UNIVERSITY OF CALIFORNIA
UNIVERSITY OF CALIFORNIA SANTA CRUZ

TWO PROBLEMS IN SUB-RIEMANNIAN GEOMETRY

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

by

Corey R. Shanbrom

June 2013

The Dissertation of Corey R. Shanbrom
is approved:

Professor Richard Montgomery, Chair

Professor Anthony Tromba

Professor Jie Qing

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Two Problems in Sub-Riemannian Geometry

by

Corey R. Shanbrom

In this thesis we study two interesting problems in sub-Riemannian geometry. First, we pose and partially solve the Kepler Problem on the Heisenberg Group. Second, we present a formula for computing the Puiseux characteristic corresponding to a Goursat germ with prescribed small growth vector.

The Kepler Problem is among the oldest and most fundamental problems in mechanics. It has been studied in curved geometries, such as the sphere and hyperbolic plane. Here, we formulate the problem on the Heisenberg group, the simplest sub-Riemannian manifold. We take the sub-Riemannian Hamiltonian as our kinetic energy, and our potential is the fundamental solution to the Heisenberg sub-Laplacian. We record many interesting properties of the system, prove the existence of periodic orbits, deduce a version of Kepler's third law, and reduce the integration of a fundamental integrable subsystem to the parametrization of a family of algebraic plane curves.

Germs of Goursat distributions can be classified according to a geometric coding called an RVT code. Jean and Mormul have shown that this coding carries precisely the same data as the small growth vector. Montgomery and Zhitomirskii have shown that such germs correspond to finite jets of Legendrian curve germs, and that the RVT coding corresponds to the classical invariant in the singularity theory of planar curves: the Puiseux characteristic. Here we derive a simple formula for the Puiseux characteristic of the curve corresponding to a Goursat germ with given small growth vector.

To the memory of

Edward Shanbrom

Acknowledgments

I would like to express my deepest gratitude to my advisor Richard Montgomery for his patience, guidance, and kindness throughout the preparation of this thesis. Richard introduced me to entire worlds of beautiful mathematics, helped immensely in formulating and approaching new problems, and has fully supported my transformation into a mathematician. I have learned so much from him that I cannot hope to express here the extent to which he influenced my academic development. I am very lucky to have been his student.

I am also sincerely indebted to my committee members, Debra Lewis, Jie Qing, Tony Tromba, and Marty Weissman, for their advice and support. Each has greatly enriched my education both inside and outside the classroom. The entire mathematics faculty at UC Santa Cruz deserves my thanks, as I have enjoyed so many fantastic teachers over my long stay here. In particular, Richard Mitchell was profoundly inspirational during my undergraduate career, and remains a great mentor and friend. My high school teacher Ken Umholtz is also deserving of much credit for steering a young and restless mind towards the beauty and quiet peace of mathematics.

Further, I am very grateful to the UC Santa Cruz mathematics graduate students, past and present, who have helped me succeed both academically and socially. Shawn O'Hare, Wyatt Howard, and Rob Laber have been especially helpful over the years and merit particular recognition. I am also grateful to Victor Dods for sharing his programming expertise, which helped greatly with the research contained herein.

I also wish to thank all of my family and friends for their unconditional support along this long and perhaps unlikely path. I am especially appreciative of Dena Spatz for her love, trust and the numerous sacrifices she has made for me and my work. Finally, I am ever thankful for the love and support of my parents, Bill and Jill, and my brother Evan.

Part I

The Kepler Problem on the Heisenberg Group

Chapter 1

Introduction

1.1 Overview

In Hamiltonian mechanics, one typically begins with a Riemannian manifold and a choice of potential energy function. The Riemannian metric induces a kinetic energy function on the cotangent bundle of the manifold, which in turn generates the geodesics of the geometry. One chooses a potential function on the manifold, and calls the sum of the potential and kinetic energies the total energy, or Hamiltonian. The flow lines of the induced Hamiltonian vector field on the cotangent bundle correspond precisely to the trajectories of particles under the influence of the affiliated mechanical system.

Here, we study one such mechanical system, albeit with one caveat. Our setting is not a Riemannian manifold, but instead a sub-Riemannian manifold: the three-dimensional Heisenberg Group. Thankfully, we still have a natural choice of kinetic energy, induced by the metric, which indeed generates the sub-Riemannian geodesics. We choose our potential function so as to model the classical Kepler Problem: that of determining the motion of a planet around a fixed sun subject only to the force of gravity.

Thus, the primary question we seek to answer is the following: *What would the orbit of a planet around a sun look like in Heisenberg geometry?* Naturally, we are interested in other properties of this system as well, and many of the these properties are recorded below. Our primary result, Theorem 2, asserts that closed orbits do indeed exist; approximations of such orbits are depicted in Figures 3.4 and 3.5. We also show that these periodic orbits must lie on the zero energy hypersurface, and that the problem is integrable on this hypersurface. Moreover, we prove a version of Kepler's third law for these periodic orbits.

More specifically, we choose our (gravitational) potential to be the fundamental solution to the Heisenberg sub-Laplacian. The delta function source, acting as our sun, lies at the origin. (The Heisenberg group is topologically a vector space, whose origin corresponds to the group identity.) The explicit form of this function was discovered by Folland in 1970 ([12]).

1.2 Motivation

Our motivations for studying this problem are multiple. Originally, we sought to study the question: *Can we do sub-Riemannian mechanics?* That is, we were curious whether classical mechanics could be formulated on a sub-Riemannian, rather than a Riemannian, manifold. While this is a broad question which we hope to pursue further in future work, it seemed natural to start with the simplest sub-Riemannian geometry: the Heisenberg group. The Kepler Problem was chosen as it is simple to formulate and interpret but retains many of the most fundamental aspects of mechanical systems in the large. In other words, the classical Kepler Problem is very well understood but simultaneously non-trivial and subtle.

This line of thinking fits with the historical study of the problem. Isaac Newton studied the Euclidean Kepler Problem in the 17th century and derived Kepler's three laws of planetary motion. But the problem was posed on spaces of constant curvature much later. In 1835, Lobachevsky ([19]) posed the Kepler Problem in three-dimensional hyperbolic space. Bolyai did similar work (independently) in the same time period. Paul Joseph Serret posed and solved the Kepler Problem on the two-sphere in 1860. Schering, Lipschitz, Killing, and Liebmann studied the Kepler Problem on hyperbolic and spherical three-space between 1870 and 1902. For more information, and the relevant references, see Florin Diacu's wonderful paper [11].

With this historical background in mind, it seems natural to continue efforts to pose and solve the Kepler Problem in more general geometries (sub-Riemannian geometry encompasses the Riemannian sort.) But let us revisit Sir Newton momentarily. While Kepler wrote down his laws based on observational data alone, Newton derived them from more elementary laws. The first law (that planetary orbits are ellipses) and second law ("equal areas in equal times," equivalent to conservation of angular momentum) have been shown to hold in spherical and hyperbolic space. But the third law (the period T of an orbit is related to its size a by the universal relation $a^3 = cT^2$) fails in both.

Why should Kepler's third law fail in these spaces? Where might it hold? The short answer (pursued in more detail in [24]) is that these spaces do not admit dilations. In fact, it follows from Gromov's work in [17] that the only homogeneous Riemannian manifolds which admit dilations are Euclidean spaces (see again [24] for a sketch of the proof). Following Galileo we will insist that our spaces are homogeneous, so that the location of the sun does not matter. Then we conclude that to regain Kepler's third law in a non-Euclidean geometry, we must leave the world of Riemannian manifolds. Indeed, the non-Euclidean spaces admitting dilations are Carnot groups. The simplest of these is the Heisenberg group, and we derive a version of Kepler's third law here.

Chapter 2

The Problem

2.1 The Setting

2.1.1 Sub-Riemannian Geometry

A *sub-Riemannian geometry* is a triple $(M, D, \langle\cdot,\cdot\rangle)$ where M is a smooth manifold, D is a distribution (subbundle of the tangent bundle), and $\langle\cdot,\cdot\rangle$ is a fiber inner product on D. We call curves and vector fields *horizontal* if they are tangent to D.

This structure induces a distance function on M in the usual way, where we restrict our attention to horizontal curves. If $\gamma\colon [a,b] \to M$ is a horizontal curve, define its *length* by

$$l(\gamma) = \int_a^b \sqrt{\langle\dot\gamma(t), \dot\gamma(t)\rangle}dt.$$

We can then define the *distance* between the points $p, q \in M$ to be

$$d(p,q) = \inf l(\gamma),$$

where the infimum is taken over all horizontal curves connecting p to q. A horizontal curve γ is called a *geodesic* if it realizes this distance; that is, if

$$l(\gamma) = d(\gamma(a), \gamma(b)).$$

It may so happen that there are no horizontal curves connecting p to q. If this is the case, then we say the distance between the points is infinite, and d fails to be a genuine metric.

However, this is not a problem for sub-Riemannian geometries whose distributions are of the following type.

Definition 1 *A distribution D is called* non-holonomic *or* bracket-generating *if D Lie-generates TM.*

In other words, if one takes a local frame of vector fields for D, then sufficiently many Lie brackets will generate the entire tangent bundle. One can check that this definition does not depend on the choice of frame. Note that bracket-generating distributions are also known as those satisfying *Hörmander's condition*, as he proved that their induced sub-Laplacians (see Section 2.2.1) are hypoelliptic.

The main theorem on these distributions is the Chow-Rashevskii theorem, which says for M connected, D bracket-generating, our manifold is horizontally path-connected.

Theorem 1 (Rashevskii, 1938; Chow, 1939) *If D is bracket-generating, then there exists a horizontal path connecting $p \in M$ to any point in the same component.*

Thus, in this setting, the distance function d is a genuine metric, known as the *Carnot-Carathéodory metric.* Note that, in some sense, bracket-generating distributions are antithetical to involutive ones. By the Frobenius theorem, in an involutive distribution, only the points on the same leaf can be connected by a horizontal path.

