^{3/4}\} \le C_{\mathrm{iso}}(\mathbb{H})P_{\mathbb{H}}(\Omega) \tag{8.1}$$

holds. In other words,

$$C_{\mathrm{iso}}(\mathbb{H}) = \sup_{\Omega} \frac{\min\{|\Omega|^{3/4}, |\mathbb{H}\setminus\Omega|^{3/4}\}}{P_{\mathbb{H}}(\Omega)}, \tag{8.2}$$

where the supremum is taken on all Caccioppoli subsets of the Heisenberg group. In a dual manner we could define the isoperimetric constant of $\mathbb{H}$ as the value

$$\Big(\inf\{P_{\mathbb{H}}(E) : E \subset \mathbb{H} \text{ is a bounded Caccioppoli set and } |E| = 1\}\Big)^{-1}. \tag{8.3}$$

We also define the isoperimetric profile.

Definition 8.1. An *isoperimetric profile* for $\mathbb{H}$ consists of a family of bounded Caccioppoli sets $\Omega_{\mathrm{profile}} = \Omega_{\mathrm{profile}}(V)$, $V > 0$, with $|\Omega_{\mathrm{profile}}(V)| = V$ and

$$|\Omega_{\mathrm{profile}}|^{3/4} = C_{\mathrm{iso}}(\mathbb{H})P_{\mathbb{H}}(\Omega_{\mathrm{profile}}).$$

The invariance and scaling properties of the Haar measure and perimeter measure clearly imply that the sets comprising the isoperimetric profile are closed under the operations of left translation and group dilation. In [219], Pansu conjectured that any set in the isoperimetric profile of $\mathbb{H}$ is, up to translation and dilation, a bubble set $\mathcal{B}(o,R)$. Recall from Section 2.3 that $\mathcal{B}(o,R)$ is obtained by rotating around the x_3-axis a geodesic joining two points at height $\pm\pi R^2/2$. More precisely, these cylindrically symmetric surfaces have profile curve

$$x_3 = f_R(r) = \pm\frac{1}{4}(r\sqrt{R^2-r^2} + R^2 \arccos r/R).$$

Setting $u(x) = f_R(|z|) - x_3$, $x = (z, x_3)$, we easily compute

$$|\mathcal{B}(o,R)| = 4\pi \int_0^R r f(r)\, dr = \frac{3}{16}\pi^2 R^4 \tag{8.4}$$

and

$$P_{\mathbb{H}}(\mathcal{B}(o,R)) = 2\int_{B(o,R)} |\nabla_0 u| = 4\pi \int_0^R r\sqrt{f'(r)^2 + r^2/4}\, dr = \frac{1}{2}\pi^2 R^3,$$

yielding the following conjecture for the value of the Heisenberg isoperimetric constant and the isoperimetric profile.

Conjecture 8.2 (Pansu).

$$C_{\mathrm{iso}}(\mathbb{H}) = \frac{|\mathcal{B}(o,R)|^{3/4}}{P_{\mathbb{H}}(\mathcal{B}(o,R))} = \frac{3^{3/4}}{4\sqrt{\pi}} \tag{8.5}$$

for any R, and equality is obtained if and only if Ω is a bubble set.[1]

We take this opportunity to reiterate the fact that Pansu's conjecture is still unsolved in this generality, although numerous partial results and special cases have been established over the years.

The isoperimetric problem for Minkowski content may be formulated as follows: determine the value of

$$\min\{\mathcal{M}_3(\partial E) : E \subset \mathbb{H} \text{ is bounded}, |E| = 1\}. \tag{8.6}$$

Proposition 6.17 shows that the minima in (8.6) and in (8.3) coincide when the class of competitors is restricted to sets with C^2 boundary. (In Section 8.5 we sketch an argument verifying Pansu's conjecture in this category.) As we shall see, it is not currently known if the isoperimetric profile sets in $\mathbb{H}$ have C^2 boundary. However, the expression

$$\inf\{\mathcal{M}_3(\partial E) : E \subset \mathbb{H} \text{ is bounded}, |E| = 1\} \tag{8.7}$$

is equal to the isoperimetric constant of $\mathbb{H}$.

[1]Compare the conjectured value for $C_{\mathrm{iso}}(\mathbb{H})$ with the (nonsharp) value obtained in (7.15).

Overview of the chapter. The first portion of this chapter is devoted to describing evidence supporting Pansu's conjecture culminating, in Sections 8.5 and 8.6, with affirmative answers in the C^2, respectively convex, category. It is important to note that none of the machinery necessary to analyze variational problems of this sort existed at the time when the conjecture was proposed. The development of such machinery beginning in the late 1990s was instrumental in laying the groundwork needed for potential approaches to, and partial results for, this and related conjectures.

In Section 8.2, we present an important result of Leonardi and Rigot asserting the existence of an isoperimetric profile in any Carnot group, and demonstrating weak regularity (i.e., Ahlfors-type regularity and interior and exterior corkscrew condition, see [176, Definition 2.10]) for the constituent sets. As we saw in Chapter 6, the study of sub-Riemannian geometric measure theory is still in its infancy and, in particular, there is not sufficient infrastructure to bootstrap further regularity properties of the solution. This motivates the introduction of a substantial, yet useful, restriction to the class of C^2 surfaces. Under this smoothness assumption, we show, in Section 8.3, that the isoperimetric sets have constant horizontal mean curvature and, as a consequence, have particularly nice parametrizations. This allows for an analysis of the isoperimetric in this class via two different methods. First, one may use the link to sub-Riemannian constant mean curvature surfaces to introduce the methods of geometric partial differential equations. Second, the techniques of Riemannian geometric analysis of constant mean curvature surfaces may be adapted.

In Section 8.4, we present a result of Danielli, Garofalo and Nhieu showing the validity of Pansu's conjecture under an extra symmetry and C^1 smoothness assumption. In Section 8.5, we sketch the proof of a recent ground-breaking result of Ritoré and Rosales (Theorem 8.23) which allows for the removal of the symmetry assumption, showing that the isoperimeteric profile in the class of closed C^2 surfaces is indeed given by the class of bubble sets. The results in Section 8.5 rest heavily on the work of Cheng, Hwang, Malchiodi and Yang presented in Sections 4.4.1 and 4.4.2. In Section 8.6 we present a very recent result of Monti and Rickly, where Pansu's conjecture is proved with no smoothness assumptions but in the class of (Euclidean) convex sets. This sequence of results provides strong evidence for the validity of Pansu's original conjecture.

In Section 8.7, we present three other possible approaches to the isoperimetric problem which either fail or are in some way incomplete. In Section 8.7.1, we present an approach based on Riemannian approximation. Using work of Tomter which classifies cylindrically symmetric constant mean curvature surfaces in the Riemannian Heisenberg group, one can analyze the evolution of these surfaces in $(\mathbb{R}^3, g_L)$ as $L \to \infty$. While one recovers the bubble sets as L tends to infinity, the method is incomplete as it requires finer information concerning the limiting process than is currently available. In Section 8.7.2, we present an analog to the well-known approach to isoperimetry through the Brunn–Minkowski theorem, and discuss the obstruction to implementing such a scheme in the sub-Riemannian case. Last, in Section 8.7.3, we discuss motion by horizontal mean curvature flow in $\mathbb{H}$ which could potentially be used to understand the isoperimetric profile. The concluding Section 8.8 discusses two related results: Monti and Morbidelli's computation of the isoperimetric profile of the Grushin plane, and Ritoré and Rosales' classification of C^2 rotationally symmetric constant mean curvature surfaces in $\mathbb{H}^n$.

8.2 Existence of minimizers

In this section we establish the existence of an isoperimetric profile. Although our exposition is in the Heisenberg group, the result continues to hold in the setting of general Carnot groups.

Theorem 8.3 (Leonardi–Rigot). *The Heisenberg group $\mathbb{H}$ admits an isoperimetric profile. More precisely, for any $V > 0$, there exists a bounded set $\Omega \subset \mathbb{H}$ with finite perimeter so that $|\Omega| = V$ and*

$$|\Omega|^{3/4} = C_{\text{iso}}(\mathbb{H}) P_{\mathbb{H}}(\Omega).$$

The proof of this theorem follows classical lines. One considers a sequence of sets $\Omega_i \subset \mathbb{H}$ with $|\Omega_i| = 1$ whose isoperimetric ratios

$$C_i = \frac{|\Omega_i|^{3/4}}{P_{\mathbb{H}}(\Omega_i)} = \frac{1}{P_{\mathbb{H}}(\Omega_i)}$$

converge to $C_{\text{iso}}(\mathbb{H})$ as $i \to \infty$ and examines the existence and properties of a subconvergent limit. It suffices to show that

(1) the sequence (Ω_i) subconverges to a set Ω_∞, and
(2) $|\Omega_\infty| = 1$.

Step (1) requires an extension of some additional techniques from Euclidean geometric measure theory to the Heisenberg setting, in particular, the following compactness theorem.

Theorem 8.4 (Garofalo–Nhieu). *Let (Ω_i) be a sequence of measurable sets so that*

$$\sup_i P_{\mathbb{H}}(\Omega_i) < \infty.$$

Then (Ω_i) subconverges in $L^1_{\text{loc}}(\mathbb{H})$ to a measurable set Ω_∞ with finite perimeter.

Proof. In view of Lemma 5.23, we can apply Theorem 5.24 to either $\Omega = B(x, R)$ or $\Omega = B_{\mathbb{H}}(x, R)$. If (Ω_i) is a sequence of Caccioppoli sets as in the statement of Theorem 8.4, then the functions $f_i = \chi_{\Omega_i \cap \Omega}$ lie in $BV(\Omega)$ and

$$||f_i||_{BV(\Omega)} \le |B(x, R)| + P_{\mathbb{H}}(\Omega_i, \Omega) \le M < \infty.$$

In view of Theorem 5.24 one can find $f_\infty \in BV(\Omega)$ such that $f_i \to f_\infty$ in $L^1(\Omega)$. Considering pointwise convergence, we see that $f_\infty = \chi_{\Omega_\infty}$ for some set $\Omega_\infty \subset \Omega$. Since $f_\infty \in BV$, Ω_∞ is of finite perimeter. The lower semi-continuity of the perimeter functional with respect to $L^1_{\rm loc}$ convergence guarantees that $P_{\mathbb{H}}(\Omega_\infty) \leq M$. This completes the proof of Theorem 8.4. □

Concentration-compactness. The compactness result of Theorem 8.4 guarantees the existence of the limit set Ω_∞. The bulk of the proof of Theorem 8.3 focuses on step (2). A *"concentration-compactness"*argument prevents the possibility that the sets Ω_i become very thin, spread out, and in the limit lose volume at infinity. Indeed, for each member of the sequence, a fixed amount of volume must lie within a ball of radius 1.

Lemma 8.5. *Let* $\omega_{\mathbb{H}} = |B(o,1)|$*, and let* A *be a set with* $0 < |A| < \infty$ *and* $0 < P_{\mathbb{H}}(A) < \infty$*. Assume that* $m \in (0, \omega_{\mathbb{H}}/2)$ *is such that* $|A \cap B(x,1)| < m$ *for all* $x \in \mathbb{H}$*. Then there exists a constant* $c > 0$ *so that*

$$c\left(\frac{|A|}{P_{\mathbb{H}}(A)}\right)^4 \leq m. \tag{8.8}$$

Proof. We sketch the proof. Let $\mathcal{S}$ be a maximal set of points in $\mathbb{H}$ with pairwise mutual distance at least $\frac{1}{2}$ and satisfying $|A \cap B(x,1/2)| > 0$ for all $x \in S$. By the maximality of $\mathcal{S}$, we have

$$\left| A \setminus \bigcup_{x\in\mathcal{S}} B(x,1) \right| = 0.$$

Then,

$$\begin{aligned}
|A| &\leq \sum_{x\in\mathcal{S}} |A \cap B(x,1)| \\
&= \sum_{x\in\mathcal{S}} |A \cap B(x,1)|^{\frac{1}{4}} |A \cap B(x,1)|^{\frac{3}{4}} \\
&\leq m^{\frac{1}{4}} \sum_{x\in\mathcal{S}} |A \cap B(x,1)|^{\frac{3}{4}} \\
&\leq C m^{\frac{1}{4}} \sum_{x\in\mathcal{S}} P_{\mathbb{H}}(A, B(x,1)).
\end{aligned}$$

The third line follows from the hypothesis concerning m and the last line follows from the relative isoperimetric inequality for balls (Lemma 5.26). Rearranging the final inequality yields the claim. □

Using this lemma, we sketch the proof of Theorem 8.3. Consider a minimizing sequence $\{\Omega_i\}$ satisfying $|\Omega_i| = 1$ and

$$P_{\mathbb{H}}(\Omega_i) \leq C_{\rm iso}(\mathbb{H})^{-1}\left(1 + \frac{1}{i}\right). \tag{8.9}$$

Lemma 8.5 guarantees that we can pick $x_i \in \Omega_i$ so that

$$|\Omega_i \cap B(x_i, 1)| \geq m_0$$

for some absolute constant $m_0 > 0$. Indeed, if no such constant existed, then (8.8) would violate the existence of $C_{\text{iso}}(\mathbb{H})$. Right translating each Ω_i by x_i, we may assume that Ω_i contains the origin and that

$$|\Omega_i \cap B(o, 1)| \geq m_0. \tag{8.10}$$

Theorem 8.4 assures the existence of Ω_∞ and lower semi-continuity of the perimeter yields

$$P_{\mathbb{H}}(\Omega_\infty) \leq \liminf_{i \to \infty} P_{\mathbb{H}}(\Omega_i) \leq C_{\text{iso}}(\mathbb{H})^{-1}.$$

The choice of m_0 implies

$$m_0 \leq \lim_{i \to \infty} |\Omega_i \cap B(o, 1)| = |\Omega_\infty \cap B(o, 1)|$$

and lower semi-continuity implies

$$|\Omega_\infty| \leq \liminf_{i \to \infty} |\Omega_i| \leq 1.$$

Taking all of this together yields

$$m_0 \leq |\Omega_\infty| \leq 1.$$

We claim that Ω_∞ is essentially bounded. Indeed, suppose that $|\Omega_\infty \cap B(o,r)| < 1$ for some $r \geq 2$; we will show that $r \leq R_0$ for some absolute constant $R_0 < \infty$. Then $|\Omega_\infty \cap B(o, R_0)| = 1$ whence Ω_∞ is essentially bounded.

To prove the claim, introduce $m_i(\rho) = |\Omega_i \cap B(o, \rho)|$ and $m_\infty(\rho) = |\Omega_\infty \cap B(o, \rho)|$. By assumption, $m_\infty(r) < 1$. Using the standard relation between the (local) perimeter and the rate of change of the (local) volume of a set, we find

$$m_i(\rho)^{3/4} \leq C_{\text{iso}}(\mathbb{H})\, P_{\mathbb{H}}(\Omega_i \cap B(o, \rho)) \leq C_{\text{iso}}(\mathbb{H})\, \left(P_{\mathbb{H}}(\Omega_i, B(o, \rho)) + m_i'(\rho)\right) \tag{8.11}$$

and

$$(1 - m_i(\rho))^{3/4} \leq C_{\text{iso}}(\mathbb{H})\, \left(P_{\mathbb{H}}(\Omega_i, \mathbb{H} \setminus \overline{B(o, \rho)}) + m_i'(\rho)\right) \tag{8.12}$$

for almost every $\rho > 0$. Using (8.9) we deduce

$$\begin{aligned} m_i(\rho)^{3/4} + (1 - m_i(\rho))^{3/4} &\leq C_{\text{iso}}(\mathbb{H})\, \left(P_{\mathbb{H}}(\Omega_i) + 2m_i'(\rho)\right) \\ &\leq 1 + \frac{1}{i} + 2C_{\text{iso}}(\mathbb{H}) m_i'(\rho) \end{aligned} \tag{8.13}$$

for almost every $\rho > 0$.

Let $\Phi(x) = x^{3/4} + (1-x)^{3/4} - 1$ and $\epsilon = \min\{m_0, (1-m_\infty(r))/2\}$. For sufficiently large i and all $x \in [\epsilon, 1-\epsilon]$, we have $m_i(r) \le (1+m_\infty(r))/2$ and $\Phi(x) \ge \frac{1}{i}$. Since

$$\epsilon \le m_i(1) \le m_i(\rho) \le m_i(r) \le 1-\epsilon$$

for all such i and all $1 \le \rho \le r$ (see (8.10)), we may rewrite the differential inequality (8.13) in the form

$$1 \le C \frac{m_i'(\rho)}{\Phi(m_i(\rho)) - \frac{1}{i}}, \qquad \text{a.e. } \rho \in [1, r].$$

Integrating from $r/2$ to r gives

$$\begin{aligned} \frac{r}{2} &\le C \int_{r/2}^{r} \frac{m_i'(\rho)\, d\rho}{\Phi(m_i(\rho)) - \frac{1}{i}} = C \int_{m_i(r/2)}^{m_i(r)} \frac{dx}{\Phi(x) - \frac{1}{i}} \\ &\le C \int_{\epsilon}^{1-\epsilon} \frac{dx}{\Phi(x) - \frac{1}{i}} \to C \int_0^1 \frac{dx}{\Phi(x)} < \infty \end{aligned}$$

as $i \to \infty$. The conclusion holds with $R_0 = 2C \int_0^1 \Phi(x)^{-1}\, dx$.

As observed above, this argument shows that the isoperimetry extremals Ω_∞ are essentially bounded. To pass from essential boundedness to boundedness requires additional regularity properties, namely, Ahlfors regularity of the boundary. See the notes to this section for further discussion of these and other properties of the isoperimetric subsets of $\mathbb{H}$.

8.3 Smooth isoperimetric profiles have constant horizontal mean curvature

In this section we prove that if the isoperimetric profile of $\mathbb{H}$ is C^2 smooth, then it necessarily has constant horizontal mean curvature away from the characteristic set. The proof is based on the first variation for the perimeter, with an additional volume constraint.

Proposition 8.6. *Let $\Omega \subset \mathbb{H}$ be a bounded open set enclosed by an oriented C^2 immersed surface S with g_1-Riemannian normal ν_1. Denote by $d\mu$ the perimeter measure defined in Corollary 5.8, and by $\mathcal{H}_0$ the horizontal mean curvature of S. If S is volume-preserving and perimeter stationary, then*

$$\mathcal{H}_0 = (\mathcal{H}_0)_S := \frac{\int_S \mathcal{H}_0 d\mu}{P_{\mathbb{H}}(S)} \text{ in } S \setminus \Sigma(S). \tag{8.14}$$

Proof. We use Lagrange multipliers and for fixed $V > 0$ consider the functional

$$\mathcal{A}(\Omega) = P_{\mathbb{H}}(S) - b(\Omega)\Big(|\Omega| - V\Big),$$

where the Lagrange multiplier b is a function of Ω. Critical points of $\mathcal{A}(\Omega)$ are volume-preserving and perimeter stationary. Variations $t \to \mathcal{A}(\Omega_t)$ along a vector field $U = a\nu_H$ with $a \in C_0^\infty(S \setminus \Sigma(S))$ give

$$\int_S \Big(\mathcal{H}_0 - b\Big) a d\mu - (|\Omega| - V) \int_S \frac{d}{dt} b(\Omega_t)|_{t=0}\, d\mu = 0,$$

and $|\Omega| = V$. From the latter we have

$$\int_S \Big(\mathcal{H}_0 - b\Big) a d\mu = 0 \quad \text{for all } a \in C_0^\infty(S \setminus \Sigma(S)).$$

The fundamental theorem of the calculus of variations [121] guarantees that $\mathcal{H}_0 = b$ is constant in $S \setminus \Sigma(S)$, hence completing the proof of (8.14). □

In the proof we have used the following elementary result, stated here for the reader's convenience.

Lemma 8.7. *If $S \subset \mathbb{H}$ is a C^2 surface enclosing a bounded region Ω, then S is volume-preserving and perimeter stationary if and only if S is a critical point of the functional $P_{\mathbb{H}}(S) - H(|\Omega| - V)$ for all variations supported in $S \setminus \Sigma(S)$ and for some choice of $H \in \mathbb{R}$.*

Remark 8.8. We give an equivalent way of proving (8.14) based on the first variation formula (6.34). Using the notation introduced in Section 6.4.2, we consider variations $\Omega_t = \phi_t(\Omega)$ along a vector field U defined on S. The variation is volume-preserving if

$$\frac{d}{dt}|\Omega_t|\bigg|_{t=0} = \int_S \langle U, \nu_1\rangle_1 d\sigma = 0.$$

Thus u corresponds to a volume-preserving variation if and only if

$$\int_S u\, d\sigma = 0. \tag{8.15}$$

As an aside, note that if $U = a\nu_H$ where $a \in C_0^\infty(S \setminus \Sigma(S))$, then $\int_S a d\mu = 0$ if and only if the corresponding variation along U is volume-preserving (since $\langle a\nu_H, \nu_1\rangle_1 d\sigma = a d\mu$). Inserting mean zero functions compactly supported away from the characteristic locus into (6.34) and using the Riemannian divergence theorem, we find that $\mathcal{H}_0 = \operatorname{div}_S \nu_H$ must be constant in $S \setminus S(\Sigma)$.

We state without proof a related result which will be used in the following sections.

Proposition 8.9 (Minkowski formula). *There exists a positive constant $C > 0$ such that if $S \subset \mathbb{H}$ is any C^2 volume-preserving, perimeter stationary surface enclosing a bounded region Ω, then $P_{\mathbb{H}}(S) = C\mathcal{H}_0|\Omega|$, where $\mathcal{H}_0$ is the (constant) horizontal mean curvature of S.*

We emphasize that the exact value of the constant C in Proposition 8.9 depends on our choice of metric in $\mathbb{H}$ adapted to the specific coordinate system which we fixed in the beginning. Note also that the Minkowski formula has an important corollary: *The horizontal mean curvature (away from the characteristic set) of a C^2 volume-preserving, perimeter stationary surface, computed with respect to the inner normal, is strictly positive.*

8.3.1 Parametrization of C^2 CMC t-graphs in $\mathbb{H}$

The observation that C^2 isoperimetric sets must have constant mean curvature in the sub-Riemannian sense points towards a possible approach to the isoperimetric problem: the classification of sub-Riemannian constant mean curvature (CMC) surfaces. As shown in Proposition 4.24, the horizontal curves foliating a constant mean curvature surface must be lifts of circles. We can parameterize surfaces of constant mean curvature in the Heisenberg group using this observation. To ensure the simplest and cleanest presentation, we restrict our attention to C^2 CMC(ρ) surfaces in $\mathbb{H}$ which are t-graphs, i.e., graphs over the complex plane $\mathbb{C}$. We note that this is not particularly restrictive as the candidates for solutions to the isoperimetric problem are graphs over $\mathbb{C}$ for most points on the surface.