Finally, every sub-Riemannian geometry comes equipped with an operator called the *sub-Riemannian gradient.* If f is a smooth function on M, we define the horizontal field $\nabla_{sR} f$ much like the Riemannian version. For a horizontal vector field V, we put

$$\langle \nabla_{sR} f, V \rangle = df(V).$$

For further background, see the usual sub-Riemannian references [5, 21, 31].

2.1.2 The Heisenberg Group

The Heisenberg group is a well-studied and ubiquitous object in mathematics, with many discrete and higher-dimensional incarnations. Here we will content ourselves with the smooth, three-dimensional version, although we will exploit its Lie group, contact, and sub-Riemannian structures.

Lie group structure

Definition 2 *The* Heisenberg algebra *is the three-dimensional real Lie algebra*

$$\mathfrak{h} = \langle X, Y, Z \mid [X, Y] = Z,\ [X, Z] = [Y, Z] = 0 \rangle.$$

The Heisenberg algebra is so named because one can consider X and Y as self-adjoint operators measuring the position and momentum of a particle, respectively, with Z a multiple of the identity. The failure of X and Y to commute is one version of the famous uncertainty principle.

Since $\mathfrak{h}$ is nilpotent, the exponential map $\exp\colon \mathfrak{h} \to \mathbb{H}$ maps $\mathfrak{h}$ diffeomorphically onto its simply connected Lie group. This $\mathbb{H}$ will turn out to be the Heisenberg group. The exponential map endows $\mathbb{H}$ with global coordinates (x, y, z). According to the Baker-Campbell-Hausdorff formula, in these coordinates, the group law will be polynomial, with degree equal to the degree of nilpotency of $\mathfrak{h}$, which happens to be 2 in this case.

Definition 3 *The* Heisenberg group *is the Lie group $\mathbb{H}$ which is diffeomorphic to $\mathbb{R}^3$, and whose group law is*

$$(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + \tfrac{1}{2}(x_1 y_2 - x_2 y_1)).$$

Thus, the Heisenberg group is topologically a vector space endowed with global coordinates (x, y, z). In these coordinates, the origin $(0, 0, 0)$ is indeed the group identity element.

In the literature both the Lie algebra and Lie group are often realized as matrices. We have

$$\mathfrak{h} = \left\{ \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} \right\},$$

where the Lie bracket is just the matrix commutator. We also have

$$\mathbb{H} \cong \left\{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \right\},$$

where the product is matrix multiplication and the Lie group isomorphism is given by $z \mapsto z + \frac{1}{2}xy$. Here, a, b, c, x, y, z are all real numbers.

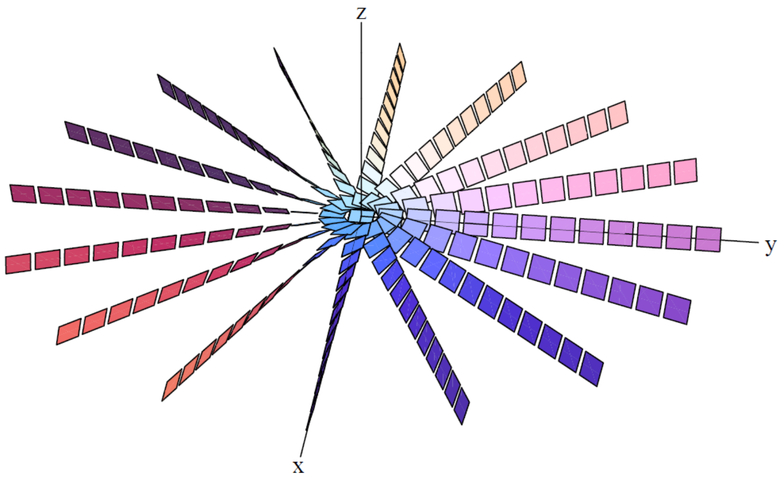

Figure 2.1: The Heisenberg distribution

Sub-Riemannian Structure

So far, we know that the Heisenberg group as a manifold is just $\mathbb{R}^3$. We now give it a distribution and fiber inner product.

By construction, X, Y, and Z are a basis for $\mathfrak{h} = T_0\mathbb{H}$. But we can push these vectors forward via left multiplication in the group and think of X, Y, and Z instead as left invariant vector fields which frame the tangent bundle. In coordinates (which we will exploit liberally in the sequel), we have

$$X = \frac{\partial}{\partial x} - \frac{1}{2}y\frac{\partial}{\partial z}, \quad Y = \frac{\partial}{\partial y} + \frac{1}{2}x\frac{\partial}{\partial z}, \quad Z = \frac{\partial}{\partial z}.$$

Now we define our 2-plane distribution to be

$$D = \operatorname{span}\{X, Y\}.$$

Then D is clearly bracket-generating (it only takes one Lie bracket to generate the missing direction). A picture of some of these distribution planes appears as Figure 2.1 – other planes are obtained by vertically translating those shown.

But D is also contact (it is in fact the canonical contact structure on $\mathbb{R}^3$). To see this, we can realize D as the kernel of the one-form

$$\Theta = dz - \tfrac{1}{2}(xdy - ydx).$$

Then the contact condition is that

$$\Theta \wedge d\Theta = -dx \wedge dy \wedge dz$$

is non-zero. It is evidently the opposite of the Lebesgue volume form on $\mathbb{R}^3$, which is certainly not zero. Moreover, this shows that $\mathbb{H}$ is endowed with a canonical volume form, a property which most sub-Riemannian manifolds do not enjoy. We will exploit this fact later. Finally, note that a curve $\gamma(t) = (x(t), y(t), z(t))$ is horizontal if and only if it satisfies

$$\dot{z} = \tfrac{1}{2}(x\dot{y} - y\dot{x}).$$

Next, we define the fiber inner product. Let $p \in \mathbb{H}$ and let $v, w \in D_p$ have coordinate expressions $v = v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} + v_3 \frac{\partial}{\partial z}$ and $w = w_1 \frac{\partial}{\partial x} + w_2 \frac{\partial}{\partial y} + w_3 \frac{\partial}{\partial z}$. Then we put

$$\langle v, w \rangle_p = v_1 w_1 + v_2 w_2.$$

In other words, the sub-Riemannian inner product is just the usual Euclidean dot product on $\mathbb{R}^3$ with the third component ignored. This is the right choice for two reasons. First, our frame X, Y is now orthonormal. Second, left multiplication in the group is now an isometry. Note that the length of a curve is equal to the length of its projection to the xy-plane.

Since D is bracket-generating and $\mathbb{H}$ is connected, the Chow-Rashevskii theorem guarantees that $\mathbb{H}$ is horizontally path-connected and we do have a genuine metric

$$ds^2_{\mathbb{H}} = dx^2 + dy^2.$$

In the following we will denote the Heisenberg distance function by d:

$$d(p, q) = \text{the sub-Riemannian distance between points } p \text{ and } q.$$

There is no known explicit form of this function. We will denote by $|| \cdot ||_{sR}$ the sub-Riemannian distance to the origin, and $|| \cdot ||_{\mathbb{H}}$ will denote the Heisenberg norm of a horizontal vector:

$$\begin{aligned} ||p||_{sR} &= d(p, 0), \qquad && p \in \mathbb{H} \\ ||v||_{\mathbb{H}} &= \sqrt{\langle v, v \rangle}, \qquad && v \in D_p. \end{aligned}$$

The sub-Riemannian gradient of a function $f\colon \mathbb{H} \to \mathbb{R}$ is the horizontal vector field

$$\nabla_{sR} f = X(f)X + Y(f)Y.$$

In the standard $\mathbb{R}^3$ basis, this takes the form

$$\nabla_{sR} f = X(f)\frac{\partial}{\partial x} + Y(f)\frac{\partial}{\partial y} + \tfrac{1}{2}(xY(f) - yX(f))\frac{\partial}{\partial z}.$$

We now briefly investigate the induced geometry. Consider $T^*\mathbb{H}$ with canonical coordinates (x, y, z, p_x, p_y, p_z). Then

$$P_X = p_x - \tfrac{1}{2} y p_z \quad \text{and} \quad P_Y = p_y + \tfrac{1}{2} x p_z$$

are dual momenta to X and Y, respectively. Let

$$K = \tfrac{1}{2}(P_X^2 + P_Y^2).$$

The function $K\colon T^*\mathbb{H} \to \mathbb{R}$ is known as the sub-Riemannian Hamiltonian, as the flow lines of its Hamiltonian vector field (symplectic gradient) are precisely the geodesics in $\mathbb{H}$. As these geodesics can be thought of as trajectories of free particles, we will call the function K our *kinetic energy*. While it is written here in coordinates, it can be defined canonically in terms of the cometric.

Proposition 1 *Heisenberg geodesics are "helices."*

The quotation marks are included since these curves are only qualitatively helices, meaning they project to circles (we consider line segments degenerate circles). The z-coordinate does not grow linearly in the angle, but is instead given by the area traced out by the projection of the curve to the xy-plane. See the proof sketched below and Figure 2.2.

Proof (Sketch) Fix two points. We may assume that one point is the origin since left multiplication in the group is a transitive isometric action. Suppose the other is $q = (x_1, y_1, z_1)$. Suppose γ is a horizontal path connecting $(0, 0, 0)$ to q, and let $\tilde{\gamma}$ denote its projection to the xy-plane. Then we want to minimize $l(\gamma) = l(\tilde{\gamma})$. But

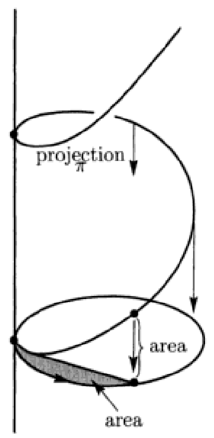

Figure 2.2: A Heisenberg geodesic

we compute

$$\begin{aligned} z_1 &= z_1 - z_0 \\ &= \int_{\tilde{\gamma}} dz \\ &= \int_{\tilde{\gamma}} \tfrac{1}{2}(xdy - ydx) \qquad \text{(horizontal condition)} \\ &= \text{area inside } \tilde{\gamma}. \qquad \text{(Stokes' Theorem)} \end{aligned}$$

Thus, the problem is to minimize the length of a planar curve given a fixed area enclosed by the curve and the line segment connecting its endpoints. This is the dual to the isoperimetric problem (Dido's problem), and solutions are known to be arcs of circles (with line segments as a degenerate case). Thus, $\tilde{\gamma}$ traces out an arc of a circle, and the z-coordinate of γ grows like the area traced out by this arc. $\square$

Note that the computation above shows that the z-coordinate of *any* horizontal curve must grow like the area traced out by its projection to the xy-plane.

To end this section, we observe that the Heisenberg group (like any Carnot group) admits dilations. For any positive real number λ, define the map

$$\begin{aligned} \delta_\lambda \colon \mathbb{H} &\to \mathbb{H} \\ (x, y, z) &\mapsto (\lambda x, \lambda y, \lambda^2 z). \end{aligned}$$

To say that this map is a dilation is to say that

$$d(\delta_\lambda(p), \delta_\lambda(q)) = \lambda d(p,q), \quad \forall\ p,q \in \mathbb{H}.$$

This map lifts to a map on the cotangent bundle, which we also denote by δ_λ:

$$(x,y,z,p_x,p_y,p_z) \mapsto (\lambda x, \lambda y, \lambda^2 z, \lambda^{-1}p_x, \lambda^{-1}p_y, \lambda^{-2}p_z).$$

Observe that this dilation on $T^*\mathbb{H}$ is generated by the function

$$J = xp_x + yp_y + 2zp_z.$$

2.2 The System

2.2.1 The Potential Energy

Recall the classical Kepler Problem on $\mathbb{R}^3$. Let $r = \sqrt{x^2+y^2+z^2}$. Then the Hamiltonian is

$$H = \tfrac{1}{2}(p_x^2+p_y^2+p_z^2) - \frac{k}{r}.$$

How do we characterize the potential $U = -\frac{k}{r}$? The usual answer is that U is (a constant times) the inverse of the distance function. However, this characterization fails to provide guidance when we attempt to study the problem on spaces without an explicit distance function, such as the Heisenberg group. A better answer is that, when $k = \frac{1}{4\pi}$, U is the fundamental solution to the Laplacian on $\mathbb{R}^3$ (see [2]). In other words, U satisfies $\Delta U = \delta_0$, where δ_0 is the Dirac delta function with source at 0.

Now consider the vector fields X and Y which form an orthonormal frame for the Heisenberg distribution. Thinking of these as first-order differential operators, we define the Heisenberg sub-Laplacian to be the second-order operator

$$\Delta_{\mathbb{H}} = X^2 + Y^2.$$

A different choice of orthonormal frame will yield a different sub-Laplacian in general. However, as $\mathbb{H}$ is equipped with a canonical volume form, we can resolve the ambiguity using integration by parts as in Riemannian geometry: we define $\Delta_{\mathbb{H}}$ by the formula

$$-\int f(\Delta_{\mathbb{H}} g) d\text{vol} = \int \langle \nabla_{sR} f, \nabla_{sR} g \rangle d\text{vol}.$$

Here, f and g are smooth functions with compact support, $d\mathrm{vol}$ denotes the canonical volume form (negative Lebesgue measure) described in Section 2.1.2, and ∇_{sR} denotes the sub-Riemannian gradient. One checks that the sub-Laplacian defined above is indeed correct.

In [12], Folland found the fundamental solution to the Heisenberg sub-Laplacian $\Delta_{\mathbb{H}}$:

$$U = -\alpha\left((x^2+y^2)^2 + \tfrac{1}{16}z^2\right)^{-\frac{1}{2}}.$$

Here, $\alpha = 2/\pi$; for our purposes it suffices to leave α as a positive constant. This constant is given by an integral in Folland's work and is computed in Section 3.1.2. We recognize that this potential has a singularity at the origin $(x, y, z) = (0, 0, 0)$, but is smooth away from this point. The singularity corresponds physically to our planet crashing into the sun, at which point the planet's potential energy is $-\infty$ and its kinetic energy is $+\infty$. For this reason we will refer to a trajectory passing through the origin as a *collision*.

For notational purposes, we set

$$\mu = (x^2+y^2)^2 + \tfrac{1}{16}z^2.$$

Note that Folland uses the notation $\rho = \mu^{1/4}$, which is homogeneous of degree 1 with respect to the dilation δ_λ defined in Section 2.1.2; we will use ρ later as a norm. Then we can write

$$U = -\alpha\mu^{-1/2} = -\alpha\rho^{-2}.$$

2.2.2 Hamiltonian Formalism

We define our Hamiltonian H in the usual way, as the sum of the kinetic and potential energies: $H = K + U$. For this reason, we will often refer to H as the *total energy* or simply the *energy*. As our Hamiltonian has a singularity at the origin (from the potential U), this function is defined on the cotangent bundle of the Heisenberg group with the origin deleted:

$$H\colon T^*(\mathbb{H} - \{(0,0,0)\}) \to \mathbb{R}.$$

Explicitly, we have

$$H = \tfrac{1}{2}(p_x - \tfrac{1}{2}yp_z)^2 + \tfrac{1}{2}(p_y + \tfrac{1}{2}xp_z)^2 - \tfrac{2}{\pi}\left((x^2+y^2)^2 + \tfrac{1}{16}z^2\right)^{-\frac{1}{2}}. \qquad (2.1)$$

Using the notation developed above, we can rewrite this equation in the less intimidating form

$$H = \tfrac{1}{2}(P_X^2 + P_Y^2) - \alpha\mu^{-\frac{1}{2}}. \tag{2.2}$$

Our goal is to investigate the flow of the induced Hamiltonian vector field on the cotangent bundle $T^*(\mathbb{H} - \{(0,0,0)\})$, called *phase space*, of the Heisenberg group with deleted origin $\mathbb{H} - \{(0,0,0)\}$, called *configuration space*.

Our first task is writing down Hamilton's equations, or the *equations of motion*:

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}. \end{aligned}$$

Here, q_i and p_i denote the ith *position* and *momentum* coordinate, respectively. That is, $(q_1, q_2, q_3, p_1, p_2, p_3) = (x, y, z, p_x, p_y, p_z) \in T^*\mathbb{H}$. We will attempt to stick with the latter notation, but will on occasion make use of the convenient indexing provided by the former.

For our Hamiltonian (2.1), these equations take the form:

$$\begin{aligned} \dot{x} &= P_X \\ \dot{y} &= P_Y \\ \dot{z} &= \frac{1}{2}xP_Y - \frac{1}{2}yP_X \end{aligned}$$

$$\begin{aligned} \dot{p}_x &= -\frac{1}{2}P_Y p_z - 2\alpha x(x^2+y^2)\mu^{-3/2} \\ \dot{p}_y &= \frac{1}{2}P_X p_z - 2\alpha y(x^2+y^2)\mu^{-3/2} \\ \dot{p}_z &= -\frac{\alpha}{16}z\mu^{-3/2}. \end{aligned}$$

Our goal, in its strongest form, is to solve this six-dimensional autonomous nonlinear system of ordinary differential equations.

For future reference, we record the second derivatives of the position coordinates:

$$\ddot{x} = \dot{P}_X = -P_Y p_z + \big(\tfrac{\alpha}{32} yz - 2\alpha x(x^2+y^2)\big)\mu^{-3/2}$$
$$\ddot{y} = \dot{P}_Y = P_X p_z - \big(\tfrac{\alpha}{32} xz + 2\alpha y(x^2+y^2)\big)\mu^{-3/2}$$
$$\ddot{z} = \tfrac{1}{2} p_z(xp_x + yp_y) - \tfrac{\alpha}{64}(x^2+y^2)z\mu^{-3/2}.$$

Polar Coordinates

Throughout this document, we will often make use of polar (cylindrical) coordinates (r, θ, z). We will now determine the appropriate conjugate momenta, then write down the Hamiltonian and equations of motion.

Define the coordinate transformation

$$\Phi\colon (x,y,z) \mapsto (r,\theta,z) = (\sqrt{x^2+y^2}, \arctan\tfrac{y}{x}, z).$$

We want Φ to be a canonical transformation, so we require $\Phi^*\Lambda = \Lambda$, where Λ denotes the canonical one-form on the cotangent bundle, and Φ^* denotes the pullback of Φ. That is, we require

$$\begin{aligned}\Lambda &= p_x dx + p_y dy + p_z dz \\ &= p_r dr + p_\theta d\theta + p_z dz,\end{aligned}$$

where p_r, p_θ, p_v are our new momenta coordinates. Solving, we define the new momenta by the equations

$$\begin{aligned}p_r &= \frac{xp_x + yp_y}{r} \\ p_\theta &= xp_y - yp_x \\ p_z &= p_z.\end{aligned}$$

We can then write the transformed Hamiltonian (in one of its many incarnations) as

$$H = \tfrac{1}{2}p_r^2 + \tfrac{1}{2}\Big(\frac{p_\theta}{r} + \tfrac{1}{2} r p_z\Big)^2 - \alpha\mu^{-1/2}, \tag{2.3}$$

where, as above, we have written $\mu = r^4 + \frac{1}{16}z^2$. It is worth noting that the Hamiltonian is invariant under the following reflections:

$$z \mapsto -z, \qquad p_r \mapsto -p_r, \qquad (p_z, p_\theta) \mapsto (-p_z, -p_\theta).$$

Finally, we can write the transformed equations of motion:

$$\begin{aligned}
\dot r &= p_r \\
\dot\theta &= \frac{p_\theta}{r^2} + \tfrac{1}{2}p_z \\
\dot z &= \tfrac{1}{2}(p_\theta + \tfrac{1}{2}p_z r^2)
\end{aligned}$$

$$\begin{aligned}
\dot p_r &= \frac{p_\theta^2}{r^3} - \tfrac{1}{4}rp_z^2 - 2\alpha r^3\mu^{-3/2} \\
\dot p_\theta &= 0 \\
\dot p_z &= -\tfrac{\alpha}{16}\left\{r^4 + \tfrac{1}{16}z^2\right\}^{-3/2} z = -\tfrac{\alpha}{16}\mu^{-3/2}z.
\end{aligned}$$

2.2.3 Lagrangian Formalism

In this section we consider our problem as a variational problem with subsidiary constraints. As usual, we define our Lagrangian $L\colon T\mathbb{H} \to \mathbb{R}$ as the difference of the kinetic and potential energies $L = K - U$. Explicitly, we have

$$L(t, q, \dot q) = \tfrac{1}{2}\dot x^2 + \tfrac{1}{2}\dot y^2 + \alpha\Big((x^2+y^2)^2 + \tfrac{1}{16}z^2\Big)^{-1/2}.$$

(Here, as in Section 2.2.2, we write $q = (x, y, z)$.)