So assume that S is a CMC(ρ) surface in $\mathbb{H}$ given as the graph of a C^2 function over a domain $\Omega \subset \mathbb{C}$. By the observation in Lemma 4.24, S must be foliated by horizontal lifts of circles. We can create an adapted parametrization of Ω, using circles in $\mathbb{C}$ as one parameter and picking a second parameter by specifying an arc length parameterized curve γ which is perpendicular to the foliation by circles at each point. More precisely, consider the parametrization

$$\begin{aligned} F(s,r) = \Bigg(&\gamma_1(s) + \gamma_1'(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_2'(s)\frac{\sin(\rho r)}{\rho}, \\ &\gamma_2(s) + \gamma_2'(s)\frac{1-\cos(\rho r)}{\rho} - \gamma_1'(s)\frac{\sin(\rho r)}{\rho}\Bigg) \end{aligned} \tag{8.16}$$

of the domain Ω. The curve γ is called a *seed curve* for S, following earlier nomenclature in the setting of minimal surfaces. If we are considering embedded surfaces, a seed curve is a smooth properly embedded planar curve. A quick inspection shows that this parametrization is not always a local diffeomorphism:

$$\begin{aligned} &\det DF \\ &= \det \begin{pmatrix} \gamma_1'(s) + \gamma_1''(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_2''(s)\frac{\sin(\rho r)}{\rho} & \gamma_2'(s) + \gamma_2''(s)\frac{1-\cos(\rho r)}{\rho} - \gamma_1''(s)\frac{\sin(\rho r)}{\rho} \\ \gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r) & \gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r) \end{pmatrix} \\ &= -\cos(\rho r) + \kappa(s)\frac{\sin(\rho r)}{\rho}, \end{aligned} \tag{8.17}$$

where κ is the curvature. In this computation, we take advantage of the facts that $\langle \gamma', \gamma'' \rangle(s) = 0$ and $\kappa(s) = \langle \gamma'', (\gamma')^\perp \rangle(s)$. We use also the convention $v^\perp = (v_2, -v_1)$ for $v = (v_1, v_2)$. The parametrization in (8.17) ceases to be a local diffeomorphism at

$$r = \frac{1}{\rho} \operatorname{arccot}\left(\frac{\kappa(s)}{\rho}\right).$$

We remark that taking the limit as $\rho \to 0$ yields the locus

$$r = \frac{1}{\kappa(s)} \tag{8.18}$$

in the case of sub-Riemannian minimal surfaces.

Next, we lift this parametrization of Ω to a parametrization of S in $\mathbb{H}$:

$$(F(s,r), h(s,r)).$$

Recalling that the circles, $F(s_0, r)$, must lift to horizontal curves, we investigate the form that h must take. Computing the derivative yields

$$\begin{aligned} F_r(s,r) &= (\gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r))\ X_1 + (\gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r))\ X_2 \\ &+ \left(h_r(s,r) - \frac{1}{2}\langle \gamma, (\gamma')^\perp \rangle(s)\sin(\rho r) + \frac{1}{2}\langle \gamma, \gamma' \rangle \cos(\rho r) - \frac{1-\cos(\rho r)}{\rho}\right) X_3. \end{aligned} \tag{8.19}$$

As $F(s_0, r)$ must be a horizontal curve, we have

$$h_r(s,r) = \frac{1}{2}\langle \gamma, (\gamma')^\perp \rangle(s)\sin(\rho r) - \frac{1}{2}\langle \gamma, \gamma' \rangle \cos(\rho r) + \frac{1-\cos(\rho r)}{\rho}.$$

Integrating from 0 to r gives

$$h(s,r) = h_0(s) + \frac{1}{2}\langle \gamma, (\gamma')^\perp \rangle(s)\frac{1-\cos(\rho r)}{\rho} - \frac{1}{2}\langle \gamma, \gamma' \rangle(s)\frac{\sin(\rho r)}{\rho} + \frac{r}{\rho} - \frac{\sin(\rho r)}{\rho^2} \tag{8.20}$$

with $h_0(s) = h(s, 0)$. We note that, taking the limit as $\rho \to 0$ yields the formula

$$h(s,r) = h_0(s) - \frac{1}{2}\langle \gamma, \gamma' \rangle(s) r \tag{8.21}$$

in accordance with the known result for sub-Riemannian minimal surfaces.

Calculating the s derivative yields:

$$\begin{aligned} F_s(s,r) = &\left(\gamma_1'(s) + \gamma_1''(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_2''(s)\frac{\sin(\rho r)}{\rho}\right)\ X_1 \\ &+ \left(\gamma_2'(s) + \gamma_2''(s)\frac{1-\cos(\rho r)}{\rho} - \gamma_1''(s)\frac{\sin(\rho r)}{\rho}\right)\ X_2 \\ &+ \left(h_0'(s) - \frac{\sin(\rho r)}{\rho} - \frac{1}{2}\langle \gamma, (\gamma')^\perp \rangle + \frac{1-\cos(\rho r)}{\rho^2}\kappa(s))\right)\ X_3. \end{aligned} \tag{8.22}$$

Note that at a characteristic point both F_s and F_r must be horizontal. Thus the characteristic locus is $\{\Sigma(s,r)=0\}$, where

$$\Sigma(s,r) = h_0'(s) - \frac{\sin(\rho r)}{\rho} - \frac{1}{2}\langle \gamma, (\gamma')^\perp\rangle + \frac{1-\cos(\rho r)}{\rho^2}\kappa(s). \tag{8.23}$$

We note that, as above, if we take the limit as $\rho \to 0$, we recover the known formula

$$h_0'(s) - r - \frac{1}{2}\langle \gamma, (\gamma')^\perp\rangle + \frac{1}{2}\kappa(s) r^2 = 0 \tag{8.24}$$

for the characteristic locus of a sub-Riemannian minimal surface.

So far, we have shown that every graphical CMC(ρ) surface has such a parametrization in a neighborhood of each noncharacteristic point. In fact, the existence of such parametrizations characterizes CMC(ρ) surfaces. To see this, we will show that, given a choice of $\gamma \in C^2$ and $h_0 \in C^1$, a patch of surface parameterized by $(F(s,r), h(s,r))$, $(s,r) \in \Omega$, has constant mean curvature. We start by computing the divergence of ν_H. From previous calculations,

$$\begin{aligned} F_r(s,r) &= (\gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r))\ X_1 \\ &\quad + (\gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r))\ X_2 \end{aligned}$$

and

$$\begin{aligned} F_s(s,r) = {}& \left(\gamma_1'(s) + \gamma_1''(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_2''(s)\frac{\sin(\rho r)}{\rho}\right)\ X_1 \\ &+ \left(\gamma_2'(s) + \gamma_2''(s)\frac{1-\cos(\rho r)}{\rho} - \gamma_1''(s)\frac{\sin(\rho r)}{\rho}\right)\ X_2 + \Sigma(s,r)X_3. \end{aligned}$$

Computing the Riemannian normal and projecting to the horizontal bundle give a nonunit horizontal normal vector:

$$\begin{aligned} &(\gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r))\Sigma(s,r)\ X_1 \\ &\quad - (\gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r))\Sigma(s,r)\ X_2. \end{aligned}$$

Normalizing gives the unit horizontal normal:

$$\begin{aligned} \nu_H &= (\gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r))\ X_1 \\ &\quad - (\gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r))\ X_2. \end{aligned}$$

To simplify notation, let

$$\overline{p} = (\gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r))$$

and

$$\overline{q} = -(\gamma_1'(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r)).$$

Note that

$$\overline{p}^2 + \overline{q}^2 = 1. \tag{8.25}$$

Then

$$\begin{aligned} \text{div } \nu_H &= \frac{\partial}{\partial x_1}\overline{p} + \frac{\partial}{\partial x_2}\overline{q} \\ &= \overline{p}_s \frac{\partial s}{\partial x_1} + \overline{p}_r \frac{\partial r}{\partial x_1} + \overline{q}_s \frac{\partial s}{\partial x_2} + \overline{q}_r \frac{\partial r}{\partial x_2} \\ &= \overline{p}_s \frac{(F_2)_r}{\det\ DF} - \overline{p}_r \frac{(F_2)_s}{\det\ DF} - \overline{q}_s \frac{(F_1)_r}{\det\ DF} + \overline{q}_r \frac{(F_1)_s}{\det\ DF} \end{aligned} \tag{8.26}$$

since $(x_1, x_2) = F(s, r)$. A direct computation shows that

$$\begin{aligned} (F_2)_r &= \gamma_2'(s)\sin(\rho r) - \gamma_1'(s)\cos(\rho r) = \overline{p}, \\ (F_1)_r &= \gamma_1(s)\sin(\rho r) + \gamma_2'(s)\cos(\rho r) = -\overline{q}, \\ -(F_2)_s &= -\gamma_2'(s) - \gamma_2''(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_1''(s)\frac{\sin(\rho r)}{\rho}, \\ (F_1)_s &= \gamma_1'(s) + \gamma_1''(s)\frac{1-\cos(\rho r)}{\rho} + \gamma_2''(s)\frac{\sin(\rho r)}{\rho}. \end{aligned} \tag{8.27}$$

Evaluating (8.26) using (8.17), (8.25) and (8.27) gives

$$\begin{aligned} \text{div } \nu_H &= \frac{1}{\det\ DF}\left(\overline{p}_s\overline{p} + \rho\overline{q}(F_2)_s + \overline{q}_s\overline{q} + \rho\overline{p}(F_1)_s\right) \\ &= \frac{1}{\det\ DF}\left(\rho\overline{q}(F_2)_s + \rho\overline{p}(F_1)_s\right) = \rho. \end{aligned} \tag{8.28}$$

To summarize, we have the following theorem:

Theorem 8.10. *If $S \subset \mathbb{H}$ is a C^2 graph over a set $\Omega \subset \mathbb{R}^2$ with empty characteristic locus, then S is CMC(ρ) if and only if for each $x_0 \in S$, there is a neighborhood of x_0, a seed curve $\gamma(s)$, and a height function $h(s,r)$, so that*

$$S = (F(\Omega), h(\Omega)),$$

where F and h are given in (8.16) *and* (8.20), *respectively.*

8.4 Existence and characterization of minimizers with additional symmetries

In this section we describe some results on the characterization of minimizers for the isoperimetric ratio in special classes of domains having suitable symmetry and regularity properties. To set the stage, we introduce the half-spaces $\mathbb{H}_+ = \{(z, x_3) \in \mathbb{H} : x_3 > 0\}$ and $\mathbb{H}_- = \{(z, x_3) \in \mathbb{H} : x_3 < 0\}$ and consider the collection

$$\mathcal{E} = \{E \subset \mathbb{H} : E \text{ satisfies (i) and (ii)}\},$$

where

(i) $|E \cap \mathbb{H}_+| = |E \cap \mathbb{H}_-|$, and

(ii) there exist $R > 0$, and functions $u, v : \overline{B_R} \to [0, \infty)$, with $u, v \in C^1(B_R) \cap C(\overline{B_R})$, $u = v = 0$ on ∂B_R, and such that

$$\begin{aligned} &\partial E \cap \mathbb{H}_+ = \{(z, x_3) \in \mathbb{H}_+ : |z| < R, x_3 = u(z)\} \\ \text{and} \qquad &\partial E \cap \mathbb{H}_- = \{(z, x_3) \in \mathbb{H}_- : |z| < R, x_3 = -v(z)\}. \end{aligned}$$

Here, and throughout this section, we write $B_R = B((0,0), R)$ for the ball of radius R in $\mathbb{R}^2$.

Note that the upper and lower portions of a set $E \in \mathcal{E}$ can be described by possibly different C^1 graphs, and that, besides C^1 smoothness, and the fact that their common domain is a metric ball, no additional assumptions are made on the functions u and v. The main result in this section characterizes the isoperimetric sets within the class $\mathcal{E}$.

Theorem 8.11 (Danielli–Garofalo–Nhieu). *Let $V > 0$, and define $R > 0$ so that $V = |\mathcal{B}(o, R)|$ (see 8.4). Then the variational problem*

$$\min_{E \in \mathcal{E} : |E| = V} P_{\mathbb{H}}(E)$$

has a unique solution in $\mathcal{E}$ given by the bubble set $\mathcal{B}(o, R)$.

We now present a step-by-step sketch of the proof of Theorem 8.11.

Step 1. First, we state without proof some invariance and symmetry properties of the horizontal perimeter. Consider the map $\mathcal{O} : \mathbb{H} \to \mathbb{H}$ defined by

$$\mathcal{O}(x_1, x_2, x_3) = (x_2, x_1, -x_3). \tag{8.29}$$

It is easy to see that $\mathcal{O}$ preserves the Lebesgue measure. Noting that the map $\mathcal{O}$ is an isometry of $(\mathbb{H}, d)$, it follows that it also preserves the horizontal perimeter: $P_{\mathbb{H}}(\mathcal{O}(E)) = P_{\mathbb{H}}(E)$ for every piecewise C^1 domain $E \subset \mathbb{H}$ with finite horizontal perimeter. (Note that the reflection alone $(x_1, x_2, x_3) \to (x_1, x_2, -x_3)$ is not an isometry of $\mathbb{H}$ and does not preserve the horizontal perimeter.)

Using a standard contradiction argument, we can establish the following symmetry result for isoperimetric sets whose intersection with the hyperplane $\{x_3 = 0\}$ is a 2-dimensional ball.

Theorem 8.12. *Let $E \subset \mathbb{H}$ be a bounded open set such that $\partial E \cap \mathbb{H}_+$ and $\partial E \cap \mathbb{H}_-$ are C^1 hypersurfaces, and assume that E satisfies the following condition:*

$$E \cap \{x_3 = 0\} = B_R \tag{8.30}$$

for some $R > 0$. Suppose E is an isoperimetric set satisfying $|E \cap \mathbb{H}_+| = |E \cap \mathbb{H}_-| = \frac{1}{2}|E|$. Then

$$P_{\mathbb{H}}(E; \overline{\mathbb{H}_+}) = P_{\mathbb{H}}(E; \overline{\mathbb{H}_-}).$$

Step 2. We now consider a domain $\Omega \subset \mathbb{R}^2$ and a C^1 function $u : \Omega \to [0, \infty)$. We assume that $E \subset \mathbb{H}$ is a C^1 domain enclosed by a t-graph, i.e., for which

$$E \cap \mathbb{H}_+ = \{(z, x_3) \in \mathbb{H} : z \in \Omega, 0 < x_3 < u(z)\}.$$

For $z = (x_1, x_2) \in \mathbb{R}^2$, we set $z^\perp = (x_2, -x_1)$. Indicating by $\phi(z, x_3) = x_3 - u(z)$ the defining function of $E \cap \mathbb{H}_+$, a simple computation gives

$$|\nabla_0 \phi| = \sqrt{\left|\partial_{x_1} u + \frac{x_2}{2}\right|^2 + \left|\partial_{x_2} u - \frac{x_1}{2}\right|^2} = \left|\nabla_z u + \frac{z^\perp}{2}\right|. \tag{8.31}$$

Invoking the representation formula (5.2) for the horizontal perimeter yields

$$P_{\mathbb{H}}(E, \mathbb{H}_+) = \int_{\partial E \cap \mathbb{H}_+} \frac{|\nabla_0 \phi|}{|\nabla \phi|} \, d\sigma,$$

and keeping in mind that $d\sigma = |\nabla \phi| \, dx_1 \, dx_2$, we obtain

$$P_{\mathbb{H}}(E, \mathbb{H}_+) = \int_\Omega \left|\nabla_z u + \frac{z^\perp}{2}\right| \, dx_1 \, dx_2 = \int_\Omega \sqrt{|\nabla_z u|^2 + \frac{1}{4}|z|^2 + \langle \nabla_z u, z^\perp \rangle} \, dx_1 \, dx_2. \tag{8.32}$$

Next, we introduce the relevant functional class for our problem. The class of competing functions is defined as follows.

Definition 8.13. We let $\mathcal{D}$ denote the set of functions $u \in C^{1,1}_{\mathrm{loc}}(B_R) \cap W^{1,1}(B_R)$ for which there exists $R > 0$ so that $u \geq 0$ in B_R,

$$\overline{B_R} = \bigcap \{B_{R+\rho} : \operatorname{supp}(u) \subset B_{R+\rho}\}.$$

We note explicitly that, if $u \in \mathcal{D}$ and R is as in Definition 8.13, then $u = 0$ on ∂B_R. Furthermore, functions in $\mathcal{D}$ may have large zero sets, e.g., the graph of such a function may touch the hyperplane $x_3 = 0$ in sets of large measure. We remark that $\mathcal{D}$ is not a vector space, nor is it a convex subset of $\mathcal{V}$. We mention that the requirement $u \in C^{1,1}_{\mathrm{loc}}(B_R)$ in the definition of the class $\mathcal{D}$, is justified by the following considerations. When we compute the Euler–Lagrange equation of the functional (8.32) we need to know that, with $\Omega = B_R$, the singular set $S_u = \{z = (x_1, x_2) \in \Omega \subset \mathbb{R}^2 : |\nabla_z u(z) + \frac{z^\perp}{2}| = 0\}$, which is the projection of the characteristic set of the graph of u (see (4.27)), has vanishing 2-dimensional Lebesgue measure. This is guaranteed by Theorem 4.48.

Step 3. Following classical ideas from the calculus of variations, we next introduce the admissible variations for the problem at hand, see [122] and [248].

Definition 8.14. Given $u \in \mathcal{D}$, we say that $\phi \in \mathcal{V}$, with $\operatorname{supp} \phi \subseteq \operatorname{supp} u$, is *$\mathcal{D}$-admissible at u* if $u + \lambda\phi \in \mathcal{D}$ for all $\lambda \in \mathbb{R}$ sufficiently small.

For $u \in \mathcal{D}$ we let

$$G[u] = \int_{\operatorname{supp}(u)} u(z) \, dx_1 \, dx_2 = \int_{B_R} u(z) \, dx_1 \, dx_2. \tag{8.33}$$

With (8.32) in mind, we define

$$J[u] = \int_{\mathrm{supp}(u)} \sqrt{|\nabla_z u|^2 + \frac{1}{4}|z|^2 + \langle \nabla_z u, z^\perp \rangle}\, dx_1\, dx_2 \tag{8.34}$$

for such u. Within the class of C^1 graphs over $\mathbb{R}^2$, the isoperimetric problem consists in minimizing the functional $J[u]$ subject to the constraint $G[u] = V$, where $V > 0$ is given and B_R is replaced by an *a priori* unknown domain Ω. We emphasize that finding the section of the isoperimetric profile with the hyperplane $\{x_3 = 0\}$, i.e., finding the domain Ω, constitutes here part of the problem. Because of the lack of an obvious symmetrization procedure, this seems a difficult question. To avoid this obstacle, we restrict the class of domains by requiring that their section with the hyperplane $\{x_3 = 0\}$ be a ball, i.e., we assume that, given $E \in \mathcal{E}$, there exists $R = R(E) > 0$ such that $\Omega = B_R$. Under this hypothesis, one can appeal to Theorem 8.12. The latter implies that it suffices to solve the following variational problem: *given* $V > 0$, *find* $R_0 > 0$ *and* $u_o \in \mathcal{D}$ *with* $\mathrm{supp}(u_o) = B_{R_0}$ *for which the following holds*:

$$J[u_o] = \min_{u \in \mathcal{D}} J[u] \quad \text{and} \quad G[u_o] = \frac{V}{2}. \tag{8.35}$$

Step 4. Next, we reduce (8.35) to an unconstrained problem using an application of the following standard version of the Lagrange multiplier theorem (see, e.g., Proposition 2.3 in [248]).

Proposition 8.15. *Let $\mathcal{D}$ be a subset of a normed vector space $\mathcal{V}$, and consider functionals $\mathcal{F}$, $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k$ defined on $\mathcal{D}$. Suppose there exist constants $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$, and $u_o \in \mathcal{D}$, such that u_o minimizes*

$$\mathcal{F} + \lambda_1 \mathcal{G}_1 + \lambda_2 \mathcal{G}_2 + \cdots + \lambda_k \mathcal{G}_k \tag{8.36}$$

(*uniquely*) *on $\mathcal{D}$. Then u_o minimizes $\mathcal{F}$* (*uniquely*) *when restricted to the set*

$$\{u \in \mathcal{D} : \mathcal{G}_j[u] = \mathcal{G}_j[u_o], j = 1, \ldots, k\}.$$

The procedure of applying the above proposition when solving a problem of the type

$$\min_{u \in \mathcal{D}} F[u]$$

subject to the constraints $\mathcal{G}_1[u] = V_1, \ldots, \mathcal{G}_k[u] = V_k$, consists of two steps. First, one shows that constants $\lambda_1, \ldots, \lambda_k$ and a function $u_o \in \mathcal{D}$ can be found so that u_o solves the Euler–Lagrange equation of (8.36), and u_o satisfies $\mathcal{G}_1[u_o] = V_1, \ldots, \mathcal{G}_k[u_o] = V_k$. Next, one proves that the solution u_o of the Euler–Lagrange equation is indeed a minimizer of (8.36). If the functional involved possesses appropriate convexity properties, then one can show in addition that such minimizer u_o is unique.

The constrained variational problem (8.35) is thus equivalent to the following one without constraint (provided the parameter λ is appropriately chosen): *minimize the functional*

$$\begin{aligned}\mathcal{J}[u] &= \int_{\text{supp}(u)} h(z, u(z), \nabla_z u(z))\, dx_1\, dx_2 \\ &= \int_{\text{supp}(u)} \left\{ \left| \nabla_z u(z) + \frac{z^\perp}{2} \right| + \lambda u(z) \right\} dx_1\, dx_2,\end{aligned} \tag{8.37}$$

over the set $\mathcal{D}$ *introduced in Definition* 8.13. It is easily recognized that the Euler–Lagrange equation of (8.37) is

$$\operatorname{div}_z \left[\frac{\nabla_z u + \frac{z^\perp}{2}}{\sqrt{|\nabla_z u|^2 + \frac{|z|^2}{4} + \langle \nabla_z u, z^\perp \rangle}} \right] = \lambda. \tag{8.38}$$

Step 5. As we pointed out above, solving (8.38) on an arbitrary domain of $\Omega \subset \mathbb{R}^2$ is a difficult task. However, when Ω is a ball in $\mathbb{R}^2$, the equation (8.38) admits a familiar class of spherically symmetric solutions. We note explicitly that for a graph $x_3 = u(z)$ with spherical symmetry in z, the only characteristic points can occur at the intersection of the graph with the x_3-axis.

Theorem 8.16. *Given* $R > 0$, *for every* $\lambda \in [-2/R, 0)$, *equation* (8.38) *with Dirichlet condition* $u = 0$ *on* ∂B_R, *admits the cylindrically symmetric solution* $u_{R,\lambda} \in \mathcal{D}$, *where*

$$u_{R,\lambda}(z) = \frac{1}{\lambda^2} \arccos\left(\frac{\lambda |z|}{2}\right) + \frac{|z|}{4\lambda} \sqrt{(4 - (\lambda |z|)^2} - C_{R,\lambda} \tag{8.39}$$

and

$$C_{R,\lambda} = \frac{1}{\lambda^2} \arccos\left(\frac{\lambda R}{2}\right) + \frac{R}{4\lambda} \sqrt{4 - (\lambda R)^2}. \tag{8.40}$$

Regarding the regularity of the functions $u_{R,\lambda}$, it suffices to consider the upper half of the "normalized" candidate isoperimetric profile $E_o \subset \mathbb{H}$, where ∂E_o is the graph of the function $x_3 = u_o(z)$, with $u_o = u_{1,-2}$. The characteristic locus of E_o is $\{(0, 0, \pm\frac{\pi}{8})\}$. Unlike its Euclidean counterpart, the hypersurface $S_o = \partial E_o$ is not C^∞ at the characteristic points $(0, 0, \pm\frac{\pi}{8})$. In fact, it is C^2, but not C^3, near its characteristic locus Σ. However, S_o is C^∞ (in fact, real-analytic) away from Σ.

One immediately obtains the following consequence.

Corollary 8.17. *Let* $V > 0$ *be given, and define* $R > 0$ *so that* $V = |\mathcal{B}(o, R)|$. *Let* $\lambda = -2/R$. *Then equation* (8.38), *with the Dirichlet condition* $u = 0$ *on* ∂B_R, *admits the radially symmetric solution* $u_R \in \mathcal{D} \cap C^2(B_R)$, *where*

$$u_R(z) = \frac{R^2}{4} \arccos\left(\frac{|z|}{R}\right) + \frac{|z|}{4} \sqrt{R^2 - |z|^2}. \tag{8.41}$$

Furthermore,

$$\int_\Omega u_R(z)\, dx_1\, dx_2 = \frac{1}{2}V. \tag{8.42}$$

At this point, recalling that (8.38) is the Euler–Lagrange equation of the unconstrained functional (8.37), we deduce the following result.