Suppose $\gamma\colon (0, T) \to \mathbb{H}$ is an absolutely continuous path. Define the *action* of this path by the functional

$$A(\gamma) = \int_0^T L(t, \gamma, \dot\gamma)dt.$$

The general theory of Lagrangian mechanics says that if γ minimizes A, then γ is a solution to the equations of motion.

Here, we have the additional constraint that our solutions must be horizontal curves for our distribution. This condition can be expressed as the differential constraint

$$\dot z = \tfrac{1}{2}x\dot y - \tfrac{1}{2}y\dot x.$$

That is, solutions must lie on the zero set of the function $G\colon T\mathbb{H} \to \mathbb{R}$ defined by

$$G(t, q, \dot q) = \tfrac{1}{2}x\dot y - \tfrac{1}{2}y\dot x - \dot z.$$

The calculus of variations (see, for example, Section 12 of [13]) tells us that if γ is a minimum of the functional A which also satisfies our constraint, then there exists a scalar $\lambda = \lambda(t)$, commonly known as a *Lagrange multiplier*, such that γ is a minimum of the functional

$$A_\lambda(\gamma) = \int_0^T L_\lambda(t, \gamma, \dot{\gamma}) dt,$$

where we have written $L_\lambda(t, q, \dot{q}) = L(t, q, \dot{q}) - \lambda(t) G(t, q, \dot{q})$. Setting the first variation of A_λ equal to zero and integrating by parts yields the Euler-Lagrange equations

$$\frac{\partial L_\lambda}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L_\lambda}{\partial \dot{q}_i}\right) = 0, \tag{2.4}$$

which we may write more explicitly as

$$\frac{\partial L}{\partial q_i} - \lambda(t)\frac{\partial G}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i} - \lambda(t)\frac{\partial G}{\partial \dot{q}_i}\right) = 0$$

with $i = 1, 2, 3$. We will verify the validity of these statements in Section 3.3.

Finally, we compute the Euler-Lagrange equations (2.4). We first write down the partial derivatives of L,

$$\begin{aligned}
\frac{\partial L}{\partial x} &= -2\alpha x(x^2 + y^2)\mu^{-3/2} \\
\frac{\partial L}{\partial y} &= -2\alpha y(x^2 + y^2)\mu^{-3/2} \\
\frac{\partial L}{\partial z} &= -\tfrac{\alpha}{16} z \mu^{-3/2}
\end{aligned}$$

$$\frac{\partial L}{\partial \dot{x}} = \dot{x}, \qquad \frac{\partial L}{\partial \dot{y}} = \dot{y}, \qquad \frac{\partial L}{\partial \dot{z}} = 0.$$

and the partial derivatives of G,

$$\begin{aligned}
\frac{\partial G}{\partial x} &= \tfrac{1}{2}\dot{y}, & \frac{\partial G}{\partial y} &= -\tfrac{1}{2}\dot{x}, & \frac{\partial G}{\partial z} &= 0 \\
\frac{\partial G}{\partial \dot{x}} &= -\tfrac{1}{2}y, & \frac{\partial G}{\partial \dot{y}} &= \tfrac{1}{2}x, & \frac{\partial G}{\partial \dot{z}} &= -1.
\end{aligned}$$

These give the Euler-Lagrange equations:

$$\begin{aligned}
\ddot{x} &= -\lambda\dot{y} - \tfrac{1}{2}\dot{\lambda}y - 2\alpha x(x^2 + y^2)\mu^{-3/2} \\
\ddot{y} &= \lambda\dot{x} + \tfrac{1}{2}\dot{\lambda}x - 2\alpha y(x^2 + y^2)\mu^{-3/2} \\
\dot{\lambda} &= -\tfrac{\alpha}{16} z \mu^{-3/2}.
\end{aligned}$$

If we take $\lambda = p_z$, then these equations read

$$\ddot{x} = -\dot{y}p_z + \left(\tfrac{\alpha}{32}yz - 2\alpha x(x^2+y^2)\right)\mu^{-3/2} \tag{2.5}$$

$$\ddot{y} = \dot{x}p_z - \left(\tfrac{\alpha}{32}xz + 2\alpha y(x^2+y^2)\right)\mu^{-3/2} \tag{2.6}$$

$$\dot{p}_z = -\tfrac{\alpha}{16}z\mu^{-3/2}. \tag{2.7}$$

Note that these equations agree with the calculations given in Section 2.2.2, so that the Euler-Lagrange equations are indeed equivalent to Hamilton's equations.

Chapter 3

Results

3.1 Dynamics

3.1.1 Integrals and Symmetries

Unfortunately, we do not have many symmetries to work with. Those which we do have have already been described above, albeit only implicitly. Looking at our polar Hamiltonian in Equation (2.3), we recognize the absence of the variable θ. Consequently, the equations of motion are independent of this angle coordinate, so the system enjoys rotational symmetry about the z-axis. As this symmetry can be expressed as the invariance under an action of the compact one-dimensional Lie group S^1, symplectic reduction reduces the dimension of our system by one.

As mentioned shortly after Equation (2.3), we have a few reflectional symmetries as well. But since these correspond to actions of discrete Lie groups, we cannot reduce the dimension of the system. We also have the family of dilations δ_λ which correspond to an action of $\mathbb{R}^+$ (see Section 2.1.2). But, alas, our Hamiltonian is not preserved under this action. Rather, it is homogeneous of degree -2:

$$\delta_\lambda \colon H \mapsto \lambda^{-2} H.$$

By construction, the total energy will be conserved in time. That is, $\dot{H} = 0$. Corresponding to the rotational symmetry (cf. Noether's theorem), we have the conservation of angular momentum p_θ. This was apparent from the polar equations of motion, but we prove it explicitly below for completeness. Also, corresponding to the 'partial' dilational symmetry, we have that the generating function J (see

Section 2.1.2) is an 'almost' integral. That is, J satisfies $\dot{J} = 2H$, so it is an integral of motion for orbits with zero energy. We will see in Section 3.2.1 that $\{H = 0\}$ is a smooth submanifold on which our system is integrable.

Proposition 2 *The angular momentum $p_\theta = xp_y - yp_x$ is an integral of motion.*

Proof Using Hamilton's equations to express the time derivatives of our coordinates, we first observe that

$$\begin{aligned}
\dot{x}p_y &= P_X p_y \\
x\dot{p}_y &= \tfrac{1}{2}xP_X p_z - \alpha\mu^{-3/2}2(x^2+y^2)xy \\
\dot{y}p_x &= P_Y p_x \\
y\dot{p}_x &= -\tfrac{1}{2}yP_Y p_z - \alpha\mu^{-3/2}2(x^2+y^2)xy.
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{d}{dt}(xp_y - yp_x) &= \dot{x}p_y + x\dot{p}_y - \dot{y}p_x - y\dot{p}_x \\
&= P_X p_y + \tfrac{1}{2}xP_X p_z - P_Y p_x + \tfrac{1}{2}yP_Y p_z \\
&= P_X(p_y + \tfrac{1}{2}xp_z) - P_Y(p_x - \tfrac{1}{2}yp_z) \\
&= P_X P_Y - P_Y P_X \\
&= 0.
\end{aligned}$$

□

Proposition 3 *The dilation generating function $J = xp_x + yp_y + 2zp_z$ satisfies $\dot{J} = 2H$.*

Proof As in the previous calculation, we first observe

$$\begin{aligned}
\dot{x}p_x &= P_X p_x \\
x\dot{p}_x &= -\tfrac{1}{2}xP_Y p_z - \alpha\mu^{-3/2}2(x^2+y^2)x^2 \\
\dot{y}p_y &= P_Y p_y \\
y\dot{p}_x &= \tfrac{1}{2}yP_X p_z - \alpha\mu^{-3/2}2(x^2+y^2)y^2 \\
2\dot{z}p_z &= xP_Y p_z - yP_X p_z \\
2z\dot{p}_z &= -\tfrac{1}{8}\alpha\mu^{-3/2}z^2.
\end{aligned}$$

Then we have

$$
\begin{aligned}
\frac{d}{dt}J &= \dot{x}p_x + x\dot{p}_x + \dot{y}p_y + y\dot{p}_y + 2\dot{z}p_z + 2z\dot{p}_z \\
&= P_X(p_x + \tfrac{1}{2}zp_z - yp_z) + P_Y(p_y - \tfrac{1}{2}xp_z + xp_z) \\
&\qquad - \alpha\mu^{-3/2}\left(2(x^2+y^2)x^2 + 2(x^2+y^2)y^2 + \tfrac{1}{8}z^2\right) \\
&= P_X^2 + P_Y^2 - 2\alpha\mu^{-3/2}\left((x^2+y^2)^2 + \tfrac{1}{16}z^2\right) \\
&= P_X^2 + P_Y^2 - 2\alpha\mu^{-1/2} \\
&= 2H.
\end{aligned}
$$

□

3.1.2 Properties

The next two propositions show that negative energy solutions are always bounded, and that periodic orbits necessarily have total energy zero. This latter fact is of particular interest as we will show that periodic orbits do indeed exist (Section 3.3) and that the dynamics restricted to the zero energy submanifold are integrable (Section 3.2).

Proposition 4 *If $H < 0$ then any solution is bounded.*

Proof Suppose $H = -h$ where h is positive. Then $K + U = -h$, so

$$-U = K + h \geq h,$$

since K is always non-negative. But

$$-U = \alpha\left((x^2+y^2)^2 + \tfrac{1}{16}z\right)^{-1/2}.$$

So a solution $(x(t), y(t), z(t))$ in configuration space must satisfy

$$0 \leq \left((x^2+y^2)^2 + \tfrac{1}{16}z\right)^{1/2} \leq \frac{\alpha}{h},$$

where α and h are positive constants. □

Proposition 5 *Periodic orbits must have zero energy.*

Proof If $\gamma(t) = (x(t), y(t), z(t), p_x(t), p_y(t), p_z(t))$ satisfies $\gamma(0) = \gamma(T)$ for some $t = T$, then $J = xp_x + yp_y + 2zp_z$ is also periodic; that is, $J(\gamma(0)) = J(\gamma(T))$. But we know the time derivative of J is constant, given by $\dot{J} = 2H$. Since, by virtue of its periodicity, J cannot be monotonically increasing nor decreasing in time, we must have $\dot{J} = 2H = 0$, so $H = 0$. □

The next two propositions describe the only two families of solutions which we have in explicit form: lines through the origin in the xy-plane, and constant paths lying on the z-axis.

Proposition 6 *The only solutions in the plane $z = 0$ are lines through the origin.*

Proof The equations for $\dot{z}$ and $\dot{\theta}$ satisfy the relation

$$\dot{z} = \tfrac{1}{2} r^2 \dot{\theta}.$$

For a path lying entirely in the plane $z = 0$, this implies either $r = 0$ or $\dot{\theta} = 0$. In the first case, the path is trivial. In the second, it lies on a line through the origin. Such a curve may be parametrized by

$$\gamma(t) = (c_1 t^{1/2}, c_2 t^{1/2}, 0, \tfrac{1}{2} c_1 t^{-1/2}, \tfrac{1}{2} c_2 t^{-1/2}, 0).$$

It is easy to verify that the desired equations are satisfied. Here, c_1 and c_2 must satisfy the relation

$$c_1^2 + c_2^2 = \sqrt{8\alpha}.$$

However, as t assumes all positive values, this relation places no restriction on admissible planar lines through the origin. Thus, we have a circle's worth of solutions γ. It is easy to check that for any such γ, we have

$$H = 0$$
$$p_\theta = 0$$
$$J = \sqrt{2\alpha},$$

which do not depend on our particular choice of γ. □

Proposition 7 *The only solutions which are constant in the configuration space lie on the z axis.*

Proof Suppose $\gamma(t) = (c_1, c_2, c_3, p_x(t), p_y(t), p_z(t))$ is a solution to Hamilton's equations. Then the equations immediately imply $c_1 = c_2 = p_x = p_y = 0$. The equations for $\dot{x}$ and $\dot{y}$ imply $P_X = P_Y = 0$, and therefore $p_y = \frac{1}{2}c_1 p_z$ and $p_x = -\frac{1}{2}c_2 p_z$. Then the $\dot{p}_i$ are all constant, so the curves $p_i(t)$ are lines. In fact, if we write $p_z(t) = mt+b$, then we must have

$$p_x(t) = -\tfrac{1}{2}c_2 mt - \tfrac{1}{2}c_2 b \quad \text{and} \quad p_y(t) = \tfrac{1}{2}c_1 mt + \tfrac{1}{2}c_1 b.$$

The equations for the $\dot{p}_i$ force

$$m = -\frac{\alpha}{16}\mu^{-3/2}c_3$$

as well as

$$c_1 c_3 = 64(c_1^2 + c_2^2)c_2 \quad \text{and} \quad c_2 c_3 = -64(c_1^2 + c_2^2)c_1.$$

The last two equations imply $c_1 = c_2 = 0$, which in turn force $p_x = p_y = 0$.

Now, for any $k \neq 0$, the path

$$\gamma(t) = (0, 0, k, 0, 0, -\tfrac{4\alpha}{k^2}t)$$

is a solution. It is easy to verify that the equations for $\dot{x}, \dot{y}, \dot{z}, \dot{p}_x$, and $\dot{p}_y$ are satisfied (all equal to zero), and we see that

$$\begin{aligned} -\tfrac{\alpha}{16}\mu^{-3/2}z &= -\tfrac{\alpha}{16}\{\tfrac{1}{16}k^2\}^{-3/2}k \\ &= -\tfrac{4\alpha}{k^2} \\ &= \dot{p}_z. \end{aligned}$$

Thus, such a γ is indeed a solution. Note that the momentum, surprisingly, is not constant, and that the total energy

$$H = U = -\frac{4\alpha}{|k|}$$

is strictly negative. $\square$

Next, we explicitly integrate the equations of motion on a codimension 3 submanifold, and recover conics reminiscent of the Euclidean Kepler Problem. Consider the smooth submanifold $N = \{z = p_z = p_\theta = 0\}$. This submanifold is invariant under the dynamics, since $\dot{z} = \dot{p}_z = \dot{p}_\theta = 0$ on N. The Hamiltonian is

$$H|_N = \tfrac{1}{2}p_r^2 - \frac{\alpha}{r^2},$$

which has the form of a classical central force problem in the plane. Fix an energy level $H|_N = h$. Then since $p_r = \dot{r}$, we can explicitly solve for $r(t)$ as follows.

Proposition 8 *On N, $r(t)$ traces out a hyperbola if $h > 0$, an ellipse if $h < 0$, and a parabola if $h = 0$.*

Proof The Hamiltonian may be rewritten as the simple ODE

$$\tfrac{1}{2}\left(\frac{dr}{dt}\right)^2 = \frac{\alpha}{r^2} + h,$$

which we rewrite

$$dr = \sqrt{2}\frac{\sqrt{\alpha + r^2 h}}{r} dt.$$

Assume temporarily that $h \neq 0$. Integrating, we find

$$\begin{aligned} t &= \int dt \\ &= \tfrac{1}{\sqrt{2}} \int \frac{r}{\sqrt{\alpha + r^2 h}} dr \\ &= \tfrac{1}{h\sqrt{2}} \sqrt{\alpha + r^2 h}. \end{aligned}$$

Inverting this equation, we find

$$r(t) = \sqrt{2ht^2 - \tfrac{\alpha}{h}}.$$

This equation may be rewritten

$$r^2 - 2ht^2 = -\tfrac{\alpha}{h}.$$

Since $\alpha > 0$, this curve in the r, t-plane is an ellipse for $h < 0$ and a hyperbola for $h > 0$. Now, if $h = 0$, we find that

$$t = \tfrac{1}{\sqrt{2\alpha}} \int r dr = \tfrac{1}{2\sqrt{2\alpha}} r^2,$$

and thus

$$r^2 = \sqrt{8\alpha}\, t.$$

$\square$

This seems like a good place to compute the constant α.

Proposition 9 *We have $\alpha = \frac{2}{\pi}$.*

Proof Note that Folland uses the variable $t = z/4$. Consequently, our Laplacian is (negative, by choice of convention) 4 times his Laplacian, and our constant α will be 4 times his constant, which he denotes by c_1. Thus, according to [12], we have

$$\begin{aligned}
1/c_1 &= 3\int_{\mathbb{H}} \frac{r^2}{(r^4+t^2+1)^{5/2}} d\mathrm{vol} \\
&= 3\int_0^{\infty}\int_0^{2\pi}\int_{-\infty}^{\infty} \frac{r^2}{(r^4+t^2+1)^{5/2}} r dr d\theta dt \\
&= 6\pi \int_0^{\infty}\int_{-\infty}^{\infty} \frac{r^3}{(r^4+t^2+1)^{5/2}} dr dt \\
&= \frac{3\pi}{2}\int_0^{\infty}\int_{-\infty}^{\infty} \frac{du dt}{(u+t^2+1)^{5/2}} \\
&= \pi \int_{-\infty}^{\infty} (1+t^2)^{-3/2} dt \\
&= \pi \int_{-\pi/2}^{\pi/2} (1+\tan^2\theta)^{-3/2}\sec^2\theta d\theta \\
&= \pi \int_{-\pi/2}^{\pi/2} \cos\theta d\theta = 2\pi,
\end{aligned}$$

where we made the substitutions $u = r^4$ and $\tan\theta = t$. Thus,

$$\alpha = 4c_1 = 2/\pi.$$

□

3.1.3 Kepler's Third Law

Recall that in the classical (Euclidean) Kepler Problem, Kepler's third law says that the period T of an orbit and its size a (semi-major axis) are related by a universal monomial relation $a^3 = CT^2$.

Any homogeneous potential V on a Euclidean space satisfies a version of Kepler's third law. Homogeneity can be thought of as a scaling symmetry: if $x \mapsto \lambda x$ then $V(x) \mapsto V(\lambda x) = \lambda^{-\kappa} V(x)$. We attempt to extend the symmetry to time and velocities by a power law ansatz: $(x, t, v) \mapsto (\lambda x, \lambda^{\beta} t, \lambda^{-\nu} v)$. Balancing the resulting scalings of potential and kinetic energies implies that $\nu = \kappa/2$. Forcing $v = dx/dt$ gives $\beta = 1 + (\kappa/2)$. This yields the extended scaling

$$(x, t, v) \mapsto (\lambda x, \lambda^{1+\kappa/2} t, \lambda^{-\kappa/2} v).$$

For curves $\gamma(t)$, the scaling operation is

$$\gamma(t) \mapsto \gamma_\lambda(t) = \lambda\gamma(\lambda^{-(1+\kappa/2)}t).$$

One can check that γ_λ satisfies Newton's equation $\ddot{\gamma} = -\nabla V(\gamma)$ if γ does. The scaling symmetry thus takes solutions of energy h to solutions of energy $\lambda^{-\kappa}h$. So, if γ is periodic of period T and size a, then γ_λ is periodic of period $\lambda^\beta T = \lambda^{1+\kappa/2}T$ and size λa. This implies the generalized Kepler's third law for homogeneous potentials on Euclidean space: $T^2 = Ca^{2+\kappa}$.

Now, on the Heisenberg group, with respect to the dilation

$$\delta_\lambda \colon (x, y, z) \mapsto (\lambda x, \lambda y, \lambda^2 z),$$

the potential U is homogeneous of degree -2. That is, $\delta_\lambda \colon U \mapsto \lambda^{-2}U$. Following procedure above, we find that time scales as $t \mapsto \lambda^2 t$, and we can explicitly prove our version of Kepler's third law.