Theorem 8.18. *Let J and G be as in* (8.34) *and* (8.33) *respectively. Given $V > 0$ there exists $R = R(V) > 0$ so that the function $u_o = u_R$ in* (8.41) *is a critical point on $\mathcal{D}$ of the functional $J[u]$ subject to the constraint $G[u] = V/2$.*

Step 6. Our next objective is to prove that the function u_o in (8.41) is, first, a global minimizer of the variational problem (8.35), and second, the unique global minimizer. We will need some basic facts from the calculus of variations, which we now recall.

Definition 8.19. Let $\mathcal{V}$ be a normed vector space, and $\mathcal{D} \subset \mathcal{V}$. Given a functional $\mathcal{F} : \mathcal{D} \to \mathbb{R}$, $u \in \mathcal{D}$, and if ϕ is $\mathcal{D}$-admissible at u, one calls

$$\delta\mathcal{F}(u;\phi) \stackrel{\text{def}}{=} \lim_{\epsilon\to 0} \frac{\mathcal{F}[u+\epsilon\phi] - \mathcal{F}[u]}{\epsilon}$$

the *Gâteaux derivative* of $\mathcal{F}$ at u in the direction ϕ if the limit exists.

Definition 8.20. Let $\mathcal{V}$ be a normed vector space, and $\mathcal{D} \subset \mathcal{V}$. Consider a functional $\mathcal{F} : \mathcal{D} \to \bar{\mathbb{R}}$. $\mathcal{F}$ is said to be *convex* over $\mathcal{D}$ if for every $u \in \mathcal{D}$, and every $\phi \in \mathcal{V}$ such that ϕ is $\mathcal{D}$-admissible at u, and $u + \phi \in \mathcal{D}$, one has

$$\mathcal{F}[u+\phi] - \mathcal{F}[u] \geq \delta\mathcal{F}(u;\phi),$$

whenever the right-hand side is defined. We say that $\mathcal{F}$ is *strictly convex* if strict inequality holds in the above inequality except when $\phi \equiv 0$.

We then have the following result.

Theorem 8.21. *Suppose $\mathcal{F}$ is convex and proper over a nonempty convex subset $\mathcal{D}^* \subset \mathcal{V}$ (i.e., $\mathcal{F} \not\equiv \infty$ over $\mathcal{D}^*$), and suppose that $u_o \in \mathcal{D}^*$ is such that $\delta\mathcal{F}(u_o;\phi) = 0$ for all ϕ which are $\mathcal{D}^*$-admissible at u_o (that is, u_o is a critical point of the functional $\mathcal{F}$), then $\mathcal{F}$ has a global minimum in u_o. If moreover $\mathcal{F}$ is strictly convex at u_o, then u_o is the unique element in $\mathcal{D}^*$ satisfying*

$$\mathcal{F}[u_o] = \inf\big\{\mathcal{F}[v] : v \in \mathcal{D}^*\big\}\ .$$

Our next goal is to adapt the above results to (8.35). Given $V > 0$ consider the number $R = R(V) > 0$ defined in Corollary 8.17, the corresponding fixed ball B_R, and the normed vector space $\mathcal{V}(R) = \{u \in C(\overline{B_R}) : u = 0 \,\text{on}\, \partial B_R\}$. Let $\mathcal{D}(R)$ be the collection of functions $u \in \mathcal{V}(R)$ with $u \geq 0$, $u \in C^2(B_R) \cap W^{1,1}(B_R)$, and

$$\overline{B_R} = \bigcap\{B_{R+\rho} : \operatorname{supp}(u) \subset B_{R+\rho}\}\}.$$

We note that $\mathcal{D}(R)$ is a nonempty convex subset of $\mathcal{V}(R)$, and that $u = 0$ on ∂B_R for every $u \in \mathcal{D}(R)$. Consider the functional (8.37). Given $u \in \mathcal{D}(R)$ and ϕ which is $\mathcal{D}(R)$-admissible at u, in view of Theorem 4.48, we see that $\mathcal{J}$ is Gâteaux differentiable at u in the direction of ϕ, and

$$\begin{aligned} \delta\mathcal{J}(u;\phi) &= \int_{B((0,0),R)} \Big\{ h_u(z,u(z),\nabla u(z))\,\phi(z) \\ &\qquad + \langle \nabla_p h(z,u(z),\nabla u(z)), \nabla\phi(z)\rangle \Big\}\, dx_1 dx_2 \\ &= \int_{B_R} \left\{ \frac{\langle \nabla_z u + z^\perp/2, \nabla_z \phi\rangle}{|\nabla_z u + z^\perp/2|} + \lambda\,\phi \right\} dx_1 dx_2. \end{aligned} \tag{8.43}$$

At this point, using an algebraic agument, it is possible to show the existence of a global minimizer of $\mathcal{J}$. Such global minimizer is indeed provided by the spherically symmetric function u_R in (8.41).

Proposition 8.22. *Given $V > 0$, let $R = R(V) > 0$ be as in Corollary 8.17. The functional $\mathcal{J}$ in (8.37) is convex on $\mathcal{D}(R)$. As a consequence, the function u_R in (8.41) is a global minimizer of $\mathcal{J}$ on $\mathcal{D}(R)$.*

Finally, the proof of Theorem 8.11 is complete once one shows that u_o is the unique minimizer of the variational problem (8.35). This, in turn, follows from the fact that for every function ϕ, not identically zero, which is $\mathcal{D}(R)$-admissible at u_R, the strict inequality

$$\mathcal{J}[u_R + \phi] \;>\; \mathcal{J}[u_R]$$

holds.

8.5 The C^2 isoperimetric profile in $\mathbb{H}$

In this section we sketch an argument showing that the bubble sets are the isoperimetric minimizers in the category of C^2 surfaces.

Theorem 8.23 (Ritoré–Rosales). *If Ω is an isoperimetric region in $\mathbb{H}$ which is bounded by a C^2 smooth surface S, then S is congruent (i.e., equivalent by the composition of a Heisenberg isometry and a dilation) to the boundary of a bubble set.*

In contrast to the previous sections, which are quite analytic or measure theoretic in spirit, the proof of Theorem 8.23 relies heavily on techniques of differential geometry. We give a step-by-step outline of the proof and present some of the main ideas. Note that in the results presented below, only the bare minimum is shown to illustrate the ideas of the proof (see the notes at the end of the chapter for detailed references).

Step 1. The starting point is the study of the characteristic set of a C^2 isoperimetric profile.

Proposition 8.24. *Let $S \subset \mathbb{H}$ be an oriented volume-preserving perimeter stationary C^2 compact surface enclosing a region Ω. If the characteristic locus $\Sigma(S)$ contains a C^1 curve C, then the rules of the Legendrian foliation of S meet C orthogonally.*

Proof. Recall that by Proposition 8.6, S is CMC. In view of Section 8.3, CMC surfaces are ruled by horizontal curves of fixed curvature. To prove the proposition assume that C is a curve in $\Sigma(S)$. By Lemma 4.32, C is C^1. Let $B \subset S$ be such that $B \setminus C$ is the union of two open connected sets $B_\pm$. Let n_+ be the inward pointing normal to B_+. Finally, let $u : B \to \mathbb{R}$ be a mean zero function with compact support in B so that $u|_C$ is supported on $C \cap B$. Using this u in (6.33) and taking into account the assumption that S is perimeter stationary, we have

$$\begin{aligned}
0 &= \int_{B\setminus\Sigma(S)} u \operatorname{div}_S \nu_H \, d\sigma - \int_{B\setminus\Sigma(S)} \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) \, d\sigma \\
&= -\int_{B\setminus\Sigma(S)} \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) \, d\sigma
\end{aligned}$$

(since $\operatorname{div}_S \nu_H$ is constant in $S \setminus \Sigma(S)$)

$$\begin{aligned}
&= -\int_{B_+\setminus\Sigma(S)} \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) \, d\sigma - \int_{B_-\setminus\Sigma(S)} \operatorname{div}_S(u(\nu_H)_{\mathrm{tang}}) \, d\sigma \\
&= \int_C u\langle n_+, \nu_H^+\rangle_1 \, d\sigma - \int_C u\langle n_+, \nu_H^-\rangle_1 \, d\sigma
\end{aligned}$$

(by the divergence theorem)

$$= 2\int_C u\langle n_+, \nu_H^+\rangle_1 \, d\sigma,$$

where $\nu_H^\pm(q) = \lim_{p\in B_\pm, p\to q} \nu_H(p)$ for $q \in C$, and $\nu_H^+(q) = -\nu_H^-(q)$ by Proposition 4.34. Since we may choose u so that $u|_C$ is arbitrary, we conclude that $\langle n_+, \nu_H^+\rangle_1 = 0$ on all of $C \cap B$. Hence ν_H^+ is tangent to C in B, so the rules approaching C meet C orthogonally. □

Step 2. The second step of the proof is to show improved regularity of curves in the characteristic locus.

Proposition 8.25. *Let $S \subset \mathbb{H}$ be an oriented perimeter stationary C^2 compact surface. If the characteristic locus $\Sigma(S)$ contains a C^1 curve C, then C is in fact C^2 regular.*

This result follows from Lemma 4.32 and an analysis of the surface using the seed curve/height function parametrization of Section 8.3.
The regularity of characteristic curves leads to a crucial property of the characteristic locus of oriented volume-preserving perimeter stationary compact surfaces.

Theorem 8.26. *Let S be a complete, oriented C^2 immersed volume-preserving perimeter stationary surface in $\mathbb{H}$ with nonvanishing mean curvature. Then any connected curve in $\Sigma(S)$ is a geodesic.*

This theorem follows from a careful study of the behavior of the rules of the surface emanating from a characteristic curve to determine when they return to the characteristic locus. With the observation that the rules meet the characteristic locus orthogonally, Theorem 8.26 may also be derived from the work in Section 8.3. We sketch the proof and indicate some of the main computations.

Sketch of the proof. By (8.23), we know that, for a given seed curve/height function pair (γ, h_0), the characteristic locus arises when $\Sigma(s,r) = 0$. By readjusting our choice of seed curve and height function, we may assume that $\Sigma(s,0) = 0$, i.e., a portion of the characteristic locus occurs at $r = 0$. The cautious reader will note that the computations in Section 8.3 are performed away from the characteristic locus. However, in equation (8.23) we show the behavior of the characteristic locus if we extend the parametrization to all values of r. It is not hard to see that (8.23) and the seed curve/height function representation may be used in this context as well. Note that here we must invoke Proposition 8.25 and assert that curves in the characteristic locus are C^2. In fact, our seed curves must be at least C^2. Inspection of (8.23) leads to the following condition on h_0:

$$h_0'(s) = \frac{1}{2}\langle \gamma, (\gamma')^{\perp}\rangle(s).$$

Consider the case where $\Sigma(S)$ has another component that contains a C^2 smooth curve. There exists $r = r(s)$ so that this curve is given by $(F(s,r(s)), h(s,r(s)))$, i.e., $\Sigma(s,r(s)) = 0$. Since $h_0'(S) = \frac{1}{2}\langle \gamma, (\gamma')^{\perp}\rangle(s)$, we have

$$\Sigma(s,r) = -\frac{\sin(\rho r)}{\rho} + \frac{1-\cos(\rho r)}{\rho^2}\kappa(s) \tag{8.44}$$

and hence $r(s)$ is a solution of $\Sigma(s,r) = 0$. Next, we wish to use the fact that the rules meet Σ orthogonally. We first note that

$$0 = \langle F_s, F_r\rangle = \sin(\rho r(s)) - \kappa(s)\Big(\frac{1-\cos(\rho r(s))}{\rho}\Big)$$

using (8.19) and (8.22). Hence, for $r = r(s_0)$, $F_s(s_0, r(s_0))$ is perpendicular to $F_r(s_0, r(s_0))$. Since

$$\partial_s F(s,r(s)) = F_s(s,r(s)) + F_r(s,r(s))r'(s)$$

we conclude that the rules meet the characteristic locus orthogonally if and only if $r(s)$ is constant. Inspection of (8.44) shows that this may only happen if $\kappa(s)$ is also constant. Hence the seed curve is a portion of a straight line or a circle and, as we have arranged that it lifts to a horizontal curve, we conclude that $(F(s,0), h_0(s))$ is a Heisenberg geodesic. □

Step 3. By the previous steps, any compact C^2 solution to the isoperimetric problem cannot have a nontrivial curve in its characteristic locus, as any such curve would be a geodesic which would leave any bounded domain in a finite time. Thus, the characteristic locus may contain only isolated points. The third and final step of the proof is to show that if the characteristic locus of the surface contains an isolated point, then it is congruent to the boundary of a bubble set.

Theorem 8.27. *Let S be a complete, connected C^2 oriented immersed surface in $\mathbb{H}$ with nonvanishing constant mean curvature. If $\Sigma(S)$ contains an isolated point, then S is congruent to the boundary of a bubble set.*

Sketch of the proof. The nucleus of the argument is contained in Theorem 4.38, which indicates that, at each isolated characteristic point p, there are values $r > 0$ and $\rho \in \mathbb{R}$ so that

$$\{\gamma_{p,v,\rho}(s) : v \in H_pS, |v| = 1, s \in [0, r)\}$$

is a proper subset of S. Here, $\gamma_{p,v,\rho}$ denotes the sub-Riemannian geodesic passing through p with tangent v and curvature ρ. Direct computation shows that if $r = \pi/|\rho|$, then the resulting surface is congruent with the boundary of a bubble set. The theorem follows by analyzing the cases $r < \pi/|\rho|$ and $r > \pi/|\rho|$ and using the structure of the geodesics of the Heisenberg group. □

These facts impose significant constraints on the class of C^2 CMC surfaces:

Theorem 8.28. *Let S be a compact, connected C^2 immersed volume-preserving perimeter stationary surface in $\mathbb{H}$. Then S is congruent to the boundary of a bubble set.*

Sketch of the proof. From the Minkowski formula (Proposition 8.9) we conclude that the horizontal mean curvature of S computed with respect to the inner normal is strictly positive, hence nonvanishing. Next, by compactness we deduce that S contains at least one characteristic point. If $\Sigma(S)$ contains a curve C, then Theorem 8.26 implies that C is a complete geodesic; as such geodesics leave any bounded set in finite time, this would violate the compactness assumption. Thus $\Sigma(S)$ may contain only isolated points. By Theorem 8.27, S is congruent to the boundary of a bubble set. □

Coupled with Theorem 8.3 and the characterization of isoperimetric sets as CMC surfaces, Theorem 8.28 gives a proof of Theorem 8.23.

8.6 The convex isoperimetric profile of $\mathbb{H}$

In this section we present a proof of Pansu's conjecture within the category of Euclidean convex sets.

Theorem 8.29 (Monti–Rickly). *If Ω is a (Euclidean) convex region in $\mathbb{H}$ which is an isoperimetric set, then Ω is congruent to a bubble set.*

From the point of view of regularity, the assumption in Theorem 8.29 (Euclidean convexity) is weaker than that in Theorem 8.23 (Euclidean Lipschitz vs. C^2 regularity). The general strategy in the proof of Theorem 8.29 is to show that patches of $\partial\Omega$ can be parameterized by lifts of circles, and then use the convexity hypothesis to conclude the argument.

For the remainder of this section, convexity always refers to convexity in the Euclidean sense.

Convex domains in $\mathbb{H}$. Let $\Omega \subset \mathbb{H}$ be a bounded convex open set. A supporting plane for Ω at $p \in \partial\Omega$ for Ω is simply a plane Π through p which has empty intersection with Ω. The notion of supporting plane generalizes the concept of tangent plane to the setting of non-smooth convex sets.

The characteristic set of a convex domain (with no further regularity assumptions) may be defined as the set of $p \in \partial\Omega$ for which the horizontal plane $H(p)$ at p is a supporting plane for Ω.

We sketch the proof of Theorem 8.29 in a series of steps.

Step 1. In the first stage, we study the structure of the characteristic locus for convex domains, proving the following results.

1. Convex and bounded C^1 sets have at least one characteristic point, strictly convex bounded sets have at most two.
2. In general, if Ω is a convex domain containing the origin, then $\Sigma(\partial\Omega)$ can be written as the union of two disjoint components $\Sigma_\pm$, separated by the horizontal plane at the origin, each of which is a (possibly degenerate) horizontal segment.
3. If two geodesic (hence horizontal) arcs with the same curvature $H > 0$ are contained in $\partial\Omega$ and intersect at $p \in \partial\Omega$, then they must coincide in a neighborhood of p.

Step 2. Assume now that Ω is a convex isoperimetric set in $\mathbb{H}$. Decompose $\partial\Omega$ in four patches, which are graphs over convex planar domains $A \subset \mathbb{R}^2$ of the form $x_3 = f(x_1, x_2)$(for the "top" and "bottom" portions), and $x_1 = g(x_2, x_3)$ (for the "lateral" portion), with $f, g : A \to \mathbb{R}$ convex (hence Lipschitz continuous with BV derivatives). Recalling (4.27), we denote by S_f the singular set of points $(x_1, x_2) \in A$ such that $(x_1, x_2, f(x_1, x_2)) \in \Sigma(\partial\Omega)$. In a similar fashion, we define the singular set S_g.

Step 3. Use a variational argument similar to the ones described in the previous sections to derive curvature equations for convex isoperimetric profiles. Additional difficulties arise in this non-smooth setting, as one needs first to show that for any compact set K disjoint from S_f or S_g, there exists $\delta > 0$ such that

$$|\nabla_0(x_3 - f(x_1,x_2))|, |\nabla_0(x_1 - g(x_2,x_3))| > \delta \tag{8.45}$$

a.e. in K. With such an estimate at hand, and in view of Remark 5.10, we may represent the perimeter of the graph of f in integral form:

$$\int_A |\nabla_0(x_3 - f(x_1,x_2))| dx_1 dx_2 = \int_A \Big| - \nabla f + \frac{1}{2}(-x_2, x_1) \Big| dx_1 dx_2.$$

If we write $P = P_{\mathbb{H}}(\Omega)$ and $V = |\Omega|$ for the perimeter and measure of Ω, respectively, and consider the portion of $\partial\Omega$ arising as the graph of f, we find that f must satisfy the PDE

$$\operatorname{div}_{x_1,x_2}\left(\frac{-\nabla f + \frac{1}{2}(-x_2,x_1)}{|-\nabla f + \frac{1}{2}(-x_2,x_1)|} \right) = \frac{3P}{4V}, \tag{8.46}$$

in the distributional sense in $A \setminus S_f$. On the other hand, again in view of Remark 5.10, we may represent the perimeter of the graph of g in integral form:

$$\int_A |\nabla_0(x_1 - g(x_2,x_3))| dx_2 dx_3 = \int_A \left| \Big(1 + \frac{1}{2} x_2 \partial_{x_3} g, -\partial_{x_2} g - \frac{1}{2} g \partial_{x_3} g\Big) \right| dx_2 dx_3.$$

If we consider the portion of $\partial\Omega$ arising as the graph of g, we find that g must satisfy the PDE

$$\frac{1}{2} g(x_2,x_3) \frac{\partial}{\partial x_3}\left(\frac{1 + \frac{1}{2} x_2 \partial_{x_3} g(x_2,x_3)}{|\nabla_0(x_1 - g(x_2,x_3))|} \right) + X_2 \left(\frac{-\partial_{x_2} g - \frac{1}{2} g \partial_{x_3} g(x_2,x_3)}{|\nabla_0(x_1 - g(x_2,x_3))|} \right) = \frac{3P}{4V}, \tag{8.47}$$

in the distributional sense in $A \setminus S_g$. We will refer to the value $3P/4V$ as the *curvature* of the isoperimetric profile Ω.

Step 4. Next, one realizes that the distributional interpretation of the PDEs (8.46) and (8.47) is not sufficient to make progress towards the final result. What is needed is an extra measure of regularity for the solutions, namely, one wants the PDEs to hold in the weak Sobolev sense rather than in the distributional sense. Such an improvement in regularity is achieved through the following proposition.

Proposition 8.30. *Let $A \subset \mathbb{R}^2$ be a bounded open set and let $\vec{u} = (u_1, u_2) : A \to \mathbb{R}^2$ be a vector field whose components u_1, u_2 are BV functions. Assume that*

(i) *there exists $\delta > 0$ such that $|\vec{u}| > \delta$ a.e. in A,*
(ii) $\operatorname{div} \vec{u}^{\perp} \in L^1(A)$,
(iii) $\operatorname{div}(\vec{u}/|\vec{u}|) \in L^1(A)$.

Then $\vec{u}/|\vec{u}| \in W^{1,1}(A, \mathbb{R}^2)$.

We apply this proposition to the vector field $\vec{u} = \nabla_0(x_3 - f(x_1, x_2)) = -\nabla f + 1/2(-x_2, x_1)$ in any compact set which avoids S_f. Observe that (i) follows from (8.45). A direct computation yields $\operatorname{div} \vec{u}^{\perp} = -1$, hence (ii) is satisfied. Assumption (iii) follows from (8.46). We deduce that

$$\frac{\nabla_0(x_3 - f(x_1, x_2))}{|\nabla_0(x_3 - f(x_1, x_2))|} = \frac{-\nabla f + \frac{1}{2}(-x_2, x_1)}{|-\nabla f + \frac{1}{2}(-x_2, x_1)|} \in W^{1,1}(A \setminus S_f, R^2).$$

A similar result holds for the patch given by $x_1 = g(x_2, x_3)$:

$$\frac{\nabla_0(x_1 - g(x_2, x_3))}{|\nabla_0(x_1 - g(x_2, x_3))|} \in W^{1,1}(A \setminus S_g, \mathbb{R}^2). \tag{8.48}$$

Step 5. The crucial step in the proof of Theorem 8.29 is the study of the Legendrian foliation of a convex isoperimetric profile. Observe that for any convex function $f : A \subset \mathbb{R}^2 \to \mathbb{R}$, the vector fields $\vec{u} = -\nabla f + 1/2(-x_2, x_1)$ and $\vec{v} = 1/2(x_1, x_2) - (-\partial_{x_2} f, \partial_{x_1} f)$ are in $BV_{\text{loc}}(A, \mathbb{R}^2) \cap L^{\infty}(A)$. Moreover $\operatorname{div} \vec{v} = -1 \in L^{\infty}(A)$. These facts suffice to apply Ambrosio's extension [8] of results of DiPerna–Lions [87] on flows generated by Sobolev vector fields: if $\vec{v} \in BV_{\text{loc}}(\mathbb{R}^2, \mathbb{R}^2) \cap L^{\infty}(\mathbb{R}^2, \mathbb{R}^2)$ has bounded divergence, then for any compact set $K \subset \mathbb{R}^2$ and $\rho > 0$ there exists a Lagrangian flow $\phi : K \times [-\rho, \rho] \to \mathbb{R}^2$ starting from K and relative to $\vec{v}$, i.e., $s \mapsto \phi(q, s)$ solves $\phi(q, s) = q + \int_0^s \vec{v}(\phi(q, t))\, dt$ for all $s \in [-\rho, \rho]$. The resulting flow is stable with respect to smooth approximations of $\vec{v}$ in the L^1 norm.