Fix a solution curve γ which satisfies the equations of motion. For any positive real number λ, define a new scaled curve by

$$\gamma_\lambda := \delta_\lambda(\gamma(\lambda^{-2}t).$$

It is easy to prove that γ_λ also satisfies Hamilton's equations.

The following result would be vacuous if there were no periodic orbits. Luckily, they do in fact exist (see Section 3.3).

Proposition 10 (Kepler's Third Law on the Heisenberg Group) *Let γ be a periodic orbit with period T. Choosing a suitable notion of the 'size' a of a periodic orbit yields the Heisenberg version of Kepler's third law:*

$$T^2 = Ca^4. \tag{3.1}$$

Proof Suppose γ is periodic with period T and 'size' a. Then we have

$$\gamma_\lambda(\lambda^2 T) = \delta_\lambda(\gamma(T)) = \delta_\lambda\gamma(0) = \gamma_\lambda(0),$$

so γ_λ has period $T_\lambda := \lambda^2 T$. That is, for any given periodic orbit γ with period T, we can scale to obtain a family of periodic orbits γ_λ with periods $\lambda^2 T$. Now, for our fixed γ, there exists a constant C such that $T^2 = Ca^4$. Then

$$T_\lambda^2 = (\lambda^2 T)^2 = \lambda^4(Ca^4) = C(\lambda a)^4 = Ca_\lambda^4.$$

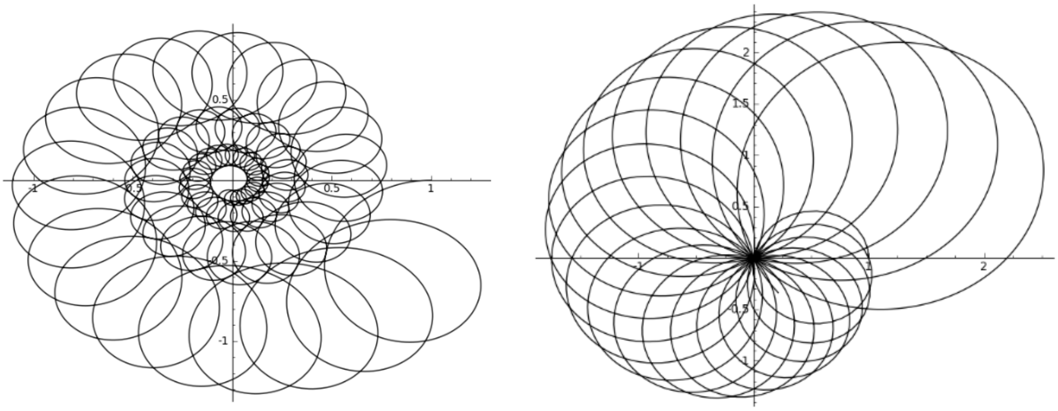

Figure 3.1: Projections of zero-energy orbits to the xy-plane.

So the law (3.1) holds for any orbit γ_λ as well (with the same constant C), where its size is $a_\lambda := \lambda a$. $\square$

3.2 An Integrable Subsystem

In this section we show that our dynamics are integrable on a smooth hypersurface, and we reduce the integration of the equations of motion there to the problem of parametrizing a family of algebraic plane curves. Approximations of two orbits are shown in Figures 3.1 and 3.2. Note that these orbits 'look integrable.' Only the projections to the xy-plane are shown in Figure 3.1, while the corresponding three-dimensional orbits appear in Figure 3.2. We have numerically found examples of such orbits which 'spiral in' towards the origin, and others which 'spiral out.' Note the resemblance to the helical Heisenberg geodesics – the geometry here clearly influences the dynamics. Throughout this section let $\{\cdot, \cdot\}$ denote the Poisson bracket.

3.2.1 The Zero Energy Hypersurface

The following lemma shows that the subvariety $\{H = 0\}$ is indeed a smooth submanifold of $T^*\mathbb{H}$.

Lemma 11 *The only critical points of H are of the form $(q, p) = (0, 0, 0, 0, 0, p_z)$.*

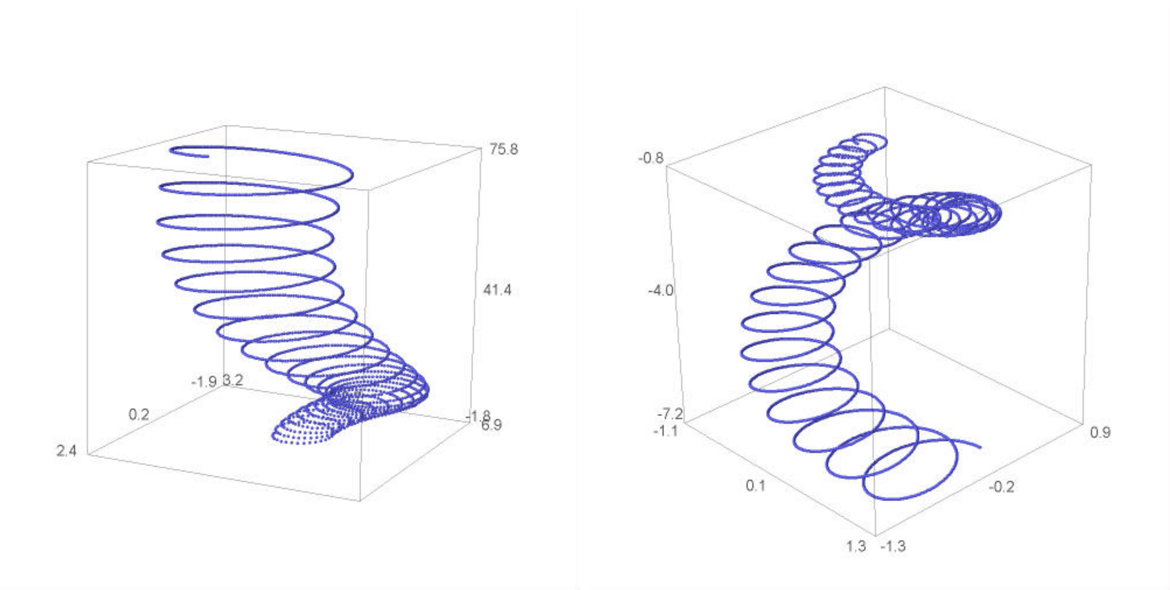

Figure 3.2: Zero-energy orbits in the Heisenberg group.

Proof Suppose $(q, p) = (x, y, z, p_x, p_y, p_z)$ is a critical point of H. Then Hamilton's equations imply

$$DH(q,p) = \begin{bmatrix} -\dot{p}_x & -\dot{p}_y & -\dot{p}_z & \dot{x} & \dot{y} & \dot{z} \end{bmatrix} = 0.$$

Now, $\dot{p}_z = 0$ implies $z = 0$. Also, $\dot{y} = 0$ and $\dot{p}_x = 0$ together imply $x = 0$. Similarly, $\dot{x} = 0$ and $\dot{p}_y = 0$ together imply $y = 0$. Then $y = 0$ and $\dot{x} = 0$ force $p_x = 0$, and similarly, $x = 0$ and $\dot{y} = 0$ force $p_y = 0$. $\square$

Note that the value of H is undefined at these points.

Corollary 12 *All level sets of H are smooth 5-dimensional submanifolds of $T^*\mathbb{H}$.*

Next, we show that the dynamics are integrable by quadratures on this submanifold.

Proposition 13 *Hamilton's equations are integrable by quadratures on $\{H = 0\} \cap \{x = y = p_x = p_y = 0\}^c$.*

Proof We invoke Theorem 5.1 from [4]. Our system, which happens to be autonomous, has three functions on phase space, $F_1 = H, F_2 = p_\theta$, and $F_3 = J$, satisfying the relations

$$\{F_1, F_2\} = \{F_2, F_3\} = 0$$
$$\{F_1, F_3\} = 2F_1.$$

Thus, our only non-zero structure constant is $c_{13}^1 = 2$. A lengthy computation shows that $dH \wedge dp_\theta \wedge dJ$ is non-zero whenever x, y, p_x, and p_y are not all zero, so that the differentials dF_i are linearly independent away from the surface $x = y = p_x = p_y = 0$. Next, the fact that c_{13}^1 is our only non-zero structure constant implies that the theorem will hold on the set $F_1 = H = 0$. Finally, we note that the Lie algebra $\mathcal{A}$ spanned by the F_i (with multiplication given by the Poisson bracket) is solvable (although not nilpotent), which is evident from the bracket relations above. Theorem 5.1 then guarantees that for $H = 0$ and x, y, p_x, and p_y not all zero, the system

$$(\dot{q}, \dot{p}) = I(dH)$$

can be integrated by quadratures, where I denotes the inverse of the vector bundle isomorphism determined by the canonical symplectic form on $T^*\mathbb{H}$. $\square$

Remark 1 *We pause to explain the details omitted from the end of the above proof for those unfamiliar with symplectic geometry. Like any cotangent bundle, $T^*\mathbb{H}$ is equipped with a natural symplectic form ω. This two-form induces a map*

$$\tilde{\omega} \colon T(T^*\mathbb{H}) \to T^*(T^*\mathbb{H})$$
$$v \mapsto \iota_\omega v = \omega(v, \cdot).$$

The non-degeneracy of ω ensures that this map is a linear bundle isomorphism, so we may define $I = \tilde{\omega}^{-1}$. Then we have $I(dH) = \operatorname{sgrad} H = X_H$, the symplectic gradient or Hamiltonian vector field of H, so solutions to the system $(\dot{q}, \dot{p}) = I(dH)$ are precisely the solutions to Hamilton's equations.

3.2.2 A Change of Coordinates

Consider a new Hamiltonian

$$\hat{H} = -\frac{K}{U} = \tfrac{\alpha}{2}(P_X^2 + P_Y^2)\mu^{1/2}.$$

It is easy to check that $\hat{H}$ is scale invariant; that is, $\delta_\lambda \colon \hat{H} \mapsto \hat{H}$. Note also that the zero energy hypersurface is precisely the level set of $\hat{H}$ of value 1:

$$\{H = 0\} = \{\hat{H} = 1\}.$$

An elementary lemma from symplectic geometry shows that the flows of H and $\hat{H}$ are the same up to reparametrization on this hypersurface (this can be thought of as a version of the Jacobi-Maupertuis principle). We will thus investigate the flow of $\hat{H}$, which we will find to be integrable.

Lemma 14 *Let (M, ω) be symplectic, and let $f, g \in C^\infty(M)$. Suppose there exist constants c_1 and c_2 and a hypersurface N such that $N = f^{-1}(c_1) = g^{-1}(c_2)$. Suppose further than the differentials of f and g do not vanish on N. Then there is a nonvanishing function $\xi \colon N \to \mathbb{R}$ such that*

$$X_f = \xi X_g$$

on N, where X_f and X_g denote the Hamiltonian vector fields of f and g, respectively.

Proof Let $p \in N$. Then $T_pN = \ker df_p = \ker dg_p$, since N is a level set of both f and g. Thus, there is some scalar $\xi(p) \neq 0$ such that $df_p = \xi(p) dg_p$. Since $p \in N$ was arbitrary, we have $df = \xi dg$ on N. But ω defines an invertible correspondence

$$X_f \mapsto \omega(X_f, \cdot) = df,$$

so we must have $X_f = \xi X_g$ on N. □

Proposition 15 *We have $\{\hat{H}, J\} = 0$.*

Proof This is a straightforward calculation:

$$\begin{aligned}
\{\hat{H}, J\} &= \frac{\partial \hat{H}}{\partial x}\frac{\partial J}{\partial p_x} + \frac{\partial \hat{H}}{\partial y}\frac{\partial J}{\partial p_y} + \frac{\partial \hat{H}}{\partial z}\frac{\partial J}{\partial p_z} - \frac{\partial \hat{H}}{\partial p_x}\frac{\partial J}{\partial x} - \frac{\partial \hat{H}}{\partial p_y}\frac{\partial J}{\partial y} - \frac{\partial \hat{H}}{\partial p_z}\frac{\partial J}{\partial z} \\
&= \tfrac{\alpha}{2}\big(p_z P_Y \mu^{1/2} + (P_X^2 + P_Y^2) 2xr^2 \mu^{-1/2}\big)x \\
&\quad + \tfrac{\alpha}{2}\big(- p_z P_X \mu^{1/2} + (P_X^2 + P_Y^2) 2yr^2 \mu^{-1/2}\big)y + \tfrac{\alpha}{16}(P_X^2 + P_Y^2)\mu^{-1/2} z^2 \\
&\quad - \alpha P_X \mu^{1/2} p_x - \alpha P_Y \mu^{1/2} p_y - \alpha(xP_Y - yP_X)\mu^{1/2} p_z \\
&= -\tfrac{\alpha}{2} p_z (xP_Y - yP_X)\mu^{1/2} + \alpha(P_X^2 + P_Y^2)\mu^{1/2} - \alpha(p_x P_X + p_y P_Y)\mu^{1/2} \\
&= \alpha \mu^{1/2}\big((P_X^2 + P_Y^2) + P_X(\tfrac{1}{2} y p_z - p_x) + P_Y(-\tfrac{1}{2} x p_z - p_y)\big) \\
&= 0.
\end{aligned}$$

Note that since J generates the scaling action δ_λ, this also implies that $\hat{H}$ is scale invariant. □

Proposition 16 *We have* $\{\hat{H}, p_\theta\} = 0$.

Proof This is an easy calculation:

$$\begin{aligned}
\{\hat{H}, p_\theta\} &= \tfrac{\alpha}{2}\big(p_z P_Y \mu^{1/2} + (P_X^2 + P_Y^2) 2xr^2\mu^{-1/2}\big)(-y) \\
&\quad + \tfrac{\alpha}{2}\big(- p_z P_X \mu^{1/2} + (P_X^2 + P_Y^2) 2yr^2 \mu^{-1/2}\big)x \\
&\quad - \alpha P_X \mu^{1/2} p_y - \alpha P_Y \mu^{1/2}(-p_x) \\
&= \tfrac{\alpha}{2}\big(- y p_z P_Y \mu^{1/2}\big) + \tfrac{\alpha}{2}\big(- x p_z P_Y \mu^{1/2}\big) \\
&\quad - \alpha p_y P_X \mu^{1/2} + \alpha p_x P_Y \mu^{1/2} \\
&= \alpha\mu^{1/2}\big(P_X(-p_y - \tfrac{1}{2}xp_z) + P_Y(p_x - \tfrac{1}{2}yp_z)\big) \\
&= 0.
\end{aligned}$$

□

These last two propositions show that both J and p_θ are conserved by the flow of $\hat{H}$. Thus, we have three integrals for the flow of $\hat{H}$: J, p_θ, and $\hat{H}$ itself. The flow is therefore integrable. When $\hat{H} = 1$, the flow corresponds to a multiple of the flow for H on the submanifold $H = 0$. The first step in finding explicit solutions is changing variables. Define the coordinate transformation

$$\Phi\colon (\theta, r, z) \mapsto (\theta, r, v = z/r^2).$$

We want Φ to be a canonical transformation, so we require $\Phi^*\Lambda = \Lambda$, where Λ denotes the canonical one-form. That is, we require

$$\begin{aligned}
\Lambda &= p_x dx + p_y dy + p_z dz \\
&= p_\theta d\theta + p_r dr + p_z dz \\
&= \tilde{p}_\theta d\theta + \tilde{p}_r dr + p_v dv,
\end{aligned}$$

where $\tilde{p}_\theta, \tilde{p}_r, p_v$ are our new momentum coordinates. Solving, we define the new momenta by the equations

$$\begin{aligned} p_\theta &= \tilde{p}_\theta \\ p_r &= \tilde{p}_r - \frac{2vp_v}{r} \\ p_z &= \frac{p_v}{r^2}. \end{aligned}$$

Proposition 17 *In these coordinates, $J = r\tilde{p}_r$.*

Proof We have

$$\begin{aligned} J &= rp_r + 2zp_z \\ &= r(\tilde{p}_r - 2vp_v/r) + 2(vr^2)(p_v/r^2) \\ &= r\tilde{p}_r. \end{aligned}$$

$\square$

Note that, in this form, the function J resembles the analogous generating function for Euclidean dilations.

Proposition 18 *We have $\hat{H} = \hat{H}(J, p_\theta, v, p_v)$. More specifically,*

$$\hat{H} = \tfrac{1}{2\alpha}\big((J - 2vp_v)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\big)(1 + \tfrac{1}{16}v^2)^{1/2}.$$

Proof We first compute the kinetic energy:

$$\begin{aligned} 2K &= p_r^2 + \Big(\frac{p_\theta}{r} + \tfrac{1}{2}rp_z\Big)^2 \\ &= \Big(\tilde{p}_r - \frac{2vp_v}{r}\Big)^2 + \Big(\frac{p_\theta}{r} + \tfrac{1}{2}r\frac{p_v}{r^2}\Big)^2 \\ &= \frac{1}{r^2}\Big((J - 2vp_v)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\Big). \end{aligned}$$

Next, we compute the potential energy:

$$\begin{aligned} U &= -\alpha(r^4 + \tfrac{1}{16}z^2)^{-1/2} \\ &= -\alpha(r^4 + \tfrac{1}{16}(vr^2)^2)^{-1/2} \\ &= -\frac{\alpha}{r^2}(1 + \tfrac{1}{16}v^2)^{-1/2}. \end{aligned}$$

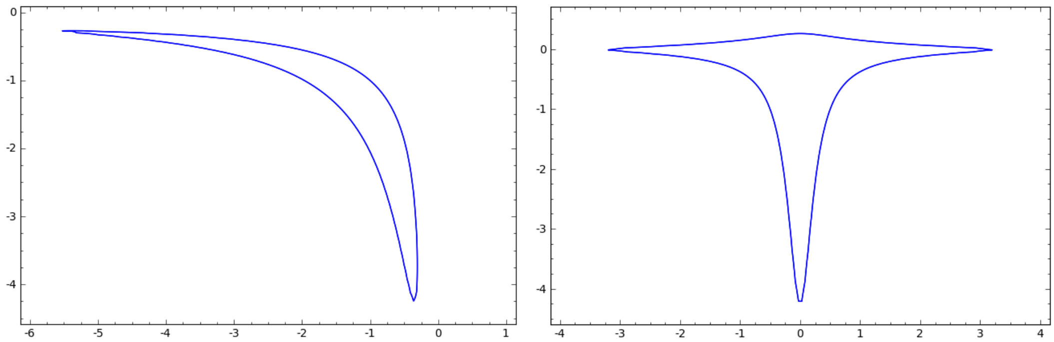

Figure 3.3: Curves in the v, p_v-plane corresponding to $J = 3,\ p_\theta = 1$ (left) and $J = 0,\ p_\theta = .5$ (right).

Thus, we find

$$\begin{aligned}
\hat{H} &= -\frac{K}{U} \\
&= \frac{\frac{1}{2r^2}\Big((J - 2vp_v)^2 + (p_\theta + \frac{1}{2}p_v)^2\Big)}{\frac{\alpha}{r^2}(1 + \frac{1}{16}v^2)^{-1/2}} \\
&= \tfrac{1}{2\alpha}\big((J - 2vp_v)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\big)(1 + \tfrac{1}{16}v^2)^{1/2}.
\end{aligned}$$

□

On the submanifold $\{H = 0\}$, we have $\hat{H} = 1$. Also, the initial conditions determine the constants J and p_θ. Thus, given initial conditions, $\hat{H}$ is a function of v and p_v only. We thus arrive at the following result.

Corollary 19 *For $\hat{H} = 1$, any solution must project to an algebraic curve in the v, p_v-plane.*

In other words, the equation

$$\tfrac{1}{2\alpha}\big((J - 2vp_v)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\big)(1 + \tfrac{1}{16}v^2)^{1/2} = 1 \tag{3.2}$$

defines a two-parameter family of plane curves in the variables v and p_v, parametrized by the quantities J and p_θ. Examples of these curves are shown in Figure 3.3.