We apply these results to the convex functions f, g in the graphical patch representation of a convex isoperimetric profile to obtain:

Theorem 8.31. *Let $\Omega \subset \mathbb{H}$ be a convex isoperimetric profile with mean curvature H. If we locally represent a portion of $\partial\Omega$ as a convex graph $x_3 = f(x_1, x_2)$ over a convex region $A \subset \mathbb{R}^2$, then for all compact sets $K \subset A \setminus S_f$ and open neighborhoods of K, $K \subset O \subset A \setminus S_f$, there exists a sufficiently small $\rho > 0$ and a Lagrangian flow $\phi : K \times [-\rho, \rho] \to O$ relative to $\vec{v} = 1/2(x_1, x_2) - (-\partial_{x_2} f, \partial_{x_1} f)$ such that for a.e. $z \in K$, the curve $s \to \phi(z, s)$ is an arc of a circle (oriented clockwise) of radius $1/H$.*

The flow ϕ is regular in the following sense: there exists a constant $\lambda \geq 1$ such that for all measurable sets $A \subset K$ and $s \in [-\rho, \rho]$ one has

$$\frac{1}{\lambda}|A| \leq |\phi(A, s)| \leq \lambda |A|. \tag{8.49}$$

Note that for each $z \in K$, the vector $\vec{v}(z)$ is the projection to $\mathbb{C}$ of the horizontal vector

$$v_1 X_1 + v_2 X_2 = \left(\frac{1}{2}x_1 + \partial_{x_2} f, \frac{1}{2}x_2 - \partial_{x_1} f, \frac{-1}{2}[x_1 \partial_{x_1} f + x_2 \partial_{x_2} f]\right)$$

which is a.e. tangent to $\partial\Omega$ in the graphical patch determined by f. Consequently the Lagrangian flow ϕ lifts to a geodesic foliation of the coordinate patch, away from characteristic points.

Aside: sketch of the proof of Theorem 8.31. The main problem is to compute the second derivatives of an integral curve of $\vec{v}$ and interpret the PDE (8.46) pointwise. This is not trivial because while $\vec{v}$ admits a regular Lagrangian flow (in view of the prior discussion), it is only in BV_{loc}. On the other hand, we might be tempted to use integral lines of the normalized vector field $\vec{v}/|\vec{v}|$ which is in $W^{1,1}$, and so such curves would be twice differentiable a.e. However one cannot directly define a Lagrangian flow of $\vec{v}/|\vec{v}|$ since its divergence is only in L^1 and not in L^∞.

To resolve this problem one first observes that, since $\vec{v}$ and its normalization are parallel, one can find an integral curve for one by reparametrizing integral curves of the other. In standard fashion, consider a suitable reparametrization of the flow ϕ,

$$\gamma(s) = \phi(z, \tau(s)) \tag{8.50}$$

defined so that γ is an integral curve of the vector field $\vec{v}/\lambda$ and $\lambda : A \to \mathbb{R}$ is a measurable function (to be chosen later) satisfying $0 < c_1 \leq \lambda \leq c_2$, for some constants $c_1, c_2 > 0$.

At this point, we observe that in compact sets outside of S_f we may set $\lambda = |\vec{v}|$, and recall that the normalized vector field $\vec{w} = \vec{v}/|\vec{v}|$ is in $W^{1,1}$. To conclude that the integral curves γ of $\vec{w}$ defined above through a reparametrization have second derivatives a.e., we need to use a special *chain rule* for the composition of $W^{1,1}$ vector fields $\vec{w}$ and curves γ defined as in (8.50), namely:

$$\vec{w} \circ \gamma \in W^{1,1}$$

and

$$\frac{d}{ds}(\vec{w} \circ \gamma)(s) = (\nabla \vec{w} \circ \gamma)\gamma'(s) \qquad \text{a.e.}$$

Because of our choice of $\vec{w}$ and in view of equation (8.46) we immediately obtain

$$\gamma'' = -H(\gamma') \qquad \text{a.e.}$$

From this ODE we immediately deduce the smoothness of γ and the desired result. □

An analogous result holds for the graphical patches determined by g.

Step 6. Thanks to the above results and in view of basic extension and uniqueness arguments, every convex isoperimetric profile $\Omega \subset \mathbb{H}$ with curvature H has boundary $\partial\Omega$ foliated by geodesics, which are lifts of circles of radius $1/H$ and which have one endpoint in Σ_+ and the other in Σ_-. The only thing left to show is that $\Sigma_\pm$ consist each of a single point (the poles of the bubble set). To prove this we argue by contradiction and assume without loss of generality that Σ_- contains a horizontal segment of the form $\{(0, x_2, 0) \ : |x_2| < \mu\}$ for some $\mu > 0$.

Because of convexity, and from the definition of characteristic points, one must have

$$\Omega \subset H\left(\left(0, \frac{\mu}{2}, 0\right)\right)^+ \cap H\left(\left(0, -\frac{\mu}{2}, 0\right)\right)^+, \tag{8.51}$$

where we denote by $H(x)^+$ the component of $\mathbb{H} \setminus H(x)$ containing $\{(0,0,x_3) : x_3 > R\}$ for some sufficiently large $R > 0$. The right-hand side of (8.51) forms a wedge, whose edge is in $\partial\Omega$ and contains the origin. An elementary computation shows that any smooth geodesic arc emanating from the origin o must necessarily have a horizontal tangent at o; the continuity of the tangent then contradicts (8.51).

Step 7. The two points $\Sigma_\pm$ must both lie on the x_3 axis, since otherwise there would be only one geodesic arc joining them and we know that any curve in the geodesic foliation joins these two points.

Step 8. In conclusion, $\partial\Omega$ is foliated by geodesics, lifts of circles with radius $1/H$ touching the origin o, with two isolated characteristic points $\Sigma_\pm$. It follows that Ω is congruent with a bubble set.

8.7 Other approaches

8.7.1 Riemannian approximation approach to the isoperimetric problem

Pansu's conjectured extremals can also be recovered by solving the CMC equation on the Riemannian approximants and taking the limit as $L \to \infty$. The derivation of the cylindrically symmetric constant mean curvature surfaces in $(\mathbb{R}^3, g_1)$ was given by Tomter [247]. In this section, we reproduce his result in the slightly more general setting of $(\mathbb{R}^3, g_L)$ and study the sub-Riemannian limit.

The key idea in Tomter's proof is to utilize the rotational invariance of the Heisenberg group to reduce the CMC equation on $(\mathbb{R}^3, g_L)$ to a system of ODEs on the orbit (2-)manifold. This system may be solved explicitly by quadrature, generating a foliation of $(\mathbb{R}^3, g_L)$ by closed $U(1)$-invariant surfaces with constant mean curvature $H > 0$. We present the derivation of this foliation, together with asymptotic formulas describing the volume and surface area of the resulting domains as series in the curvature parameter H and in the metric parameter L.

Consider the action of the circle group $U(1)$ on $\mathbb{R}^3$ given by

$$e^{\mathbf{i}\,\theta}(z, x_3) = (e^{\mathbf{i}\,\theta} z, x_3), \qquad z = x_1 + \mathbf{i}\,x_2.$$

Observe that this action is by isometries on the manifolds $(\mathbb{R}^3, g_L)$, where g_L is the metric defined in Section 2.4. We denote by

$$\overline{M}_L = (\mathbb{R}^3, g_L)/U(1) = \{(r, x_3) : r = |z| \geq 0, x_3 \in \mathbb{R}\}$$

the orbit manifold, equipped with orbital distance metric[2]

$$\overline{g}_L(r, x_3) = dr^2 + \frac{4L}{4 + Lr^2} dx_3^2. \tag{8.52}$$

The length of the orbit $U(1) \cdot (r, x_3)$ is

$$\operatorname{Length} U(1) \cdot (r, x_3) = \int_0^{2\pi} \left\langle \frac{\partial}{\partial\theta}, \frac{\partial}{\partial\theta} \right\rangle^{1/2} d\theta,$$

where $\theta = \arg z = \arctan(x_2/x_1)$. From the choice of the frame in $(\mathbb{R}^3, g_L)$, we find

$$\frac{\partial}{\partial\theta} = -x_2 X_1 + x_1 X_2 + \frac{1}{2}\sqrt{L} r^2 \widetilde{X_3}$$

so

$$\operatorname{Length} U(1) \cdot (r, x_3) = \int_0^{2\pi} \sqrt{r^2 + \frac{1}{4} L r^4}\, d\theta = \pi r \sqrt{4 + Lr^2}. \tag{8.53}$$

Our goal is to describe a family of closed $U(1)$-invariant CMC surfaces in $(\mathbb{R}^3, g_L)$. Let S be a $U(1)$-invariant surface in $(\mathbb{R}^3, g_L)$, obtained as the orbit of a curve γ in $\overline{M}_L$. We assume that $\gamma = \gamma(t)$ is parameterized by arc length. By a formula from equivariant geometry, the mean curvature of S is

$$H = k_\gamma \circ \pi - \partial_\nu \log \operatorname{Length} U(1) \cdot \pi, \tag{8.54}$$

where k_γ denotes the curvature of γ in $\overline{M}_L$, $\pi : (\mathbb{R}^3, g_L) \to \overline{M}_L$ is the quotient map, and ∂_ν is the normal derivative. If we denote by $\alpha = \alpha(t)$ the angle between $\partial/\partial r$ and the tangent vector $\dot\gamma(t)$, then (8.54) yields the following criterion for S to be a CMC surface in $(\mathbb{R}^3, g_L)$.

Lemma 8.32. *S has constant mean curvature H if and only if $(r(t), x_3(t), \alpha(t))$ is a solution to the system*

$$\dot r = \cos\alpha, \qquad \dot{x_3} = \frac{1}{2}\frac{\sqrt{4 + Lr^2}}{\sqrt{L}} \sin\alpha, \qquad \dot\alpha = H - \frac{\sin\alpha}{r}. \tag{8.55}$$

Proof. The map

$$(r, x_3) \overset{\iota}{\mapsto} (\varphi(r)\cos x_3, \varphi(r)\sin x_3, \psi(r)),$$

where

$$\varphi(r) = \frac{2\sqrt{L}}{\sqrt{4 + Lr^2}} \quad \text{and} \quad \psi(r) = \int_0^r \sqrt{1 - \varphi'(\rho)^2}\, d\rho$$

[2]The expression for $\overline{g}_L$ in (8.52) is easy to compute as the inverse of the matrix of inner products of the gradients of the coordinate functions in $\overline{M}_L$:

$$\nabla r = \frac{x_1}{r} X_1 + \frac{x_2}{r} X_2 \quad \text{and} \quad \nabla x_3 = -\frac{1}{2} x_2 X_1 + \frac{1}{2} x_1 X_2 + L^{-1/2} \widetilde{X_3}.$$

is an isometric embedding of $\overline{M}_L$ in $\mathbb{R}^3$. Via this embedding, it is straightforward to compute the curvature of γ:

$$k_\gamma(r, x_3) = \dot{\alpha} - r\varphi(r)^2 \sin\alpha.$$

The logarithmic normal derivative of the orbital length function is

$$\begin{aligned}\partial_{\nu(x)} \log \operatorname{Length} U(1)\cdot\pi(x) &= \left(-\sin\alpha\frac{\partial}{\partial r} + \frac{\cos\alpha}{\varphi(r)}\frac{\partial}{\partial x_3}\right)\log(\pi r\sqrt{4+Lr^2})\\ &= -\sin\alpha\left(\frac{1}{r} + r\varphi(r)^2\right)\end{aligned}$$

whence

$$H(x) = (\dot{\alpha} - r\varphi(r)^2\sin\alpha) + \sin\alpha\left(\frac{1}{r} + r\varphi(r)^2\right) = \dot{\alpha} + \frac{\sin\alpha}{r}. \tag{8.56}$$

(8.55) follows from (8.56) and the condition

$$1 = |\dot{\gamma}|^2 = \dot{r}^2 + \varphi(r)^2\dot{x_3}^2 = \cos^2\alpha + \varphi(r)^2\dot{x_3}^2. \qquad \square$$

Observe that

$$J = J(r,\alpha) = r\sin\alpha - \frac{1}{2}Hr^2$$

is an invariant of the system (8.55). For a solution curve $(r(t), x_3(t), \alpha(t))$ with $J(r(t), \alpha(t)) = C$, we have

$$\sin\alpha = \frac{C}{r} + \frac{Hr}{2}$$

and

$$\frac{dx_3}{dr} = \frac{\dot{x_3}}{\dot{r}} = \frac{\sqrt{4+Lr^2}}{2\sqrt{L}}\tan\alpha = \pm\frac{\sqrt{4+Lr^2}}{2\sqrt{L}}\frac{C/r + Hr/2}{\sqrt{1-(C/r+Hr/2)^2}}.$$

In particular, for $C = 0$ we find

$$\sin\alpha = \frac{Hr}{2}, \qquad \frac{dx_3}{dr} = \pm\frac{Hr}{2}\sqrt{\frac{4+Lr^2}{L(4-H^2r^2)}}.$$

Choosing the negative sign and integrating, we find

$$x_3(r) = -\int_r^{2/H}\frac{H\rho}{2}\sqrt{\frac{4+L\rho^2}{L(4-H^2\rho^2)}}\,d\rho.$$

To evaluate the integral we substitute $w = \arcsin(K\sqrt{4+L\rho^2})$, where $K = H/2\sqrt{L+H^2}$, to obtain

$$x_3(r) = \frac{H^2+L}{H^2L}\cdot\left(\arccos\left(\frac{H}{2}\sqrt{\frac{4+Lr^2}{H^2+L}}\right) + \frac{H\sqrt{L(4+Lr^2)(4-H^2r^2)}}{4(H^2+L)}\right) \tag{8.57}$$

for $0 \le r \le 2/H$.

Remark 8.33. In the limit as $L \to \infty$ we obtain

$$\begin{aligned} x_3(r) &= \frac{1}{H^2}\left(\arccos(\frac{Hr}{2}) + \frac{Hr\sqrt{4-H^2r^2}}{4}\right) \\ &= \frac{1}{4}\left(R^2\arccos(\frac{r}{R}) + r\sqrt{R^2-r^2}\right), \qquad R = \frac{2}{H}, \end{aligned}$$

which agrees with Pansu's conjectured isoperimetry extremal in the Heisenberg group with Carnot–Carathéodory metric.

We denote by S_{HL} the closed surface in $(\mathbb{R}^3, g_L)$ obtained as the lift of the curve $\gamma_{HL} \subset \overline{M}_L$ which is given by the graphs of $x_3 = \pm x_3(r)$ for $0 \le r \le R = 2/H$. Denote by V_{HL} the volume of the region bounded by S_{HL}, and by A_{HL} the surface area of S_{HL}. The parameterized curve

$$H \mapsto (A_{HL}, V_{HL})$$

yields an upper bound for the isoperimetric constant of $(\mathbb{R}^3, g_L)$ and, for sufficiently large H, coincides with that isoperimetric constant (by unpublished work of Kleiner). Direct computation gives the volume

$$\begin{aligned} V_{HL} &= 2\pi\int_0^{2/H}\int_0^{x_3(r)} r\sqrt{4+Lr^2}\left(\frac{2\sqrt{L}}{\sqrt{4+Lr^2}}\right) dt\,dr \\ &= 4\pi\sqrt{L}\int_0^{2/H} r x_3(r)\,dr \\ &= \frac{2\pi}{H^3L}\left(3L + H^2 + \frac{(3L-H^2)(L+H^2)\arctan(\sqrt{L}/H)}{H\sqrt{L}}\right) \end{aligned}$$

(use the substitution $v = K\sqrt{4+Lr^2}$) and the surface area

$$\begin{aligned} A_{HL} &= 2\pi\int_0^{2/H} r\sqrt{4+Lr^2}\left(\frac{2}{\sqrt{4-H^2r^2}}\right) dr \\ &= \frac{8\pi}{H^2}\left(1 + \frac{(L+H^2)\arctan(\sqrt{L}/H)}{H\sqrt{L}}\right). \end{aligned}$$

For fixed L, the series expansions for V_{HL} and A_{HL} in inverse powers of H take the form

$$\begin{aligned} V_{HL} &= \frac{32\pi}{3H^3} + \frac{64\pi L}{15H^5} + O(H^{-7}), \\ A_{HL} &= \frac{16\pi}{H^2} + \frac{16\pi L}{3H^4} + O(H^{-6}), \end{aligned}$$

while for fixed H, the series expansions for V_{HL} and A_{HL} in inverse powers of $L^{1/2}$ take the form

$$V_{HL} = \frac{3\pi^2\sqrt{L}}{H^4} + \frac{2\pi^2}{H^2\sqrt{L}} + O(L^{-3/2}),$$

$$A_{HL} = \frac{4\pi^2\sqrt{L}}{H^3} + \frac{4\pi^2}{H\sqrt{L}} + O(L^{-1}).$$

From these asymptotic expansions, the failure of this method is evident. As stated above, for fixed L and sufficiently large H, say, $H > H_0$ the quantities A_{HL} and V_{HL} yield the isoperimetric constant (and profile) of $(\mathbb{R}^3, g_L)$. However, H_0 depends on L and indeed $H_0 \to \infty$ as $L \to \infty$. Thus, in essence, two limits must be computed: one as $L \to \infty$ and the other as $H \to \infty$. To recover the conjectured isoperimetric constant and profile in the sub-Riemannian limit, one must compute the limits simultaneously with a specific interdependence given by the manner in which H_0 depends on L. Unfortunately, there is currently an insufficient understanding of this functional dependence to perform this computation.

8.7.2 Failure of the Brunn–Minkowski approach to isoperimetry in $\mathbb{H}$

In this section we discuss the classical proof of the isoperimetric inequality in $\mathbb{R}^n$ via convex geometry and the Brunn–Minkowski inequality, and describe work of Monti which demonstrates an essential obstruction to this approach in the Heisenberg group.

The *Brunn–Minkowski inequality* in $\mathbb{R}^n$ asserts that

$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n} \tag{8.58}$$

whenever A and B are nonempty measurable sets in $\mathbb{R}^n$. Here $A+B = \{a+b : a \in A, b \in B\}$ denotes the Minkowski sum of A and B. There are several approaches to prove inequality (8.58). Let us recall one of the most well-known arguments. By approximation, it suffices to prove (8.58) in the case when A and B are closed rectilinear parallelpipeds. In this case, if $A = \prod_j I_j$ and $B = \prod_j K_j$ are products of closed intervals, then $A + B = \prod_j (I_j + K_j)$ and (8.58) reduces to

$$\prod_j |I_j + K_j|^{1/n} = \prod_j (|I_j| + |K_j|)^{1/n} \geq \prod_j |I_j|^{1/n} + \prod_j |K_j|^{1/n},$$

which is an immediate consequence of the arithmetic-geometric mean inequality applied to the collections $u_j = |I_j|/(|I_j| + |K_j|)$ and $v_j = |K_j|/(|I_j| + |K_j|)$, $j = 1, \dots, n$.

The Brunn–Minkowski inequality leads to a quick solution to the isoperimetric problem for the Minkowski content

$$\mathcal{M}_{n-1}(E) = \lim_{\epsilon \to 0} |E_\epsilon \setminus E|/\epsilon$$

in $\mathbb{R}^n$. For any $E \subset \mathbb{R}^n$, $E_\epsilon = E + B(o,\epsilon)$ so

$$\begin{aligned} |E_\epsilon \setminus E| &= |E + B(o,\epsilon)| - |E| \\ &\geq (|E|^{1/n} + |B(o,\epsilon)|^{1/n})^n - |E| \\ &= (|E|^{1/n} + |B|^{1/n}\epsilon)^n - |E| \\ &= n|B|^{1/n}|E|^{1-1/n}\epsilon + O(\epsilon^2). \end{aligned} \tag{8.59}$$

Thus

$$\liminf_{\epsilon\to 0} \frac{|E_\epsilon \setminus E|}{\epsilon} \geq n|B|^{1/n}|E|^{1-1/n}$$

so

$$\frac{\mathcal{M}_{n-1}(\partial E)}{|E|^{(n-1)/n}} \geq n|B|^{1/n} = \frac{\mathcal{M}_{n-1}(\partial B)}{|B|^{(n-1)/n}}$$

whenever $\mathcal{M}_{n-1}(\partial E)$ exists.

Monti [205] showed that the corresponding Brunn–Minkowski inequality in the Heisenberg group,

$$|A \cdot B|^{1/4} \geq |A|^{1/4} + |B|^{1/4}, \tag{8.60}$$

$A \cdot B = \{ab : a \in A, b \in B\}$ fails to hold. His argument is indirect, using the fact that Carnot–Carathéodory balls are not solutions to the isoperimetric problem. Indeed, it can be shown by direct calculation that

$$\frac{P_{\mathbb{H}}(B)}{|B|^{3/4}} > \frac{P_{\mathbb{H}}(\mathcal{B})}{|\mathcal{B}|^{3/4}} \tag{8.61}$$

for any CC ball $B \subset \mathbb{H}$ and any bubble set $\mathcal{B}$. However, the validity of (8.60) would imply equality in (8.61), by exactly the same argument as given above in the Euclidean case. Under the assumption that (8.60) holds, one would obtain

$$\frac{\mathcal{M}_3(\partial E)}{|E|^{3/4}} \geq \frac{\mathcal{M}_3(\partial B)}{|B|^{3/4}} = \frac{P_{\mathbb{H}}(B)}{|B|^{3/4}} \tag{8.62}$$

for all bounded open sets $E \subset \mathbb{H}$. Equations (8.7) and (8.62), taken together, contradict (8.61).

8.7.3 Horizontal mean curvature flow

This section is of a far more speculative nature than the remainder of the chapter, and is intended merely to provide a rough sketch of another potential avenue towards Pansu's conjecture. The idea is to implement an analog of the Euclidean mean curvature flow discussed in Chapter 1 in the sub-Riemannian geometry of $\mathbb{H}$. More precisely, we seek to deform in a continuous fashion a bounded open set $\Omega \subset \mathbb{H}$ in a flow of bounded sets $\{\Omega_t\}$ so that the isoperimetric ratio

$$\frac{P_{\mathbb{H}}(\Omega_t)^{4/3}}{|\Omega_t|}$$

decreases with time. Among all possible flows which may accomplish this, we restrict our attention to flows which satisfy the following properties:

(i) the perimeter $P_{\mathbb{H}}(\Omega_t)$ is non-increasing with respect to time, and strictly decreasing unless $\partial\Omega_t$ satisfies some curvature conditions (e.g., has constant mean curvature), and

(ii) the volume $|\Omega_t|$ is constant in time.

More precisely, we consider the following analog of the mean curvature flow:

$$\frac{\partial F^N}{\partial t} = -\mathcal{H}_0 \nu_H^N \tag{8.63}$$

Here $F : \partial\Omega \times [0,T) \to \mathbb{H}$ is a family of immersions, $\mathcal{H}_0$ and ν_H are respectively the horizontal mean curvature and the horizontal normal defined in Sections 4.3 and 4.3.1, and for any vector $V \in \mathbb{R}^3$, we let V^N denote the projection of V on the Euclidean unit normal $\vec{n}$ to $\partial\Omega$. We remind the reader that the tangential component of the velocity is not relevant from a geometric point of view since it will give rise to a reparametrization of the surface and will not contribute to variation of the perimeter. The nonlinear evolution PDE system (8.63) describes a flow of the initial surface $\partial\Omega$ in which the normal velocity at any point is given by $-\mathcal{H}_0\nu_H^N$.

The flow (8.63) does not satisfy condition (ii) above. In order to impose a volume constraint we argue as in Section 8.3 and perturb equation (8.63) by subtracting the variation of the volume to obtain

$$\frac{\partial F^N}{\partial t} = \left(-\mathcal{H}_0 + \frac{\int_{\partial\Omega_t} \mathcal{H}_0 d\mu}{P_{\mathbb{H}}(\Omega_t)} \right) \nu_H^N. \tag{8.64}$$

It is easy to check that if $\{\Omega_t\}$ evolves according to (8.64), then $P_{\mathbb{H}}(\Omega_t)$ decreases unless $\mathcal{H}_0$ is constant and $|\Omega_t|$ is constant for all t. Hence (i) and (ii) are satisfied.

The Euclidean analogue of (8.64) has been studied by Gage [111] and Huisken [153]. One of the main results in these papers states that convex initial surfaces (or curves) remain convex and converge asymptotically to spheres (or circles). A sub-Riemannian version of these results would not provide new information regarding Pansu's conjecture (as the convex and C^2 cases have been already established), but would provide an analytic approach to the non-smooth, non-convex setting via appropriate notions of weak solutions.

At the moment there are essentially no results on existence, regularity and asymptotic behavior of solutions of (8.64). This is primarily due to the lack of an adequate toolkit of differential geometry for hypersurfaces in the sub-Riemannian context.

The characteristic locus introduces an additional complication into the analysis of these flows, not present in the Euclidean case. Both systems (8.63) and

(8.64) are not well defined at characteristic points, and one needs to find a suitable interpretation of the velocity $\mathcal{H}_0\nu_H^N$ along $\Sigma(\partial\Omega_t)$. To conclude this section, we propose a possible approach to this latter problem. For simplicity we assume that Ω_t is a sublevel set of a C^1 function $u(\cdot,t):\mathbb{H}\to\mathbb{R}$. We use the Riemannian approximants $(\mathbb{R}^3, g_L)$ to approximate solutions of (8.63) or (8.64) with solutions of the corresponding Riemannian flows. By the results in Section 4.3.1, the Riemannian velocity field can be rewritten as

$$(\vec{\mathcal{H}}_L)^N = \mathcal{H}_L|\nabla_L u|,$$

where $\vec{\mathcal{H}}_L$ denotes the Riemannian mean curvature vector and $\mathcal{H}_L$ its length. From (4.21) we recall that

$$\lim_{L\to\infty} \mathcal{H}_L|\nabla_L u| = \mathcal{L}u \tag{8.65}$$

at characteristic points, where $\mathcal{L}u = X_1^2u + X_2^2u$ is the sub-Laplacian operator in $\mathbb{H}$. Consequently, in order to define smooth solutions one could use as normal velocity the quantity $\mathcal{H}_0\nu_H^N$ at noncharacteristic points, and the quantity $\mathcal{L}u$ at characteristic points.

8.8 Further results

8.8.1 The isoperimetric problem in the Grushin plane

The Grushin plane is closely connected to the first Heisenberg group, and in this setting the isoperimetric problem has been recently settled by R. Monti and D. Morbidelli [207]. In this section we briefly recount their results and sketch the main ideas of the proof.

The Grushin plane G is the space $\mathbb{R}^2$ equipped with the sub-Riemannian structure arising from the vector fields $\overline{X}_1 = \partial_{x_1}$ and $\overline{X}_2 = x_1\partial_{x_2}$, where we have denoted points of G by $x = (x_1, x_2)$. For the perimeter measure in G we use the set function

$$P_G(E) = \sup \int_E (\overline{X}_1\varphi_1 + \overline{X}_2\varphi_2)\, dx_1\, dx_2,$$

where the supremum is taken over all compactly supported functions

$$\varphi = (\varphi_1, \varphi_2) \in C^1(\mathbb{R}^2 : \mathbb{R}^2)$$

satisfying $\varphi_1^2 + \varphi_2^2 \le 1$.

Theorem 8.34 (Monti–Morbidelli). *For any measurable set $E \subset G$,*

$$|E| \le \frac{1}{3} P_G(E)^{3/2}. \tag{8.66}$$

Moreover, equality holds in (8.66) *only for the sets* $E_1(h,R)$ *consisting of all points* $(x_1,x_2) \in G$ *for which*

$$|x_2 + h| < \frac{1}{2}(R^2 \arccos\left(\frac{|x_1|}{R}\right) + |x_1|\sqrt{R^2 - x_1^2})$$

and $|x_1| < R$, *where* $h \in \mathbb{R}$ *and* $R > 0$.

Observe the connection between the conjectured isoperimetric profile of $\mathbb{H}$ and the isoperimetric profile of G given in Theorem 8.34: the bubble set $\mathcal{B}(o,R)$ in $\mathbb{H}$ is obtained by rotating the isoperimetry extremal $E_1(0,R)$ in G (identified as the x_1x_3-plane in $\mathbb{H}$) about the x_3-axis.

We sketch the proof of Theorem 8.34 in a sequence of steps.

Step 1. Under the nonlinear change of variables

$$(x_1,x_2) \overset{\Psi}{\mapsto} \left(\frac{1}{2}x_1|x_1|, x_2\right),$$

the horizontal perimeter $P_G(E)$ of a set E in G coincides with the ordinary perimeter $P(\Psi(E))$ of its image $\Psi(E)$ in $\mathbb{R}^2$, and the Lebesgue measure $|E|$ of E coincides with the weighted measure $\mu(\Psi(E))$ of its image, where $d\mu(x) = |2x_1|^{-1/2}\,dx$.

Step 2. By computing the effect of suitable symmetrization procedures on the perimeter and the weighted area using the transformation and measure μ from Step 1, it can be shown that any isoperimetry extremal in G is necessarily congruent with an extremal which is symmetric and convex with respect to each of the coordinate axes. Consequently, the boundary of such an extremal is locally Lipschitz.

Step 3. A compactness argument, together with the lower semi-continuity of the functional P_G, guarantees the existence of minimizers for the functional

$$E \mapsto \frac{P_G(E)^{3/2}}{|E|}.$$

(Compare Theorem 8.4.)

Step 4. A variational argument determines the precise character of the minimizers. Let U be a neighborhood of the point $(0,b) \in \partial E$, $b > 0$, and write $\partial E \cap U = \{(x_1,x_2) : x_2 = \varphi(x_1)\}$ for a suitable Lipschitz function φ on an interval $I = (-\delta,\delta)$, $\delta > 0$. The resulting Euler–Lagrange equation is

$$\frac{3}{2}|E|\frac{\varphi'(x_1)}{\sqrt{\varphi'(x_1)^2 + x_1^2}} + P_G(E)x_1 = c \tag{8.67}$$

for a.e. x_1, $|x_1| < \delta$, where $c \in \mathbb{R}$ is a constant. Since E is symmetric with respect to x_1, φ must be even, hence φ' is odd and $c = 0$. Solving (8.67) for $\varphi'(x_1)$ gives

$$\varphi'(x_1) = -\frac{\lambda x_1 |x_1|}{\sqrt{1 - \lambda^2 x_1^2}}$$

for a.e. x_1, $|x_1| < \delta$, where $\lambda = \frac{3}{2} P_G(E)/|E|$. Then in fact φ' is continuous and an integration gives

$$\begin{aligned}\varphi(x_1) &= \int_{x_1}^{R} \frac{t^2}{R\sqrt{1 - t^2/R^2}}\, dt = R^2 \int_{\arcsin |x_1/R|}^{\pi/2} \sin^2(t)\, dt \\ &= \frac{1}{2}(R^2 \arccos(|x_1|/R) + |x_1|\sqrt{R^2 - x_1^2})\end{aligned}$$

as asserted. Once the extremal domain has been identified, the derivation of the constant in (8.66) is a simple computation.

8.8.2 The classification of symmetric CMC surfaces in $\mathbb{H}^n$

In this section, we present Ritoré and Rosales' [232] classification of all cylindrically symmetric constant mean curvature surfaces in the higher-dimensional Heisenberg groups. Assume that S is a C^2 hypersurface in $\mathbb{H}^n$ which is invariant under the group of rotations in $\mathbb{R}^{2n+1}$ about the x_{2n+1}-axis. Any such surface can be generated by rotating a curve in the $\{x_1 \geq 0\}$ half-plane of the $x_1 x_{2n+1}$-plane, $g(t) = (x(t), f(t))$, where t varies over an interval I in the x_1-axis, around the x_{2n+1}-axis. We may parametrically realize such a surface as follows. Letting $B = I \times \mathbb{S}^{n-1}$, the map

$$\phi(t, \omega) = (x(t)\omega, f(t))$$

parameterizes the rotationally invariant surface formed by rotating the curve g about the x_{2n+1}-axis. Computing the unit Riemannian normal yields

$$\frac{((x(t)x'(t)\omega_{n+k} - f'(t)\omega_k)\, X_k + (-x(t)x'(t)\omega_k - f'(t)\omega_{n+k})\, X_{k+1} + x'(t)\, X_{2n+1})}{\sqrt{|g'(t)|^2 + x(t)^2 x'(t)^2}}.$$

From this, one can compute the horizontal mean curvature of such a surface, yielding,

$$H = \frac{1}{2n} \frac{x^3(x'f'' - x''f') + (2n-1)(f')^3 + 2(n-1)x^2(x')^2 f'}{x(x^2(x')^2 + (f')^2)^{\frac{3}{2}}}$$

Denoting by $\sigma(t)$ the angle between g' and the vertical direction, $\frac{\partial}{\partial x_{2n+1}}$, the formula for the mean curvature yields that if the surface of rotation is of constant mean curvature, then $g(t)$ satisfies the following system of ordinary differential equations:

$$\begin{aligned} x' &= \sin(\sigma), \\ f' &= \cos(\sigma), \\ \sigma' &= (2n-1)\frac{\cos^3(\sigma)}{x^3} + 2(n-1)\frac{\sin^2(\sigma)\cos(\sigma)}{x} \\ &\quad - 2nH\frac{(x^2\sin^2(\sigma)+\cos^2(\sigma))^{\frac{3}{2}}}{x^2}, \end{aligned} \tag{8.68}$$

whenever $x > 0$. Using Noether's theorem, the authors compute the first integral of this system, showing that

$$E = \frac{x^{2n-1}\cos(\sigma)}{\sqrt{x^2\sin^2(\sigma)+\cos^2(\sigma)}} - Hx^{2n}$$

is constant along any solution to the system. E is called the energy of the system. Using certain geometric properties of the solutions, Ritoré and Rosales then classify all cylindrically symmetric constant mean curvature surfaces.

Theorem 8.35. *Let $g(s)$ be a complete solution to* (8.68) *with energy E. Then the surface, $S \subset \mathbb{H}^n$, generated by rotation about the x_{2n+1}-axis, is one of the five following types:*

1. *If $H = 0$ and $E = 0$, then $g(s)$ is a straight line orthogonal to the x_{2n+1}-axis and S is a Euclidean hyperplane.*
2. *If $H = 0$ and $E \neq 0$, then S is an embedded surface of catenoidal type.*
3. *If $H \neq 0$ and $E = 0$, then S is a compact hypersurface homeomorphic to the sphere.*[3]
4. *If $EH > 0$, then $g(s)$ is a periodic graph over the x_{2n+1}-axis. S is a cylinder or an embedded hypersurface of unduloid type.*
5. *If $EH < 0$, then $g(s)$ is a locally convex curve and S is a nodoid type hypersurface with self-intersections.*

8.9 Notes

In the last decade, there has been an explosion of research on analogs of the minimal and constant mean curvature equations and associated variational problems in the setting of Carnot–Carathéodory spaces. While some of the work most closely related to the isoperimetric problem is covered in this chapter, we point out that there is a wealth of other material that is beyond the scope of this discussion. For the study of minimal surfaces in the Heisenberg groups, see [28, 80, 82, 83, 117, 221, 223, 231, 232]. For the roto-translation group, see [68, 146]. For three-dimensional

[3]For $n = 1$, these are precisely the bubble sets $\mathcal{B}(o, R)$.

pseudo-hermitian manifolds (which include both the Heisenberg groups and the roto-translation group), see [63, 64]. For general Carnot groups, see [56, 78]. For general sub-Riemannian spaces, see [116, 144, 147].

Notes for Section 8.1**.** Pansu's conjecture was first posed in [217] and [219]. The observation regarding the equivalence with the isoperimetric problem for Minkowski content is due to Monti and Serra-Cassano [211].

Notes for Section 8.2**.** Theorem 8.3 is due to Leonardi and Rigot [176], who established existence results in the general class of Carnot groups. Their proof is based on Garofalo and Nhieu's Theorem 8.4 and on several results established in [116]. In that paper the setting is Carnot–Carathéodory metrics generated by special systems of Lipschitz vector fields. We have presented a simplified proof valid in the Heisenberg group. Section 8.2 contains the *concentration-compactness* argument of Leonardi and Rigot (Lemma 8.5) as well as their ingenious method for demonstrating the (essential) boundedness of the isoperimetric sets. The relation between perimeter and rate of change of the volume used in (8.11) and (8.12) was proved by Ambrosio in [7, Lemma 3.5].

We note that Leonardi and Rigot also investigate some properties of isoperimetric sets Ω, showing that such sets are Ahlfors regular and satisfy a synthetic regularity condition known as Condition B. Moreover, in the setting of the Heisenberg group, such sets are also *domains of isoperimetry*, that is, a relative isoperimetric inequality of the form

$$\min\{|S|^{3/4}, |\Omega \setminus S|^{3/4}\} \leq C P_{\mathbb{H}}(S, \Omega)$$

holds for all sets $S \subset \Omega$ and a suitable constant $C < \infty$. As a consequence, isoperimetric sets are connected. As discussion of these facts would take us away from the main points of this survey, we refer the interested reader to the original paper [176].

Notes for Section 8.3**.** The results in this section were independently proved by many authors. Our presentation loosely follows the one in [231].

If, in addition to the hypotheses in Proposition 8.6, we also assume that $\mathcal{H}_0 \in L^1(S, d\sigma)$ with respect to the surface measure, then we rule out the possibility that $\mathcal{H}_0$ is a distribution with mass supported on $\Sigma(S)$. In this case, we may easily deduce that

$$\int_S u(\operatorname{div}_S \nu_H)\, d\sigma = \int_{S\setminus\Sigma(S)} u(\operatorname{div}_S \nu_H)\, d\sigma = \mathcal{H}_0 \int_S u\, d\sigma = 0 \tag{8.69}$$

for all volume-preserving C^1 vector fields U with compact support on S, where $u = \langle U, \nu_1 \rangle_1$.

Notes for Subsection 8.3.1**.** The derivation in this section is an original contribution of this survey and represents a generalization of techniques in [117], where the minimal surface case was considered. Formulas (8.18), (8.21) and (8.24) can be found in [117]. Lemma 8.7 and Proposition 8.9 are proved in [231], where the corollary is pointed out as well. Both [37], and [231, 232] contain (different) proofs of Theorem 8.6.

Notes for Section 8.4**.** Theorems 8.11 and 8.36 are proved by Danielli, Garofalo and Nhieu in [81]. These results continue to hold, appropriately reformulated, in any Heisenberg group $\mathbb{H}^n$.

One immediate consequence of Theorem 8.11 is the following isoperimetric inequality.

Theorem 8.36. *Let* $\mathcal{E}$ *be as in Section* 8.4*, and denote by* $\tilde{\mathcal{E}}$ *the class of sets of the form* $L_y\delta_\lambda(E)$ *for some* $E \in \mathcal{E}$*,* $\lambda > 0$ *and* $y \in \mathbb{H}$*. Then*

$$|E|^{3/4} \leq C_{\text{iso}}(\mathbb{H}) P_{\mathbb{H}}(E) \tag{8.70}$$

for all $E \in \tilde{\mathcal{E}}$*, where* $C_{\text{iso}}(\mathbb{H}) = 3^{3/4}/(4\sqrt{\pi})$*, with equality if and only if* $E = L_y\mathcal{B}(o, R)$ *for some* $R > 0$ *and* $y \in \mathbb{H}$*.*

In the interesting work [177], Leonardi and Masnou show, among other things, that such u_o is a critical point (but not the unique minimizer) of the horizontal perimeter, when the class of competitors is restricted to C^2 domains with defining function $x_3 = \pm f(|z|)$. The same result has been also noted in [232]. We also want to point out related results in a recent preprint by Ritoré [230], which considers the analog of the bubble sets in higher-dimensional Heisenberg groups and proves a sharp isoperimetric inequality yielding the isoperimetric profile of $\mathbb{H}^n$ within the class of C^1 sets contained in a cylinder with axis along the center of the group.

Theorem 4.48, which plays a role in the derivation in this section, is a result of Balogh, see Theorem 3.1 in [20]. It is worthwhile noting that the result of Theorem 4.48 fails if $C^{1,1}_{\text{loc}}$ is replaced by $C^{1,\alpha}_{\text{loc}}$ for any $\alpha < 1$; examples to this effect are also given in [20].

Unfortunately, effective symmetrization procedures in the Heisenberg group (and other Carnot groups) are noticeably lacking. An approach to symmetrization via polarization has been developed in the classical space forms, see Baernstein [18] or Brock–Solynin [44]. Simply put, this program seeks to realize certain well-studied symmetrization procedures (such as Steiner symmetrization) as limits of sequences of polarizations, i.e., reflective symmetrizations in hyperplanes. Preliminary attempts to generalize this program to the Heisenberg setting encounter significant obstructions; polarizations in vertical hyperplanes (the obvious candidates for producing cylindrical symmetry) are not well behaved. Ultimately, this stems from the fact that reflections in such planes are not isometries of the CC metric. (Compare the discussion following the definition of reflection in (8.29).)

Notes for Sections 8.5 **and** 8.6**.** The brief sketches of Theorems 8.23 and 8.29 are based on the more complete arguments given by Ritoré and Rosales [231] and Monti and Rickly [210], respectively. In comparing the proof of Theorem 8.23

with the proof of [231, Theorem 4.16] the reader will notice a slight difference. Our argument incorporates ideas from the proof of Theorem 5.3 in that paper; this choice is motivated by the need for consistency throughout this monograph. We would also like to note that the results in [231] are more general than those presented here. Proposition 8.24 is also proved in [64, Theorem 6.3] for t-graphs.

Notes for Section 8.7**.** The case $L = 1$ of the argument in Subsection 8.7.1 is due to Tomter [247]. Our presentation is based in part on unpublished notes on Tomter's work by Manfredi and Gong [127]. Subsection 8.7.2 summarizes the work by Monti in [205]. The classical Brunn–Minkowski inequality is a staple of convex geometry and can be found in many textbooks, see for example Ball [19] or Burago–Zalgaller [48].

Subsection 8.7.3 is based on some preliminary results by Bonk and Capogna [37] and Capogna and Citti [51]. In particular, regarding the PDE (8.63), comparison principles for (8.63) have been established in [37] for classical solutions and in [51] for weak viscosity solutions. (See also [34] for other comparison theorems for viscosity solutions of nonlinear parabolic equations in $\mathbb{H}^n$.). Coupled with the existence of explicit self-similar solutions, used as barriers, these comparison principles imply that any bounded surface evolving according to (8.63) will shrink to a point in a finite time. Citti has observed that (8.65) allows a way to define smooth flows even at characteristic points. Alternatively, one can define a notion of weak solutions for (8.63) using the concept of viscosity solutions, following an idea of Evans–Spruck [90], [91], [92], [93] and Chen–Giga–Goto [62].

In the Riemannian case, both for (8.63) and (8.64) weak solutions can be defined using Brakke's method of currents, see [42]. In the sub-Riemannian setting the geometric measure theoretic machinery needed to set up and study this kind of solutions has yet to be fully developed.

Notes for Subsection 8.8.1**.** Theorem 8.34 is a special case of a theorem of Monti and Morbidelli [207], who consider a one-parameter family of Grushin-type spaces. Fixing $\alpha \geq 0$, let G_α be the sub-Riemannian structure arising from the vector fields $\overline{X}_{1,\alpha} = \partial_{x_1}$ and $\overline{X}_{2,\alpha} = x_1|x_1|^{\alpha-1}\partial_{x_2}$ on $\mathbb{R}^2$. Let

$$P_{G_\alpha}(E) = \sup \int_E (\overline{X}_{1,\alpha}\varphi_1 + \overline{X}_{2,\alpha}\varphi_2)\, dx_1\, dx_2$$

be the corresponding perimeter measure, where once again the supremum is taken over all pairs of compactly supported C^1 functions φ_1, φ_2 on $\mathbb{R}^2$ satisfying $\sup_{\mathbb{R}^2}(\varphi_1^2 + \varphi_2^2) \leq 1$. Theorem 1.1 of [207] asserts that

$$|E| \leq C_{\mathrm{iso}}(G_\alpha) P_{G_\alpha}(E)^{(2+\alpha)/(1+\alpha)} \tag{8.71}$$

for all measurable sets $E \subset \mathbb{R}^2$ with finite measure, where

$$C_{\mathrm{iso}}(G_\alpha) = \frac{1+\alpha}{2+\alpha}\left(2\int_0^\pi \sin^\alpha t\, dt\right)^{-1/(1+\alpha)}.$$

The case of equality can also be characterized; see (1.3) in [207].

Further results relating isoperimetric-type problems in G_1 and $\mathbb{H}$ can be found in [12], while the isoperimetric problem in the sub-Riemannian S^3 sphere has been addressed in [154].

Notes for Subsection 8.8.2**.** The results of this section are taken from the manuscript of Ritoré and Rosales [232]. Ni [214] has given another derivation of the cylindrically symmetric constant mean curvature surfaces in the Heisenberg groups by using the Riemannian approximation scheme.

Chapter 9

Best Constants for Other Geometric Inequalities on the Heisenberg Group

As the point of departure for this final chapter, we return to the equivalence of the isoperimetric inequality with the geometric (L^1-) Sobolev inequality. As shown in Section 7.1, the best constant for the isoperimetric inequality agrees with the best constant for the geometric (L^1-) Sobolev inequality. Recall that in the context of the Heisenberg group, the L^p-Sobolev inequalities take the form

$$||u||_{4p/(4-p)} \le C_p(\mathbb{H})||\nabla_0 u||_p, \qquad u \in C_0^\infty(\mathbb{H}). \tag{9.1}$$

In this chapter we discuss sharp constants for other analytic/geometric inequalities in the Heisenberg group and the Grushin plane. These include the L^p-Sobolev inequality (9.1) in the case $p = 2$, the Trudinger inequality (9.14), which serves as a natural substitute for (9.1) in the limiting case $p = 4$, and the Hardy inequality (9.24), a weighted inequality of Sobolev type on the domain $\mathbb{H} \setminus \{o\}$.

9.1 L^2-Sobolev embedding theorem

In this section, we present a proof of the following theorem of Jerison and Lee [156].

Theorem 9.1 (Jerison–Lee). *The inequality*

$$||u||_4 \le \pi^{-1/2}||\nabla_0 u||_2 \tag{9.2}$$

holds for all $u \in C_0^\infty(\mathbb{H})$, *with equality if* u *is a translate or multiple of* $u_0(z, x_3) = \frac{2}{\sqrt{\pi}}((1 + |z|^2)^2 + 16x_3^2)^{-1/2}$.

For simplicity, we present here a partial proof of Theorem 9.1, valid in the restricted class of *first-layer radially symmetric functions*. By this terminology we mean those functions $u : \mathbb{H} \to \mathbb{R}$ which satisfy $u(z, x_3) = U(|z|, x_3)$ for some $U : [0, \infty) \times \mathbb{R} \to \mathbb{R}$.

Denote by

$$\Lambda = \inf \left\{ \int_{\mathbb{H}} |\nabla_0 u|^2 : \begin{array}{c} u \in C_0^\infty(\mathbb{H}) \text{ is first-layer radially symmetric} \\ \int_{\mathbb{H}} u^4 = 1 \end{array} \right\} \tag{9.3}$$

the best constant in the dual formulation of the Sobolev inequality (9.2). The Euler–Lagrange equation associated to this variational problem is

$$\mathcal{L}u = -\Lambda u^3. \tag{9.4}$$

We note that if u is an entire solution to (9.4), then the translates $u \circ l_p$, $p \in \mathbb{H}$ and dilates $\lambda u \circ \delta_s$, $\lambda > 0$, are also solutions to (9.4). To prove Theorem 9.1 for first-layer radially symmetric functions, it thus suffices to establish the following proposition.

Proposition 9.2. $\Lambda = \pi$ *and the only nontrivial positive entire first-layer radially symmetric solutions are vertical translates and dilates of the extremal function* u_0 *in Theorem* 9.1*.*

Let $u : \mathbb{H} \to (0, \infty)$ be an entire solution to (9.4). We begin by rewriting (9.4) in terms of $\Phi = F(u) = (4\Lambda)^{-1} u^{-2}$. From the identity $\mathcal{L}\Phi = F''(u)|\nabla_0 u|^2 + F'(u)\mathcal{L}u$ we see that (9.4) takes the form

$$\mathcal{L}\Phi = \frac{3}{2}\frac{|\nabla_0 \Phi|^2}{\Phi} + \frac{1}{2}.$$

We will classify the first-layer radially symmetric solutions to this equation. Thus let

$$\Phi(z, x_3) = \phi(|z|, x_3)$$

for some function $\phi : [0, \infty) \times \mathbb{R} \to (0, \infty)$. Setting $r = |z|$ and $t = x_3$ we calculate

$$|\nabla_0 \Phi|^2 = \phi_r^2 + \frac{1}{4} r^2 \phi_t^2, \qquad \mathcal{L}\Phi = \phi_{rr} + \frac{1}{r}\phi_r + \frac{1}{4} r^2 \phi_{tt}.$$

Thus ϕ must solve the PDE

$$\phi_{rr} + \frac{1}{r}\phi_r + \frac{1}{4} r^2 \phi_{tt} = \frac{3}{2}\frac{\phi_r^2 + \frac{1}{4} r^2 \phi_t^2}{\phi} + \frac{1}{2}.$$

The substitutions $x = t$ and $y = \frac{1}{4} r^2$ result in the PDE

$$\triangle \phi = \frac{3}{2}\frac{|\nabla \phi|^2}{\phi} - \frac{\phi_y}{y} + \frac{1}{2y}, \tag{9.5}$$

for which we seek solutions $\phi(x, y)$ in the Poincaré half-plane $\Omega = \{(x, y) : y > 0\}$.

Following the method in [119], we employ Weinberger's technique of P-functions to solve (9.5). To simplify the notation in what follows, we introduce the auxiliary function $\psi = \phi - \frac{1}{2}y$ and denote its gradient by $\nabla\psi = (v, w) = (\phi_x, \phi_y - \frac{1}{2})$. Let $P = |\nabla\phi|^2 - w = |\nabla\psi|^2 + \frac{1}{4}$ and $A = \frac{|\nabla\phi|^2}{\phi}$. In this notation, (9.5) reads

$$\triangle\psi = \frac{3}{2}A - \frac{w}{y}. \tag{9.6}$$

Observe that

$$\triangle P = 2(||\mathrm{Hess}\psi||_2^2 + \langle\nabla\psi, \nabla\triangle\psi\rangle), \tag{9.7}$$

where $||\mathrm{Hess}\psi||_2$ denotes the L^2 norm of the Hessian of ψ. We next introduce the function $h = y/\phi^2$ and the vector field

$$V = (F, G) = \nabla P - A\nabla\psi.$$

Our goal is to prove the following divergence formula:

Proposition 9.3. $\mathrm{div}(hV) = h[2||\mathrm{Hess}\psi||_2^2 - (\triangle\psi)^2 + 3(\triangle\psi - A)^2]$.

Note that the term $2||\mathrm{Hess}\phi||_2^2 - (\triangle\phi)^2$ in the preceding formula is nonnegative; thus the right-hand side is the sum of two nonnegative terms.

To prove Proposition 9.3, we will break up the computation in a series of lemmas.

Lemma 9.4. $\langle V, \nabla\phi\rangle = \phi\langle\nabla A, \nabla\psi\rangle$.

This is an easy consequence of the identity $P + w = \phi A$.

Lemma 9.5. $\langle V, \nabla h\rangle = \frac{G}{\phi^2} - 2h\langle\nabla A, \nabla\psi\rangle$.

This follows from Lemma 9.4 and the definition of h.

Lemma 9.6. $2\langle\nabla\psi, \nabla\triangle\psi\rangle = 3\langle\nabla A, \nabla\psi\rangle - \frac{P_y}{y} + 2\frac{w^2}{y^2}$.

This follows from the PDE (9.6) and a short computation.

The verification of the divergence identity in Proposition 9.3 is now an easy calculation:

$$\begin{aligned}
\mathrm{div}(hV) &= h\triangle P - hA\triangle\psi - h\langle\nabla A, \nabla\psi\rangle + \langle\nabla h, V\rangle \\
&= 2h||\mathrm{Hess}\psi||_2^2 - h\frac{P_y}{y} + 2h\frac{w^2}{y^2} + \frac{G}{\phi^2} - hA\triangle\psi \\
&= h\left(2||\mathrm{Hess}\psi||_2^2 - A\frac{w}{y} + 2\frac{w^2}{y^2} - A\triangle\psi\right) \\
&= h\left(2||\mathrm{Hess}\psi||_2^2 - (\triangle\psi)^2 + 3(\triangle\psi - A)^2\right).
\end{aligned}$$

In the second line here we used Lemmas 9.5 and 9.6 and (9.7), while in the third line we used the definitions $G = P_y - A\psi_y = P_y - Aw$ and $h = y/\phi^2$. The final line follows from another application of (9.6).

Integrating this divergence formula over $\Omega_R = \Omega \cap B(o,R)$ gives

$$\int_{\Omega \cap \partial B(o,R)} h(F\,dx + G\,dy) = \int_{\Omega_R} h[2||\mathrm{Hess}\psi||_2^2 - (\triangle\psi)^2 + 3(\triangle\psi - A)^2]\,dxdy$$

where we have used the fact that $h = 0$ on $\partial\Omega$. The (easy) growth estimates $\psi(x,y) \approx |(x,y)|^2$, $|\nabla\psi(x,y)| = O(|(x,y)|)$, and $||\mathrm{Hess}\psi||_2 = O(1)$ imply that

$$\lim_{R\to\infty} \left| \int_{\Omega \cap \partial B(o,R)} h(F\,dx + G\,dy) \right| = 0.$$

Thus

$$\int_{\Omega} h[2||\mathrm{Hess}\psi||_2^2 - (\triangle\psi)^2 + 3(\triangle\psi - A)^2]\,dxdy = 0$$

i.e.,

$$2||\mathrm{Hess}\psi||^2 = (\triangle\psi)^2, \quad \text{and} \quad \triangle\psi = \triangle\phi = A = \frac{|\nabla\phi|^2}{\phi}.$$

The first identity yields $\psi_{xy} = 0$ and $\psi_{xx} = \psi_{yy}$, whence

$$\phi(x,y) = \psi(x,y) + \frac{1}{2}y = c^2(x^2 + y^2) + d_1 x + d_2 y + f.$$

Then the second identity yields $4c^2 f = d_1^2 + d_2^2$, so

$$\phi(x,y) = \left(cx + \frac{d_1}{2c}\right)^2 + \left(cy + \frac{d_2}{2c}\right)^2.$$

However, since ϕ must solve (9.5), we see that $d_2 = \frac{1}{2}$. Substituting $x = t = x_3$ and $y = \frac{r^2}{4} = \frac{|z|^2}{4}$ gives

$$\Phi(z,x_3) = \left(cx_3 + \frac{d_1}{2c}\right)^2 + \frac{1}{16}\left(c|z|^2 + \frac{1}{c}\right)^2.$$

After a suitable vertical translation and dilation, we obtain

$$\Phi_0(z,x_3) = \frac{1}{16}\left((1 + |z|^2)^2 + 16x_3^2\right)$$

and

$$u_0 = (4\Lambda\Phi_0)^{-1/2} = \frac{2}{\sqrt{\Lambda}}\left((1 + |z|^2)^2 + 16x_3^2\right)^{-1/2}.$$

The value $\Lambda = \pi$ is easily evaluated from the normalization condition $||u_0||_4 = 1$.

9.2 Moser–Trudinger inequality

The borderline case in the L^p-Sobolev inequality

$$||f||_{L^{np/(n-p)}(\mathbb{R}^n)} \leq C_p(\mathbb{R}^n)||\nabla f||_{L^p(\mathbb{R}^n)} \tag{9.8}$$

occurs when $p = n$ is the dimension of the ambient space. Passing to the limit as $p \to n$ in (9.8) suggests that the L^n norm of the gradient of f should control the L^∞ norm of f. However, this conjecture is false: there exist unbounded functions in dimensions $n \geq 2$ with (locally) n-integrable gradient. (For a simple example, consider $f(x) = \log\log|x|$.) A natural substitute for the Sobolev inequality in this borderline case is *Trudinger's inequality*, which asserts the local exponential integrability of $W^{1,n}$ functions. More precisely, there exist constants $\beta \geq 1$, $A > 0$ and $C_0 < \infty$ depending only on n so that

$$\frac{1}{|\Omega|}\int_\Omega \exp\left(A\left(\frac{|f|}{||\nabla f||_{n,\Omega}}\right)^\beta\right) \leq C_0 \tag{9.9}$$

for all domains $\Omega \subset \mathbb{R}^n$ of finite measure and all nonconstant functions $f \in C_0^\infty(\Omega)$. Note that the lack of good scaling properties for the Orlicz functional $\Phi(t) = \exp(At^\beta)$ which occurs in (9.9) precludes the possibility of writing (9.9) as an inequality between norms as in (9.8).

Let us give an elementary argument showing the validity of (9.9) for $\beta = 1$ with some positive constant A. First, expand the exponential in a power series:

$$\exp\left(A\frac{|f(x)|}{||\nabla f||_{n,\Omega}}\right) = \sum_{m=0}^\infty \frac{1}{m!}A^m\frac{|f(x)|^m}{||\nabla f||^m_{L^n(\Omega)}} \tag{9.10}$$

for $x \in \Omega$. For each $m \in \mathbb{N}$, define an exponent p_m so that

$$m = \frac{np_m}{n-p_m}. \tag{9.11}$$

We make use of the simple estimate

$$C_p(\mathbb{R}^n) \leq \frac{(n-1)p}{n-p}C_1(\mathbb{R}^n) \tag{9.12}$$

(compare 5.21). Integrating (9.10) over Ω and using (9.11) gives

$$\frac{1}{|\Omega|}\int_\Omega \exp\left(A\frac{|f|}{||\nabla f||_{n,\Omega}}\right) \leq \sum_{m=0}^\infty \frac{1}{m!}A^m\frac{\int_\Omega |f|^{np_m/(n-p_m)}}{|\Omega|\,||\nabla f||^{np_m/(n-p_m)}_{L^n(\Omega)}}.$$

Extending f by zero in $\mathbb{R}^n \setminus \Omega$ and applying the Sobolev inequality (9.8) together with (9.11) and the estimate in (9.12), we obtain

$$\begin{aligned}
\frac{1}{|\Omega|}\int_\Omega \exp\left(A\frac{|f|}{||\nabla f||_{n,\Omega}}\right) &\le \sum_{m=0}^{\infty}\frac{1}{m!}A^m\frac{1}{|\Omega|}\frac{\int_\Omega |f|^{np_m/(n-p_m)}}{||\nabla f||_{n,\Omega}^{np_m/(n-p_m)}} \\
&\le \sum_{m=0}^{\infty}\frac{1}{m!}(AC_{p_m}(\mathbb{R}^n))^m \\
&\le \sum_{m=0}^{\infty}\frac{1}{m!}(A\frac{(n-1)p_m}{n-p_m}C_1(\mathbb{R}^n))^m \\
&= \sum_{m=0}^{\infty}\frac{1}{m!}(A\frac{n-1}{n}mC_1(\mathbb{R}^n))^m.
\end{aligned}$$

By Stirling's formula, the summand is approximately

$$\frac{(Ae\frac{n-1}{n}C_1(\mathbb{R}^n))^m}{\sqrt{2\pi m}}$$

for large m, which shows that the series converges whenever $A < (e\frac{n-1}{n}C_1(\mathbb{R}^n))^{-1}$.

With some more work the estimate (9.12) can be improved to

$$C_p(\mathbb{R}^n) \le \frac{c(n)}{(n-p)^{1-1/p}}$$

which allows the preceding argument to be carried out with $\beta = n/(n-1)$. (In fact, the best constant in the L^p-Sobolev inequality on $\mathbb{R}^n$ for $1 \le p < n$ has been computed by Talenti; see the Notes to this chapter for more details.) By using suitable truncations of the logarithmic potential $f(x) = \log 1/|x|$ in $\Omega = B(0,1)$, it is easy to see that β cannot be chosen greater than $n/(n-1)$.

Henceforth, we refer to inequality (9.9) with $\beta = n/(n-1)$ as *Trudinger's inequality*. With the value of the optimal exponent β resolved, attention naturally shifts to the coefficient A. By the best coefficient we mean the supremum $A(\mathbb{R}^n)$ of those values A so that (9.9) holds for all Ω and f. One may also ask whether the inequality persists or not in the critical case $A = A(\mathbb{R}^n)$.

The preceding power series argument clearly does not realize the optimal coefficient. The first computation of the best coefficient in Trudinger's inequality was given by Moser, who showed that

$$A(\mathbb{R}^n) = n\omega_{n-1}^{1/(n-1)}$$

and that (9.9) holds with $A = A(\mathbb{R}^n)$. The logarithmic potential mentioned above also shows that no larger value for A is admissible. In this sharp form (with the best possible exponent β and coefficient A), we refer to (9.9) as the *Moser–Trudinger inequality*.

In this section, we sketch the proof of the sharp Moser–Trudinger inequality in the first Heisenberg group $\mathbb{H}$, giving the best exponent and coefficient for exponential integrability of $S^{1,4}(\mathbb{H})$ functions. This result is due to Cohn and Lu.

Theorem 9.7 (Cohn–Lu). *For*

$$A = A(\mathbb{H}) = 4\pi^{2/3}, \tag{9.13}$$

the inequality

$$\frac{1}{|\Omega|}\int_\Omega \exp\left(A(\mathbb{H})\frac{|f|}{||\nabla_0 f||_{4,\Omega}}\right)^{4/3} \le C_0 \tag{9.14}$$

holds with an absolute constant C_0, for all domains $\Omega \subset \mathbb{H}$ of finite measure and all nonconstant functions $f \in C_0^\infty(\Omega)$. Moreover, the expression in (9.14) admits no uniform upper bound over such Ω and f for any $A > A(\mathbb{H})$.

We begin the proof of Theorem 9.7 by observing that the fundamental solution to the conformally invariant 4-Laplace operator

$$\mathcal{L}_4 u = X_1(|\nabla_0 u|^2 X_1 u) + X_2(|\nabla_0 u|^2 X_2 u)$$

takes the form

$$\Gamma_4 = -\frac{1}{\pi^2}\cdot \log\frac{1}{N},$$

where, as before, $N(x) = ||x||_{\mathbb{H}}$ denotes the Korányi norm. This can be proved by an adaptation of the argument we used in Section 5.2 for the standard (2-)Laplace operator. Integrating by parts yields the sharp representation formula

$$f(o) = -\frac{1}{\pi^2}\int_{\mathbb{H}} \frac{|\nabla_0 N|^2}{N^3}\langle \nabla_0 N, \nabla_0 f\rangle_1,$$

valid for compactly supported smooth functions f on $\mathbb{H}$. By precomposition with a group translation, we obtain

$$f(x) = \frac{1}{\pi^2}\int_{\mathbb{H}} \frac{|\nabla_0 N(y)|^2}{N(y)^3}\langle \nabla_0 N(y), \nabla_0 f(xy^{-1})\rangle_1 \, dy \tag{9.15}$$

for all $f \in C_0^\infty(\mathbb{H})$ and all $x \in \mathbb{H}$. Thus

$$|f| \le \frac{1}{\pi^2} K * |\nabla_0 f|, \tag{9.16}$$

where $*$ denotes convolution and

$$K = \frac{|\nabla_0 N|^3}{N^3}. \tag{9.17}$$

The estimate (9.16) reduces the problem from an optimal embedding inequality for the first-order Sobolev space $S^{1,4}(\mathbb{H})$ to an optimal embedding inequality for convolutions with the potential K in the Lebesgue space $L^4(\mathbb{H})$. More precisely, we see that (9.14) would follow if we could prove the implication

$$\int_\Omega F^4 \le 1 \quad \Rightarrow \quad \frac{1}{|\Omega|}\int_\Omega \exp\left(\frac{4}{\pi^2}(K*F)^{4/3}\right) \le C_0 \tag{9.18}$$

for all $\Omega \subset \mathbb{H}$ of finite measure and all non-zero, nonnegative $F \in C_0^\infty(\Omega)$.

The proof of (9.18) uses an argument of Adams involving O'Neil's Lemma on rearrangements of convolutions. For a nonnegative function F on $\mathbb{H}$ we consider the non-increasing rearrangement

$$F^*(t) = |\{s > 0 : |\{x \in \mathbb{H} : F(x) > s\}| \le t\}|.$$

As with any rearrangement, we have the identity

$$\int_{\mathbb{H}} u \circ F = \int_0^\infty u \circ F^*(t)\,dt \tag{9.19}$$

for any measurable $u : [0,\infty) \to [0,\infty)$. An elementary computation using (9.17) gives

$$K^*(t) = (\frac{\pi}{2})^{3/2} t^{-3/4}.$$

Recast in terms of the rearrangements of F and K, (9.18) reads

$$\int_0^{|\Omega|} F^*(t) \le 1 \quad \Rightarrow \quad \frac{1}{|\Omega|}\int_0^{|\Omega|} \exp\left(\frac{4}{\pi^2}(K*F)(t)^{4/3}\right) \le C_0. \tag{9.20}$$

An application of O'Neil's Lemma [259, Chapter 1]:

$$(h*g)^*(t) \le \frac{1}{t}\int_0^t h^*(s)\,ds \int_0^t g^*(s)\,ds + \int_t^\infty h^*(s)g^*(s)\,ds$$

with $h = K$ and $g = |\nabla_0 f|$ shows that (9.20) follows in turn from

$$\begin{aligned}&\int_0^{|\Omega|} F^*(t) \le 1 \\ &\Rightarrow \frac{1}{|\Omega|}\int_0^{|\Omega|} \exp\left(\left(\int_0^t F^*(s)\,ds + \left(\frac{2}{\pi}\right)^{3/2}\int_t^\infty s^{-3/4}F^*(s)\,ds\right)^{4/3}\right) \le C_0.\end{aligned} \tag{9.21}$$

After an ingenious change of variables $t = |\Omega|e^{-\tau}$, (9.21) reduces to a lemma of Adams [2]. Similarly to the Euclidean case, suitable truncations of the logarithmic potential $f(x) = \log 1/||x||_{\mathbb{H}}$ in $\Omega = \{x \in \mathbb{H} : ||x||_{\mathbb{H}} < 1\}$ show that no larger value for A can be chosen.

9.3 Hardy inequality

The classical Hardy inequalityasserts that

$$\int_{\mathbb{R}^n} \frac{|u(x)|^2}{|x|^2}\,dx \le \left(\frac{2}{n-2}\right)^2 \int_{\mathbb{R}^n} |\nabla u(x)|^2\,dx \tag{9.22}$$

for all $u \in C_0^\infty(\mathbb{R}^n \setminus \{0\})$ and $n \ge 3$, where the constant $\left(\frac{2}{n-2}\right)^2$ is best possible. Inequality (9.22) plays a key role in the analysis of the inhomogeneous heat equation

$$\begin{cases} u_t - \triangle u = \lambda |x|^{-2} u, x \in \mathbb{R}^n, 0 < t < T, \\ u(x,0) = u_0(x), \end{cases} \tag{9.23}$$

where $u_0 \in L^2(\mathbb{R}^n)$ is positive and $\lambda \in \mathbb{R}$. Indeed, (9.23) has a solution if and only if $\lambda \le (\frac{2}{n-2})^2$. To conclude this chapter, we state an analog of (9.22) in the first Heisenberg group.

Theorem 9.8. *For $u \in C_0^\infty(\mathbb{H} \setminus \{o\})$, the inequality*

$$\int_{\mathbb{H}} \frac{|z|^2 u(z,x_3)^2}{||(z,x_3)||_H^4}\,dz\,dx_3 \le \int_{\mathbb{H}} |\nabla_0 u(z,x_3)|^2\,dz\,dx_3 \tag{9.24}$$

holds. Moreover, (9.24) *is sharp in the following sense:*

$$\sup_u \frac{\int_{\mathbb{H}} |z|^2 ||(z,x_3)||_H^{-4} u(z,x_3)^2\,dz\,dx_3}{\int_{\mathbb{H}} |\nabla_0 u(z,x_3)|^2\,dz\,dx_3} = 1,$$

where the supremum is taken over all $u \in C_0^\infty(\mathbb{H} \setminus \{o\})$.

Proof. As in the preceding section, we denote by N the Korányi norm. Set $u = N^{-1}v$. Then

$$|\nabla_0 u|^2 = N^{-4}|\nabla_0 N|^2 v^2 - 2N^{-3}\langle \nabla_0 N, \nabla_0 v\rangle_1 + N^{-2}|\nabla_0 v|^2. \tag{9.25}$$

Since $|\nabla_0 N| = |z|/N$, we find

$$\begin{aligned} \int_{\mathbb{H}} |\nabla_0 u|^2 &= \int_{\mathbb{H}} |z|^2 N^{-6} v^2 - 2\int_{\mathbb{H}} N^{-3} v \langle \nabla_0 N, \nabla_0 v\rangle_1 + \int_{\mathbb{H}} N^{-2}|\nabla_0 v|^2 \\ &= \int_{\mathbb{H}} |z|^2 N^{-6} v^2 + \frac{1}{2}\int_{\mathbb{H}} \langle \nabla_0(N^{-2}), \nabla_0(v^2)\rangle_1 + \int_{\mathbb{H}} N^{-2}|\nabla_0 v|^2 \end{aligned}$$

by integrating (9.25) over $\mathbb{H}$. Integrating by parts in the middle term and using Theorem 5.15 gives

$$\int_{\mathbb{H}} |\nabla_0 u|^2 \ge \int_{\mathbb{H}} |z|^2 N^{-6} v^2 = \int_{\mathbb{H}} |z|^2 N^{-4} u^2.$$

(Recall that u, hence also v, was compactly supported in the complement of o.) To see that the inequality is sharp, consider suitable (smooth) approximations to the functions $u_\epsilon := N^{-1}\chi_{\{\epsilon < N < 1/\epsilon\}}$, $\epsilon > 0$. □

9.4 Notes

Notes for Section 9.1**.** The best constant in the Sobolev inequality (9.8) was computed by Talenti [245] and Aubin [17]. For each $1 \le p < n$, the inequality (9.8) holds for all $f \in C_0^\infty(\mathbb{R}^n)$ with

$$C_p(\mathbb{R}^n) = \left(\frac{1}{n}\right)^{(n-p)/np} \left(\left(\frac{p-1}{n-p}\right)^{n/q} \frac{\Gamma(n)}{\Gamma(n/p)\Gamma(1+n/q)\omega_{n-1}} \right)^{1/n} \tag{9.26}$$

if $1 < p < n$, or

$$C_1(\mathbb{R}^n) = \frac{1}{n^{(n-1)/n}\omega_{n-1}^{1/n}}.$$

Here Γ denotes the Gamma function, ω_{n-1} is the surface area of the unit sphere in $\mathbb{R}^n$, and $q = p/(p-1)$ is the Hölder conjugate. Moreover, equality is attained in (9.8) only for functions of the form $f(x) = (a + b|x|^q)^{-n/p}$ for suitable $a, b > 0$.

The L^2-Sobolev inequality is of interest in view of its connection with the classical Yamabe problem, which asks which metrics on a compact Riemannian manifold are conformally equivalent with a metric of constant scalar curvature. In the CR analog of this problem, the usual (Hermitian) scalar curvature is replaced by the (pseudo-Hermitian) Webster–Tanaka curvature. Jerison and Lee show that the only contact forms on $\mathbb{S}^3 \subset \mathbb{C}^2$ with constant Webster curvature are the images of the standard contact form $\hat\theta = \frac{i}{2}(\overline{\partial} - \partial)|z|^2$ under CR automorphisms of $\mathbb{S}^3$. For a fixed $p_0 \in \mathbb{S}^3$, $\mathbb{H}$ is CR equivalent with $\mathbb{S}^3 \setminus \{p_0\}$ via the Cayley transform. (See Section 3.3.) Under this transform, $\hat\theta$ is taken to $2u_{JL}^2(dx_3 - \frac{1}{2}x_1\,dx_2 + \frac{1}{2}x_2\,dx_1)$, where

$$u_{JL}(z, x_3) = ((1+|z|^2)^2 + 16x_3^2)^{-1/2}. \tag{9.27}$$

The images of the conformal factor u_{JL} under left translations and dilations of the Heisenberg group are precisely the minimizers for the L^2-Sobolev inequality on $\mathbb{H}$. In this way, Jerison and Lee [156] were led to a proof of Theorem 9.1. Garofalo and Vassilev [119] studied conformal geometry and the analog of the CR Yamabe problem in groups of Heisenberg type, and obtained an extension of Theorem 9.1 in that setting for first-layer radially symmetric functions. The proof of Theorem 9.1 which we gave is a slightly simplified version of that of Garofalo and Vassilev, restricted to the setting of $\mathbb{H}$.

We stress the fact that the extremal u_{JL} is not a function of the Korányi gauge. Similarly, the bubble sets $\mathcal{B}(o, R)$ do not admit implicit representations in terms of this gauge. On the other hand, the extremals for Trudinger's inequality on $\mathbb{H}$ **are** expressed in terms of this gauge. This highlights an essential difficulty in determining the sharp L^p Sobolev inequalities on $\mathbb{H}$, not present in the Euclidean case: the qualitative characteristics of the extremal functions necessarily vary with p.

Monti and Morbidelli's solution to the isoperimetric problem on the Grushin spaces G_α yields the sharp constant in the L^1 Sobolev inequality. Indeed, for each $\alpha \geq 0$, the inequality

$$\int_{\mathbb{R}^2} |f|^{(2+a)/(1+a)} \leq C_{\mathrm{iso}}(G_\alpha) \left(\int_{\mathbb{R}^2} \sqrt{f_x^2 + |x|^{2\alpha} f_y^2} \right)^{(1+\alpha)/(2+\alpha)}$$

holds for all $f \in C_0^\infty(\mathbb{R}^2)$, and is sharp. In particular, the sharp L^1 Sobolev inequality on the Grushin plane $G = G_1$ is given by

$$\int_{\mathbb{R}^2} |f|^{3/2} \leq \frac{1}{3} \left(\int_{\mathbb{R}^2} \sqrt{f_x^2 + |x|^2 f_y^2} \right)^{2/3}.$$

The sharp L^2 Sobolev inequality on the Grushin plane was obtained by Beckner [31] in the form

$$\left(\int_{\mathbb{R}^2} |f|^6 \right)^{1/6} \leq \pi^{-1/3} \left(\int_{\mathbb{R}^2} (f_x^2 + 4x^2 f_y^2) \right)^{1/2} \tag{9.28}$$

for all $f \in C_0^\infty(\mathbb{R}^2)$. Beckner's proof relies on a scaling transformation which recasts (9.28) as a Sobolev inequality on the hyperbolic plane $SL(2,\mathbb{R})/SO(2)$. We also point out the recent paper by Monti [206] where symmetrization techniques are used to prove sharp Sobolev inequalities in Grushin spaces.

Notes for Section 9.2**.** The proof of (9.9) via power series is due to Trudinger [249]. The technique works in great generality, see Saloff-Coste [235] for general Carnot groups, Danielli [77] for the general Hörmander vector fields, and Hajłasz and Koskela [132] for general metric spaces, using the notion of *upper gradient*.

Cohn and Lu were the first to obtain best constants in Moser–Trudinger-type inequalities in non-abelian Carnot groups. Theorem 9.7 is the $n = 1$ case of their paper [69] on Moser–Trudinger inequalities in the Heisenberg groups $\mathbb{H}^n$. The use of the conformally invariant 4-sub-Laplacian in the proof of Theorem 9.7 which we give is inspired by the approach in Balogh, Manfredi and Tyson [24], where the sharp Moser–Trudinger inequality is obtained in arbitrary Carnot groups. The technique employed is similar to that of Cohn and Lu, but begins with a general representation formula arising from the fundamental solution to the Q-sub-Laplacian. The nonlinear potential theory of the conformally invariant Q-sub-Laplace equation on Carnot groups of homogeneous dimension Q was studied by Balogh, Holopainen and Tyson in [23], where applications to the regularity of quasiconformal maps were obtained. For explicit computations of the best constant in the Moser–Trudinger inequality in H-type groups, see Cohn–Lu [70] and Balogh–Tyson [26].

Note that on all non-abelian groups (including the Heisenberg group), Adams' technique (developed in [2]) is at present the only viable method to find the sharp form of the Trudinger inequality; Moser's original approach encounters serious difficulties due to a lack of information regarding the behavior of horizontal energy norms under symmetrization. Manfredi and Vera de Serio [193] have shown that radial symmetrization in non-abelian Carnot groups is substantially less well-behaved than its Euclidean counterpart.

In [253], sharp Young's inequalities for weighted convolution operators on Carnot groups are developed and applied to the study of sharp weighted Moser–Trudinger inequalities for first-layer symmetric functions in the Heisenberg group. By passing to a suitable quotient, these estimates imply some sharp weighted Moser–Trudinger inequalities for x_1-symmetric functions on the Grushin plane. However, the class of admissible weights for the result in [253] does not cover the original (unweighted) case, which is still open.

Notes for Section 9.3**.** The Hardy inequality in the Heisenberg group is due to Garofalo and Lanconelli [115]. Niu, Zhang and Wang [215] found an L^p analog of Theorem 9.8, while Kombe [165] generalized Theorem 9.8 in a different direction, including more general Korányi radial weights in the integrals and extending the result to more general classes of Carnot groups. For Hardy-type inequalities in the setting of the Grushin plane, see D'Ambrosio [76].

Bibliography

[1] Abramowitz, M., and Stegun, I.A., Eds. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.

[2] Adams, D. A sharp inequality of J. Moser for higher order derivatives. *Ann. of Math.* 128 (1988), 385–398.

[3] Adams, D., and Hedberg, L. *Function spaces and potential theory.* Springer-Verlag, Berlin, 1996.

[4] Aebischer, B., Borer, M.; Kälin, M., Leuenberger, C., and Reimann, H.M. *Symplectic geometry. An introduction based on the seminar in Bern,* 1992. Progress in Mathematics, 124. Birkhäuser Verlag, 1994.

[5] Agrachev, A. Introduction to optimal control theory. In *Mathematical control theory, Part* 1, 2 (*Trieste,* 2001), ICTP Lect. Notes, VIII. Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2002, pp. 453–513 (electronic).

[6] Agrachev, A., and Gauthier, J.-P. On the subanalyticity of Carnot-Carathéodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire* 18, 3 (2001), 359–382.

[7] Ambrosio, L. Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces. *Adv. Math.* 159, 1 (2001), 51–67.

[8] Ambrosio, L. Transport equation and Cauchy problem for BV vector fields. *Invent. Math.* 158, 2 (2004), 227–260.

[9] Ambrosio, L., and Kirchheim, B. Currents in metric spaces. *Acta Math.* 185, 1 (2000), 1–80.

[10] Ambrosio, L., and Rigot, S. Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.* 208, 2 (2004), 261–301.

[11] Ambrosio, L., Serra-Cassano, F., and Vittone, D. Intrinsic regular hypersurfaces in Heisenberg groups. *J. Geom. Anal.* 16, 2 (2006), 187–232.

[12] Arcozzi, N., and Baldi, A. From Gruschin to Heisenberg via an isoperimetric problem. Preprint, 2006.

[13] Arcozzi, N., and Ferrari, F. Hessian of the distance function from a surface in the Heisenberg group. Preprint, 2005.

[14] Arcozzi, N., and Ferrari, F. Metric normal and distance function in the Heisenberg group. *Math. Z.* to appear.

[15] Arcozzi, N., and Morbidelli, D. A global Inverse Map Theorem and bi-Lipschitz maps in the Heisenberg group. Preprint, 2005.

[16] Arcozzi, N., and Morbidelli, D. Stability of isometric maps in the Heisenberg group. Preprint, 2005.

[17] Aubin, T. Best constants in the Sobolev imbedding theorem: the Yamabe problem. In *Seminar on Differential Geometry.* Princeton Univ. Press, Princeton, N.J., 1982, pp. 173–184.

[18] Baernstein, II, A. A unified approach to symmetrization. In *Partial differential equations of elliptic type* (*Cortona,* 1992), Sympos. Math., XXXV. Cambridge Univ. Press, Cambridge, 1994, pp. 47–91.

[19] Ball, K. An elementary introduction to modern convex geometry. In *Flavors of geometry*, vol. 31 of *Math. Sci. Res. Inst. Publ.* Cambridge Univ. Press, Cambridge, 1997, pp. 1–58.

[20] Balogh, Z. Size of characteristic sets and functions with prescribed gradient. *J. Reine Angew. Math.* 564 (2003), 63–83.

[21] Balogh, Z., and Bonk, M. Gromov hyperbolicity and the Kobayashi metric on strictly pseudoconvex domains. *Comment. Math. Helv.* 3, 75 (2000), 504–533.

[22] Balogh, Z., Hofer-Isenegger, R., and Tyson, J. Lifts of Lipschitz maps and horizontal fractals in the Heisenberg group. *Ergodic Theory Dynam. Systems* 26 (2006), 621–651.

[23] Balogh, Z., Holopainen, I., and Tyson, J. Singular solutions, homogeneous norms and quasiconformal mappings on Carnot groups. *Math. Ann.* 324, 1 (2002), 159–186.

[24] Balogh, Z., Manfredi, J., and Tyson, J. Fundamental solution for the Q-Laplacian and sharp Moser-Trudinger inequality in Carnot groups. *J. Funct. Anal.* 204, 1 (2003), 35–49.

[25] Balogh, Z., Rickly, M., and Serra-Cassano, F. Comparison of Hausdorff measures with respect to the Euclidean and Heisenberg metric. *Publ. Mat.* 47 (2003), 237–259.

[26] Balogh, Z., and Tyson, J. Polar coordinates on Carnot groups. *Math. Z.* 241, 4 (2002), 697–730.

[27] Bandle, C. *Isoperimetric inequalities and applications.* Monographs and Studies in Mathematics, 7. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1980.

[28] Barone-Adesi, V., Serra-Cassano, F., and Vittone, D. The Bernstein problem for intrinsic graphs in the Heisenberg group and calibrations. Preprint, 2006.

[29] Beals, R., Gaveau, B., and Greiner, P. The Green function of model step two hypoelliptic operators and the analysis of certain tangential Cauchy Riemann complexes. *Adv. Math.* 121, 2 (1996), 288–345.

[30] Beals, R., and Greiner, P. *Calculus on Heisenberg manifolds*, vol. 119 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1988.

[31] Beckner, W. Sharp Sobolev inequalities on the sphere and the Moser-Trudinger inequality. *Ann. of Math.* 138 (1993), 213–242.

[32] Bellaïche, A. The tangent space in sub-Riemannian geometry. In *Sub-Riemannian geometry*. Birkhäuser, Basel, 1996, pp. 1–78.

[33] Bieske, T. On ∞-harmonic functions on the Heisenberg group. *Comm. Partial Differential Equations* 3-4, 27 (2002), 727–761.

[34] Bieske, T. Comparison principle for parabolic equations in the Heisenberg group. *Electron. J. Differential Equations* 95 (2005), 11 pp. (electronic).

[35] Biroli, M., and Mosco, U. Sobolev and isoperimetric inequalities for Dirichlet forms on homogeneous spaces. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* (9) *Mat. Appl.* 6, 1 (1995), 37–44.

[36] Bloch, A. *Nonholonomic mechanics and control. With the collaboration of J. Baillieul, P. Crouch and J. Marsden. With scientific input from P.S. Krishnaprasad, R.M. Murray and D. Zenkov.* No. 24 in Interdisciplinary Applied Mathematics, Systems and Control. Springer-Verlag, 2003.

[37] Bonk, M., and Capogna, L. Mean curvature flow in the Heisenberg group. Preprint, 2005.

[38] Bonk, M., and Kleiner, B. Rigidity for quasi-Fuchsian actions on negatively curved spaces. *Int. Math. Res. Not.* 61 (2004), 3309–3316.

[39] Bonnard, B., and Chyba, M. *Singular trajectories and their role in control theory*, vol. 40 of *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*]. Springer-Verlag, Berlin, 2003.

[40] Bourdon, M. Structure conforme au bord et flot géodésique d'un CAT(-1)-espace. *Enseign. Math.* 41 (1995), 63–102.

[41] Bourdon, M. Sur le birapport au bord des CAT(-1)-espaces. *Inst. Hautes Études Sci. Publ. Math.* 83 (1996), 95–104.

[42] Brakke, K.A. *The motion of a surface by its mean curvature*, vol. 20 of *Mathematical Notes*. Princeton University Press, Princeton, N.J., 1978.

[43] Bridson, M.R., and Haefliger, A. *Metric spaces of non-positive curvature*, vol. 319 of *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*]. Springer-Verlag, Berlin, 1999.

[44] Brock, F., and Solynin, A.Y. An approach to symmetrization via polarization. *Trans. Amer. Math. Soc.* 352, 4 (2000), 1759–1796.

[45] Brockett, R.W. Control theory and singular Riemannian geometry. In *New directions in applied mathematics* (*Cleveland, Ohio,* 1980). Springer-Verlag, New York, 1982, pp. 11–27.

[46] Buckley, S., Koskela, P., and Lu, G. Boman equals John. In *XVIth Rolf Nevanlinna Colloquium* (*Joensuu,* 1995). de Gruyter, Berlin, 1996, pp. 91–99.

[47] Burago, D., Burago, Y., and Ivanov, S. *A course in metric geometry*, vol. 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[48] Burago, Y.D., and Zalgaller, V.A. *Geometric inequalities*, vol. 285 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1988. Translated from the Russian by A.B. Sosinskiĭ, Springer Series in Soviet Mathematics.

[49] Campanato, S. Proprietà di hölderianità di alcune classi di funzioni. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (3) 17 (1963), 175–188.

[50] Capogna, L. Regularity of quasilinear equations in the Heisenberg group. *Comm. Pure Appl. Math.* 50, 9 (1997), 867–889.

[51] Capogna, L., and Citti, G. Generalized mean curvature flow in Carnot groups. Preprint, 2005.

[52] Capogna, L., Danielli, D., and Garofalo, N. An embedding theorem and the Harnack inequality for nonlinear subelliptic equations. *Comm. Partial Differential Equations* 18, 9-10 (1993), 1765–1794.

[53] Capogna, L., Danielli, D., and Garofalo, N. An isoperimetric inequality and the Sobolev embedding theorem for vector fields. *Math. Res. Lett.* 1 (1994), 263–268.

[54] Capogna, L., Danielli, D., and Garofalo, N. Capacitary estimates and the local behavior of solutions of nonlinear subelliptic equations. *Amer. J. Math.* 118 (1997), 1153–1196.

[55] Capogna, L., and Garofalo, N. Boundary behavior of nonnegative solutions of subelliptic equations in NTA domains for Carnot-Carathéodory metrics. *J. Fourier Anal. Appl.* 4, 4-5 (1998), 403–432.

[56] Capogna, L., Pauls, S., and Tyson, J. Convexity and horizontal second fundamental forms for hypersurfaces in Carnot groups. Preprint, 2006.

[57] Capogna, L., and Tang, P. Uniform domains and quasiconformal mappings on the Heisenberg group. *Man. Math.* 86 (1995), 267–281.

[58] Catlin, D. Estimates of invariant metrics on pseudoconvex domains of dimension two. *Math. Z.* 200, 3 (1989), 429–466.

[59] Chavel, I. *Isoperimetric inequalities*, vol. 145 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2001. Differential geometric and analytic perspectives.

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

[61] Cheeger, J., and Kleiner, B. Generalized differential and bi-Lipschitz nonembedding in L^1. *C. R. Math. Acad. Sci. Paris* 343, 5 (2006), 297–301.

[62] Chen, Y., Giga, Y., and Goto, S. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *J. Diff. Geom.* 33, 3 (1991), 749–786.

[63] Cheng, J.-H., Hwang, J.-F., Malchiodi, A., and Yang, P. Minimal surfaces in pseudohermitian geometry. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) 4, 1 (2005), 129–177.

[64] Cheng, J.-H., Hwang, J.-F., and Yang, P. Existence and uniqueness for p-area minimizers in the Heisenberg group. *Math. Ann.* (to appear).

[65] Citti, G., and Manfredini, M. An implicit function theorem in Carnot-Carathéodory spaces. Preprint 2005.

[66] Citti, G., Manfredini, M., and Sarti, A. On the Mumford-Shah functional in a contact manifold on the Heisenberg space. Preprint, 2006.

[67] Citti, G., Manfredini, M., and Sarti, A. Neuronal oscillation in the visual cortex: Gamma-convergence to the Riemannian Mumford-Shah functional. *SIAM Journal of Mathematical Analysis* 35, 6 (2003), 1394–1419.

[68] Citti, G., and Sarti, A. A cortical based model of perceptual completion in the roto-translation space. *J. Math. Imaging Vision* 24, 3 (2006), 307–326.

[69] Cohn, W., and Lu, G. Best constants for Moser-Trudinger inequalities on the Heisenberg group. *Indiana Univ. Math. J.* 50, 4 (2001), 1567–1591.

[70] Cohn, W., and Lu, G. Best constants for Moser-Trudinger inequalities, fundamental solutions and one-parameter representation formulas on groups of Heisenberg type. *Acta Math. Sin.* (*Engl. Ser.*) 18, 2 (2002), 375–390.

[71] Coifman, R., and Weiss, G. *Analyse harmonique non-commutative sur certains espaces homogènes.* (*French*) *Ètude de certaines intégrales singulières*, vol. 242 of *Lecture Notes in Mathematics.* Springer-Verlag, 1971.

[72] Corwin, L., and Greenleaf, F.P. *Representations of nilpotent Lie groups and their applications, Part I: Basic theory and examples.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1990.

[73] Cowling, M., Dooley, A., Korányi, A., and Ricci, F. H-type groups and Iwasawa decompositions. *Adv. Math.* 87 1 (1991), 1–41.

[74] Croke, C. Some isoperimetric inequalities and eigenvalue estimates. *Ann. Sci. École Norm. Sup.* (4) 13 (1980), 419–435.

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

[76] D'Ambrosio, L. Hardy inequalities related to Grushin type operators. *Proc. Amer. Math. Soc.* 132, 3 (2004), 725–734 (electronic).

[77] Danielli, D. Formules de représentation et théorèmes d'inclusion pour des opérateurs sous-elliptiques. *C. R. Acad. Sci. Paris Sér. I Math.* 314, 13 (1992), 987–990.

[78] Danielli, D., Garofalo, N., and Nhieu, D.-M. Minimal surfaces, surfaces of constant mean curvature and isoperimetry in Carnot groups. Preprint, 2001.

[79] Danielli, D., Garofalo, N., and Nhieu, D.-M. Sub-Riemannian calculus on hypersurfaces in Carnot groups. Preprint, 2006.

[80] Danielli, D., Garofalo, N., and Nhieu, D.-M. A notable family of entire intrinsic minimal graphs in the Heisenberg group which are not perimeter minimizing. *Amer. J. Math.* (to appear).

[81] Danielli, D., Garofalo, N., and Nhieu, D.-M. A partial solution of the isoperimetric problem for the Heisenberg group. *Forum Mathematicum* (to appear).

[82] Danielli, D., Garofalo, N., Nhieu, D.-M., and Pauls, S.D. Instability of graphical strips and a positive answer to the Bernstein problem in the Heisenberg group. Preprint, 2006.

[83] Danielli, D., Garofalo, N., Nhieu, D.-M., and Pauls, S.D. Stable c^2 complete embedded noncharacteristic h-minimal surfaces in the Heisenberg group are vertical planes. Preprint, 2006.

[84] De Giorgi, E. Su una teoria generale della misura $(r-1)$-dimensionale in uno spazio ad r dimensioni. (Italian.) *Ann. Mat. Pura Appl.* 4, 36 (1954), 191–213.

[85] De Giorgi, E. Sulla proprietà isoperimetrica dell'ipersfera, nella classe degli insiemi aventi frontiera orientata di misura finita. *Atti Accad. Naz. Lincei. Mem. Cl. Sci. Fis. Mat. Nat. Sez. I* (8) 5 (1958), 33–44.

[86] Derridj, M. Sur un théorème de traces. *Ann. Inst. Fourier* (*Grenoble*) 22, 2 (1972), 73–83.

[87] DiPerna, R., and Lions, P.-L. Ordinary differential equations, transport theory, and Sobolev spaces. *Invent. Math.* 98, 3 (1989), 511–547.

[88] do Carmo, M.P. *Riemannian geometry.* Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[89] Duren, P. *Univalent functions.* Springer-Verlag, New York, 1983.

[90] Evans, L., and Spruck, J. Motion of level sets by mean curvature. I. *J. Diff. Geom.* 33, 3 (1991), 635–681.

[91] Evans, L., and Spruck, J. Motion of level sets by mean curvature. II. *Trans. Amer. Math. Soc.* 330, 1 (1992), 321–332.

[92] Evans, L., and Spruck, J. Motion of level sets by mean curvature. III. *J. Geom. Anal.* 2, 2 (1992), 121–150.

[93] Evans, L., and Spruck, J. Motion of level sets by mean curvature. IV. *J. Geom. Anal.* 5, 1 (1995), 77–114.

[94] Falconer, K.J. *Fractal geometry.* Mathematical Foundations and Applications. John Wiley and Sons Ltd., Chichester, 1990.

[95] Federer, H. *Geometric Measure Theory*, vol. 153 of *Die Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, New York, 1969.

[96] Field, D., Hayes, A., and Hess, R. Contour integration by the human visual system: evidence for a local "association field". *Vision Res.* 33 (1993), 173–193.

[97] Fleming, W., and Rishel, R. An integral formula for total gradient variation. *Arch. Math.* 11 (1960), 218–222.

[98] Folland, G. A fundamental solution for a subelliptic operator. *Bull. Amer. Math. Soc.* 79 (1973), 373–376.

[99] Folland, G. Subelliptic estimates and function spaces on nilpotent Lie groups. *Ark. Mat.* 2, 13 (1975), 161–207.

[100] Folland, G. *Harmonic analysis in phase space*, vol. 122 of *Annals of Mathematics Studies.* Princeton University Press, Princeton, NJ, 1989.

[101] Folland, G., and Kohn, J. *The Neumann problem for the Cauchy-Riemann complex.* Princeton University Press, Princeton, N.J., 1972. Annals of Mathematics Studies, No. 75.

[102] Folland, G., and Stein, E. Estimates for the $\bar{\partial}_b$ complex and analysis on the Heisenberg group. *Comm. Pure Appl. Math.* 27 (1974), 429–522.

[103] Folland, G., and Stein, E. *Hardy spaces on homogeneous groups.* Princeton University Press, Princeton, New Jersey, 1982.

[104] Franchi, B., Gallot, S., and Wheeden, R. Sobolev and isoperimetric inequalities for degenerate metrics. *Math. Ann.* 4, 300 (1994), 557–571.

[105] Franchi, B., Serapioni, R., and Serra-Cassano, F. Regular submanifolds, graphs and area formula in Heisenberg groups. Preprint, 2004.

[106] Franchi, B., Serapioni, R., and Serra-Cassano, F. Rectifiability and perimeter in the Heisenberg group. *Math. Ann.* 321, 3 (2001), 479–531.

[107] Franchi, B., Serapioni, R., and Serra-Cassano, F. Rectifiability and perimeter in step 2 groups. *Math. Bohemica* 127, 2 (2002), 219–228.

[108] Franchi, B., Serapioni, R., and Serra-Cassano, F. On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.* 13, 3 (2003), 421–466.

[109] Franchi, B., Serapioni, R., and Serra-Cassano, F. Regular hypersurfaces, intrinsic perimeter and implicit function theorem in Carnot groups. *Comm. Anal. Geom.* 11, 5 (2003), 909–944.

[110] Gage, M. Curve shortening makes convex curves circular. *Invent. Math.* 76, 2 (1984), 357–364.

[111] Gage, M. *On an area-preserving evolution equation for plane curves.* No. 51 in Contemp. Math. Amer. Math. Soc., 1986, pp. 51–62.

[112] Gage, M., and Hamilton, R. The heat equation shrinking convex plane curves. *J. Diff. Geom.* 23, 1 (1986), 69–96.

[113] Gallot, S., Hulin, D., and Lafontaine, J. *Riemannian geometry*, second ed. Springer-Verlag, Berlin, 1990.

[114] Garofalo, N. Analysis and geometry of Carnot-Carathéodory spaces, with applications to PDE's. Birkhäuser, in preparation.

[115] Garofalo, N., and Lanconelli, E. Frequency functions on the Heisenberg group, the uncertainty principle and unique continuation. *Ann. Inst. Fourier (Grenoble)* 40, 2 (1990), 313–356.

[116] Garofalo, N., and Nhieu, D.-M. Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. *Comm. Pure Appl. Math.* 49, 10 (1996), 1081–1144.

[117] Garofalo, N., and Pauls, S. The Bernstein problem in the Heisenberg group. Preprint, 2003.

[118] Garofalo, N., and Vassilev, D. Regularity near the characteristic set in the non-linear Dirichlet problem and conformal geometry of sub-Laplacians on Carnot groups. *Math. Ann.* 318 (2000), 453–516.

[119] Garofalo, N., and Vassilev, D. Symmetry properties of positive entire solutions of Yamabe-type equations on groups of Heisenberg type. *Duke Math. J.* 106, 3 (2001), 411–448.

[120] Gaveau, B. Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. *Acta Math.* 139, 1-2 (1977), 95–153.

[121] Giaquinta, M. *Introduction to regularity theory for nonlinear elliptic systems.* Birkhäuser Verlag, Basel, 1993.

[122] Giaquinta, M., and Hildebrandt, S. *Calculus of Variations, I & II*, vol. 310, 311 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, 1996.

[123] Gilbarg, D., and Trudinger, N.S. *Elliptic partial differential equations of second order.* Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[124] Gilbert, C., Das, A., Ito, M., and Westheimer, G. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences USA* 93 (1996), 615–622.

[125] Giusti, E. *Minimal surfaces and functions of bounded variation.* No. 80 in Monographs in Mathematics. Birkhäuser Verlag, 1984.

[126] Goldman, W. *Complex hyperbolic geometry.* Oxford Mathematical Monographs. Oxford University Press, 1999.

[127] Gong, J., and Manfredi, J. personal communication.

[128] Grayson, M. The heat equation shrinks embedded plane curves to round points. *J. Diff. Geom.* 26, 2 (1987), 285–314.

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

[130] Gromov, M. Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian Geometry*, vol. 144 of *Progress in Mathematics*. Birkhäuser, Basel, 1996, pp. 79–323.

[131] Gromov, M. *Metric structures for Riemannian and non-Riemannian spaces*. Birkhäuser Boston Inc., Boston, MA, 1999. Edited by J. LaFontaine and P. Pansu. Based on the 1981 French original, translated from the French by Sean Michael Bates.

[132] Hajłasz, P., and Koskela, P. Sobolev met Poincaré. *Memoirs Amer. Math. Soc.* 145, 688 (2000).

[133] Hamilton, R. Three-manifolds with positive Ricci curvature. *J. Differential Geom.* 17, 2 (1982), 255–306.

[134] Hardy, G., Littlewood, J., and Pólya, G. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1952 edition.

[135] Heinonen, J. Calculus on Carnot groups. In *Fall School in Analysis* (*Jyväskylä,* 1994), vol. 68. Ber. Univ. Jyväskylä Math. Inst., Jyväskylä, 1995, pp. 1–31.

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

[137] Heinonen, J. Nonsmooth calculus. *Bull. Amer. Math. Soc.* (to appear).

[138] Heinonen, J., and Holopainen, I. Quasiregular maps on Carnot groups. *J. Geom. Anal.* 7, 1 (1997), 109–148.

[139] Heinonen, J., Kilpeläinen, T., and Martio, O. *Nonlinear potential theory of degenerate elliptic equations*. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.

[140] Heinonen, J., and Koskela, P. Definitions of quasiconformality. *Invent. Math.* 120 (1995), 61–79.

[141] Heinonen, J., and Koskela, P. Quasiconformal maps in metric spaces with controlled geometry. *Acta Math.* 181 (1998), 1–61.

[142] Heinonen, J., Koskela, P., Shanmugalingam, N., and Tyson, J. Sobolev spaces on metric measure spaces: an approach based on upper gradients. in preparation.

[143] Heinonen, J., Koskela, P., Shanmugalingam, N., and Tyson, J. Sobolev classes of Banach space-valued functions and quasiconformal mappings. *J. Anal. Math.* 85 (2001), 87–139.

[144] Hladky, R., and Pauls, S. Constant mean curvature surfaces in sub-Riemannian spaces. Preprint, 2005.

[145] Hladky, R., and Pauls, S. A disocclusion algorithm based on a model of the visual cortex. In preparation.

[146] Hladky, R., and Pauls, S. Minimal surfaces in the roto-translation group with applications to a neurobiological image completion model. Preprint, 2005.

[147] Hladky, R., and Pauls, S. Variations of perimeter in sub-Riemannian spaces. Preprint, 2006.

[148] Hoffman, W. The visual cortex is a contact bundle. *Appl. Math. Comput.* 32, 2-3 (1989), 137–167.

[149] Hörmander, L. Hypoelliptic second order differential equations. *Acta Math.* 119 (1967), 147–171.

[150] Hubel, D., and Weisel, T. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160 (1962), 106–154.

[151] Hubel, D., and Weisel, T. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. London* (*Biol*) 198 (1977), 1–59.

[152] Huisken, G. Flow by mean curvature of convex surfaces into spheres. *J. Diff. Geom.* 1, 20 (1984), 237–266.

[153] Huisken, G. The volume preserving mean curvature flow. *J. Reine Angew. Math.* 382 (1987), 35–48.

[154] Hurtado, A., and Rosales, C. Area-stationary surfaces inside the sub-riemannian three sphere. Preprint, 2006.

[155] Jerison, D. The Poincaré inequality for vector fields satisfying Hörmander's condition. *Duke Math. J.* 53 (1986), 503–523.

[156] Jerison, D., and Lee, J.M. Extremals for the Sobolev inequality on the Heisenberg group and the CR Yamabe problem. *J. Amer. Math. Soc.* 1 (1988), 1–13.

[157] John, F. Rotation and strain. *Comm. Pure Appl. Math.* 4 (1961), 391–414.

[158] Jurdjevic, V. Non-Euclidean elastica. *Amer. J. Math.* 117, 1 (1995), 93–124.

[159] Jurdjevic, V. Hamiltonian point of view of non-Euclidean geometry and elliptic functions. *Systems Control Lett.* 43, 1 (2001), 25–41. Lie theory and its applications in control (Würzburg, 1999).

[160] Kaplan, A. Fundamental solutions for a class of hypoelliptic pde generated by composition of quadratic forms. *Trans. Amer. Math. Soc.* 1, 258 (1980), 147–153.

[161] Kirchheim, B. Rectifiable metric spaces: local structure and regularity of the Hausdorff measure. *Proc. Amer. Math. Soc.* 121, 1 (1994), 113–123.

[162] Kirchheim, B., and Serra-Cassano, F. Rectifiability and parameterization of intrinsic regular surfaces in the Heisenberg group. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) 3, 4 (2004), 871–896.

[163] Kleiner, B. An isoperimetric comparison theorem. *Invent. Math.* 108, 1 (1992), 37–47.

[164] Kohn, R. New integral estimates for deformations in terms of their nonlinear strains. *Arch. Rational Mech. Anal.* 78, 2 (1982), 131–172.

[165] Kombe, I. Sharp Hardy type inequalities on Carnot groups. Preprint, 2005.

[166] Korányi, A. Kelvin transforms and harmonic polynomials on the Heisenberg group. *J. Funct. Anal.* 49, 2 (1982), 177–185.

[167] Korányi, A. *Geometric aspects of analysis on the Heisenberg group.* Ist. Naz. Alta Mat. Francesco Severi, 1983, pp. 209–258.

[168] Korányi, A., and Reimann, H.M. Quasiconformal mappings on the Heisenberg group. *Invent. Math.* 80 (1985), 309–338.

[169] Korányi, A., and Reimann, H.M. The complex cross ratio on the Heisenberg group. *Enseign. Math.* (2) 33, 3-4 (1987), 291–300.

[170] Korányi, A., and Reimann, H.M. Horizontal normal vectors and conformal capacity of spherical rings in the Heisenberg group. *Bull. Sci. Math.* (2) 111, 1 (1987), 3–21.

[171] Korányi, A., and Reimann, H.M. Foundations for the theory of quasiconformal mappings on the Heisenberg group. *Adv. Math.* 111 (1995), 1–87.

[172] Korányi, A., and Vági, S. Singular integrals on homogeneous spaces and some problems of classical analysis. *Ann. Scuola Norm. Sup. Pisa* (3) 25 (1971), 575–648 (1972).

[173] Krantz, S.G. *Function theory of several complex variables.* AMS Chelsea Publishing, Providence, RI, 2001. Reprint of the 1992 edition.

[174] Lanconelli, E., and Morbidelli, D. On the Poincaré inequality for vector fields. *Ark. Mat.* 38, 2 (2000), 327–342.

[175] Laumond, J., Sekhavat, S., and Lamiraux, F. Guidelines in nonholonomic motion planning for mobile robots. In *Robot motion planning and control.* Springer, London, 1998, pp. 1–53.

[176] Leonardi, G., and Rigot, S. Isoperimetric sets on Carnot groups. *Houston J. Math.* 29, 3 (2003), 609–637 (electronic).

[177] Leonardi, G.P., and Masnou, S. On the isoperimetric problem in the Heisenberg group $\mathbb{H}^n$. *Ann. Mat. Pura Appl.* (4) 184, 4 (2005), 533–553.

[178] Liu, W.-S., and Sussmann, H. *Abnormal sub-Riemannian minimizers.* No. 152 in Lecture Notes in Pure and Applied Mathematics. Dekker, New York, 1994, pp. 705–716.

[179] Liu, Y., Lu, G., and Wheeden, R. Some equivalent definitions of high order Sobolev spaces on stratified groups and generalizations to metric spaces. *Math. Ann.* 323, 1 (2002), 157–174.

[180] Lu, G. Weighted Poincaré and Sobolev inequalities for vector fields satisfying Hörmander's condition and applications. *Rev. Mat. Iberoamericana* 8, 3 (1992), 367–439.

[181] Lu, G. The sharp Poincaré inequality for free vector fields: an endpoint result. *Rev. Mat. Iberoamericana* 10, 2 (1994), 453–466.

[182] Lu, G. Potential analysis on Carnot groups. II. Relationship between Hausdorff measures and capacities. *Acta Math. Sin.* (*Engl. Ser.*) 20, 1 (2004), 25–46.

[183] Lu, G., and Wheeden, R. An optimal representation formula for Carnot-Carathéodory vector fields. *Bull. London Math. Soc.* 30, 6 (1998), 578–584.

[184] Lu, G., and Wheeden, R. Poincaré inequalities, isoperimetric estimates, and representation formulas on product spaces. *Indiana Univ. Math. J.* 47, 1 (1998), 123–151.

[185] Lu, G., and Wheeden, R. High order representation formulas and embedding theorems on stratified groups and generalizations. *Studia Math.* 142, 2 (2000), 101–133.

[186] Lu, G., and Wheeden, R. Simultaneous representation and approximation formulas and high-order Sobolev embedding theorems on stratified groups. *Constr. Approx.* 20, 4 (2004), 647–668.

[187] Magnani, V. Characteristic points, rectifiability and perimeter measure on stratified groups. Preprint, 2004.

[188] Magnani, V. Measure of non-horizontal submanifolds. Preprint, 2005.

[189] Magnani, V. Differentiability and area formula on stratified Lie groups. *Houston J. Math.* 27, 2 (2001), 297–323.

[190] Magnani, V. *Elements of geometric measure theory on sub-Riemannian groups.* PhD thesis, Scuola Normale Superiore, University of Pisa, 2002.

[191] Magnani, V. On a general coarea inequality and applications. *Ann. Acad. Sci. Fenn. Math.* 27, 1 (2002), 121–140.

[192] Magnani, V. Note on coarea formulae in the Heisenberg group. *Publ. Mat.* 48, 2 (2004), 409–422.

[193] Manfredi, J., and Vera de Serio, V. Rearrangements in Carnot groups. Preprint, 2000.

[194] Margulis, G. The isometry of closed manifolds of constant negative curvature with the same fundamental group. *Dokl. Akad. Nauk. SSSR* 192 (1970), 736–737.

[195] Massari, U., and Miranda, M. *Minimal surfaces of codimension one.* No. 91 in North-Holland Mathematics Studies. North-Holland Publishing Co., 1984.

[196] Mattila, P. *Geometry of sets and measures in Euclidean spaces*, vol. 44 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1995.

[197] Maz'ja, V. *Sobolev spaces. Translated from the Russian by T.O. Shaposhnikova.* Springer Series in Soviet Mathematics. Springer-Verlag, 1985.

[198] Montanari, A., and Morbidelli, D. Balls defined by nonsmooth vector fields and the Poincaré inequality. *Ann. Inst. Fourier* (*Grenoble*) 54, 2 (2004), 431–452.

[199] Montefalcone, F. PhD thesis, University of Bologna, 2004.

[200] Montefalcone, F. Some relations among volume, intrinsic perimeter and one-dimensional restrictions of BV functions in Carnot groups. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) 4, 1 (2005), 79–128.

[201] Montgomery, R. Abnormal minimizers. *SIAM J. Control Optim.* 32, 6 (1994), 1605–1620.

[202] Montgomery, R. *Survey of singular geodesics.* In [32], 1996, pp. 325–339.

[203] Montgomery, R. *A tour of subriemannian geometries, their geodesics and applications.* No. 91 in Mathematical Surveys and Monographs. American Mathematical Society, 2002.

[204] Monti, R. Some properties of Carnot-Carathéodory balls in the Heisenberg group. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei* (9) *Mat. Appl.* 11, 3 (2000), 155–167 (2001).

[205] Monti, R. Brunn-Minkowski and isoperimetric inequality in the Heisenberg group. *Ann. Acad. Sci. Fenn. Math.* 28, 1 (2003), 99–109.

[206] Monti, R. Sobolev inequalities for weighted gradients. *Comm. Partial Differential Equations* 31, 10 (2006), 1479–1504.

[207] Monti, R., and Morbidelli, D. Isoperimetric inequality in the Grushin plane. *J. Geom. Anal.* 14, 2 (2004), 355–368.

[208] Monti, R., and Morbidelli, D. Non-tangentially accessible domains for vector fields. *Indiana Univ. Math. J.* 54, 2 (2005), 473–498.

[209] Monti, R., and Morbidelli, D. Regular domains in homogeneous groups. *Trans. Amer. Math. Soc.* 357, 8 (2005), 2975–3011 (electronic).

[210] Monti, R., and Rickly, M. Convex isoperimetric sets in the Heisenberg group. Preprint, 2006.

[211] Monti, R., and Serra-Cassano, F. Surface measures in Carnot-Carathéodory spaces. *Calc. Var. Partial Differential Equations* 13, 3 (2001), 339–376.

[212] Mostow, G.D. *Strong rigidity of locally symmetric spaces.* Princeton University Press, Princeton, N.J., 1973. Annals of Mathematics Studies, No. 78.

[213] Müller, D., Peloso, M., and Ricci, F. Eigenvalues of the Hodge Laplacian on the Heisenberg group. In *Proceedings of the El Escorial Conference* (2004).

[214] Ni, Y. Sub-Riemannian constant mean curvature surfaces in the Heisenberg group as limits. *Ann. Mat. Pura Appl.* (4) 183, 4 (2004), 555–570.

[215] Niu, P., Zhang, H., and Wang, Y. Hardy type and Rellich type inequalities on the Heisenberg group. *Proc. Amer. Math. Soc.* 129, 12 (2001), 3623–3630 (electronic).

[216] Osserman, R. The isoperimetric inequality. *Bull. Amer. Math. Soc.* 84, 6 (1978), 1182–1238.

[217] Pansu, P. Une inégalité isopérimétrique sur le groupe de Heisenberg. *C. R. Acad. Sci. Paris Sér. I Math.* 295, 2 (1982), 127–130.

[218] Pansu, P. Croissance des boules et des geodésiques fermées dans les nilvariétès. *Ergodic Theory Dynam Systems* 3 (1983).

[219] Pansu, P. An isoperimetric inequality on the Heisenberg group. *Rend. Sem. Mat. Univ. Politec. Torino*, Special Issue (1983), 159–174 (1984). Conference on differential geometry on homogeneous spaces (Turin, 1983).

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

[221] Pauls, S. Minimal surfaces in the Heisenberg group. *Geom. Dedicata* 104 (2004), 201–231.

[222] Pauls, S. A notion of rectifiability modeled on Carnot groups. *Indiana J. Math.* 53, 1 (2004), 49–81.

[223] Pauls, S. H-minimal graphs of low regularity in the Heisenberg group. *Comm. Math. Helv.* 81 (2006), 337–381.

[224] Peloso, M., and Ricci, F. Analysis of the Kohn Laplacian on quadratic CR manifolds. *J. Funct. Anal.* 203, 2 (2003), 321–355.

[225] Perelman, G. The entropy formula for the Ricci flow and its geometric applications. Preprint, 2002.

[226] Perelman, G. Ricci flow with surgery on 3-manifolds,. Preprint, 2003.

[227] Petitot, J. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *J. Physiology* 97 (2003), 265–309.

[228] Petitot, J., and Tondut, Y. Vers une neuro-géometrie. Fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques, Informatique et Sciences Humaine, EHESS, Paris* 145 (1998), 5–101.

[229] Pólya, G., and Szegö, G. *Isoperimetric Inequalities in Mathematical Physics.* Annals of Mathematics Studies no. 27. Princeton University Press, Princeton, N.J., 1951.

[230] Ritoré, M. A proof by calibration of an isoperimetric inequality in the Heisenberg group $\mathbb{H}^n$. Preprint, 2006.

[231] Ritoré, M., and Rosales, C. Area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$. Preprint, 2005.

[232] Ritoré, M., and Rosales, C. Rotationally invariant hypersurfaces with constant mean curvature in the Heisenberg group $\mathbb{H}^n$. *J. Geom. Anal.* 16, 4 (2006), 703–720.

[233] Rothschild, L., and Stein, E. Hypoelliptic differential operators and nilpotent groups. *Acta Math.* 137, 3-4 (1976), 247–320.

[234] Rumin, M. Formes diff/'erentielles sur les vari/'et/'es de contact. *Jour. Diff. Geom.* 39, 2 (1994), 281–330.

[235] Saloff-Coste, L. Théorèmes de Sobolev et inégalités de Trudinger sur certain groupes de Lie. *C. R. Acad. Sci. Paris* 306 (1988), 305–308.

[236] Samonds, J., Shou, Z., Bernard, M., and Bonds, A. Synchronous activity in cat visual cortex encodes collinear and cocircular contours. *J. Neurophysiol* 95 (2006), 2602–2616.

[237] Santaló, L.A. *Integral geometry and geometric probability*, vol. 1 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976. With a foreword by Mark Kac.

[238] Sarti, A., Citti, G., and Manfredini, M. From neural oscillations to variational problems in the visual cortex. *Journal of Physiology* 97 (2003), 379–385.

[239] Selby, C. *Geometry of hypersurfaces in Carnot groups of step two*. PhD thesis, Purdue University, 2006.

[240] Semmes, S. Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities. *Selecta Math.* 2 (1996), 155–295.

[241] Semmes, S. An introduction to Heisenberg groups in analysis and geometry. *Notices Amer. Math. Soc.* 50, 6 (2003), 640–646.

[242] Sontag, E. *Mathematical control theory*, second ed., vol. 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1998.

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

[244] Sussman, H. *Geometry and optimal control*. Springer, 1999, pp. 140–198.

[245] Talenti, G. Best constant in Sobolev inequality. *Ann. Mat. Pura. Appl.* 110 (1976), 353–372.

[246] Tang, P. Quasiconformal homeomorphisms on CR 3-manifolds with symmetries. *Math. Z.* 219 (1995), 49–69.

[247] Tomter, P. Constant mean curvature surfaces in the Heisenberg group. In *Differential geometry: partial differential equations on manifolds* (*Los Angeles, CA,* 1990). Amer. Math. Soc., Providence, RI, 1993, pp. 485–495.

[248] Troutman, J. *Variational Calculus with Elementary Convexity.* Springer-Verlag, 1983.

[249] Trudinger, N. On imbeddings into Orlicz spaces and some applications. *J. Math. Mech.* 17 (1967), 473–483.

[250] Tukia, P., and Väisälä, J. Quasisymmetric embeddings of metric spaces. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 5 (1980), 97–114.

[251] Tyson, J. Quasiconformality and quasisymmetry in metric measure spaces. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 23 (1998), 525–548.

[252] Tyson, J. Metric and geometric quasiconformality in Ahlfors regular Loewner spaces. *Conf. Geom. Dynam.* 5 (2001), 21–73.

[253] Tyson, J. Sharp weighted Young's inequalities and Moser–Trudinger inequalities in groups of Heisenberg type and Grushin spaces. *Pot. Anal.* 24 (2006), 357–384.

[254] Varopoulos, N. Fonctions harmoniques sur les groupes de Lie. *C. R. Acad. Sci. Paris Sér. I Math.* 304, 17 (1987), 519–521.

[255] Varopoulos, N., Saloff-Coste, L., and Coulhon, T. *Analysis and geometry on groups.* No. 100 in Cambridge Tracts in Mathematics. Cambridge University Press, 1992.

[256] Warhurst, B. Contact and quasiconformal mappings on real model filiform groups. *Bull. Austral. Math. Soc.* 68, 2 (2003), 329–343.

[257] Warhurst, B. Jet spaces as nonrigid Carnot groups. *J. Lie Theory* 15, 1 (2005), 341–356.

[258] Yu, H., Farley, B., Jin, D., and Sur, M. The coordinated mapping of visual space and response features in visual cortex. *Neuron* 47 (2005), 267–280.

[259] Ziemer, W. *Weakly Differentiable Functions*, vol. 120 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1989.

Index