Now we make a change of variables to realize these curves as algebraic of degree 6. We first write the equation $\hat{H} = 1$ as the zero locus of a degree 10 polynomial by squaring both sides:

$$\big((J - 2vp_v)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\big)^2(1 + \tfrac{1}{16}v^2) = 4\alpha^2. \tag{3.3}$$

Making the substitution $w = vp_v$, one has

$$1 + \tfrac{1}{16}v^2 = \frac{16p_v^2 + w^2}{16p_v^2},$$

so that (3.3) reduces to

$$\big((J - 2w)^2 + (p_\theta + \tfrac{1}{2}p_v)^2\big)^2(16p_v^2 + w^2) = 64\alpha^2 p_v^2. \tag{3.4}$$

This defines a family of (non-homogeneous) algebraic curves of degree 6 in the variables w and p_v, parametrized by J and p_θ.

3.3 Existence of Periodic Solutions

In this section we prove our most important theorem: there exist periodic orbits in the Kepler Problem on the Heisenberg group. These orbits were originally found by numerical experiment. To prove their existence we employ the direct method in the calculus of variations, showing the existence of an action minimizing orbit with prescribed symmetry. We prove the existence of solutions with k-fold rotational symmetry for any odd integer $k \geq 3$. Approximations of one such orbit, with $k = 3$, are shown in Figures 3.4 and 3.5. For the reader's benefit we first outline the structure of the proof, then carry out the analysis.

Theorem 2 *Periodic solutions exist. For any odd integer $k \geq 3$, there exists a periodic orbit with k-fold rotational symmetry about the z-axis.*

Proof of Theorem 2. We first sketch the structure of the proof, using the direct method from the calculus of variations. For similar applications of this technique to celestial mechanics problems, see [10] and [15].

Step 1: Choose a nice function space $\mathcal{F}_k$ whose members are closed loops enjoying the desired symmetry properties. Choose a minimizing sequence of curves in $\mathcal{F}_k$ whose action approaches the infimum of the action restricted to $\mathcal{F}_k$.

Step 2: Using Arzela-Ascoli, show γ_n has a C^0-convergent subsequence converging to some γ_*. Using Banach-Alaoglu, show $\gamma_n \rightharpoonup \gamma_* \in \mathcal{F}_k$.

Step 3: Show that γ_* realizes the infimum of the action restricted to $\mathcal{F}_k$. Use Fatou's Lemma and standard functional analysis.

Step 4: Prove that γ_* does not suffer a collision. Use the Hamilton-Jacobi equation.

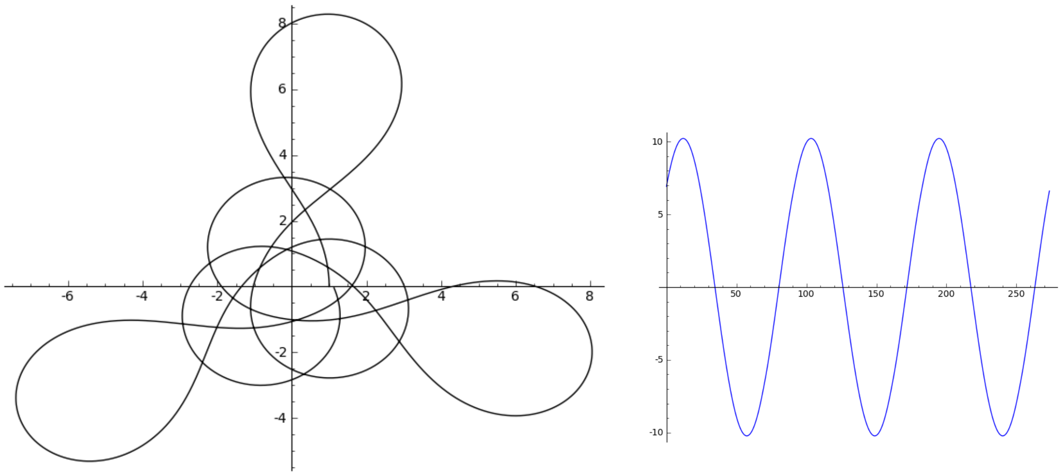

Figure 3.4: Projection of a periodic orbit to xy-plane (left). The z-coordinate over time (right).

Step 5: Show $dA(\gamma_*)(e) = 0$ for horizontal variations e. Standard analysis gives $dA(\gamma_*)|_{\mathcal{F}_k} = 0$, then use the Principal of Symmetric Criticality.
Step 6: Show γ_* satisfies the Euler-Lagrange equations, and consequently, Hamilton's equations. This is the Principle of Least Action.

3.3.1 The Function Space and a Minimizing Sequence

For a curve $\gamma(t) = (x(t), y(t), z(t))$ in the Heisenberg group parametrized by the interval $[0, T]$, let $\tilde{\gamma}$ denote its projection to the xy-plane. Let $k \geq 3$ be any odd[1] positive integer. We will restrict our attention to horizontal curves satisfying the symmetry conditions

$$\gamma(t + T/k) = R_{2\pi/k}\gamma(t) \tag{S1}$$

$$z(t + T/2) = -z(t) \tag{S2}$$

where

$$R_{2\pi/k} = \begin{bmatrix} \cos(2\pi/k) & -\sin(2\pi/k) & 0 \\ \sin(2\pi/k) & \cos(2\pi/k) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

[1]If k is even, the two symmetry conditions force z to be identically zero; such solutions are known (Proposition 6) to suffer collisions.

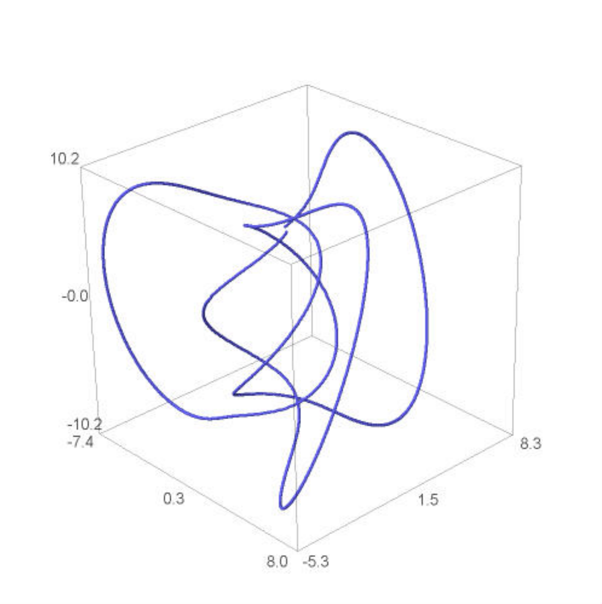

Figure 3.5: A periodic orbit in three dimensions.

denotes rotation about the z-axis by $2\pi/k$ radians counterclockwise. Note that curves satisfying condition (S1) are necessarily periodic. Also, note that the two symmetry conditions together give the z-coordinate symmetry

$$z(t + mT/2k) = (-1)^m z(t),$$

for any $m \in \mathbb{Z}$.

We will work in the function space

$$\mathcal{F}_k = \{\gamma \in H^1(S^1, \mathbb{H}) \mid \gamma \text{ horizontal and satisfies (S1) and (S2)}\},$$

where $H^1(S^1, \mathbb{H}) = W^{1,2}(S^1, \mathbb{H})$ is the completion of the space of all absolutely continuous paths in $\mathbb{H}$ whose derivative is square integrable[2]. The usual $H^1(S^1, \mathbb{R}^3)$ norm is

$$||\gamma||_{H^1} = \sqrt{\int_0^T (||\dot{\gamma}(t)||_E^2 + ||\gamma(t)||_E^2) dt},$$

where $||\cdot||_E$ denotes the usual Euclidean norm on $\mathbb{R}^3$. This norm endows $H^1(S^1, \mathbb{R}^3)$ with a Hilbert space structure.

[2]Here and in the sequel S^1 denotes the interval $[0, T]$ with endpoints identified.

To endow $\mathcal{F}_k$ with a Hilbert structure, we make the following identification:

$$\mathcal{L}^2([0,T],\mathbb{R}^2)\times\mathbb{H}\cong\mathcal{H}:=\{\text{horizontal square-integrable paths in }\mathbb{H}\}.$$

This isomorphism sends $((f,g),x_0)$ to the horizontal curve γ which solves the initial value problem

$$\dot\gamma=fX+gY,\quad \gamma(0)=x_0.$$

The existence and uniqueness of the solution γ is guaranteed by Theorem D.1 of [21]. This theorem also shows that this mapping is invertible for x_0 in some compact set. Thus, we can think of f,g as coordinates on the subspace of $\mathcal{H}$ consisting of all paths with a fixed starting point. Consequently, $\mathcal{F}_k$ is equipped with a vector space structure.

Here we will endow $\mathcal{F}_k$ with a norm similar to the H^1 norm, but slightly modified for our purposes:

$$||\gamma||_{\mathcal{F}_k}:=\sqrt{\int_0^T(\langle\dot\gamma(t),\dot\gamma(t)\rangle+||\gamma(0)||_{sR}^2dt}. \tag{3.5}$$

Remark 2 *Since $\mathbb{H}$ is a finite dimensional vector space, all norms on $\mathbb{H}$ are Lipschitz equivalent and give the same topology as the topology induced by the Carnot-Caratheodory metric. Thus, a set in $\mathbb{H}$ is bounded in the Carnot-Caratheodory metric if and only if it is bounded in the Euclidean metric. Moreover, our choice of norm $||\gamma||_{\mathcal{F}_k}$ in* (3.5) *will be the most convenient to work with. However, it is equivalent to the alternative function norms:*

$$||\gamma||=\sqrt{\int_0^T(\langle\dot\gamma(t),\dot\gamma(t)\rangle+||\gamma(t)||_{sR}^2dt} \tag{3.6}$$

$$||\gamma|=\sqrt{\int_0^T(\langle\dot\gamma(t),\dot\gamma(t)\rangle+||\gamma(0)||_{E}^2dt} \tag{3.7}$$

$$||\gamma||=\sqrt{\int_0^T(\langle\dot\gamma(t),\dot\gamma(t)\rangle+||\gamma(t)||_{E}^2dt.} \tag{3.8}$$

Finally, note that we also have the following norms:

$$||p||_{sR}=d(p,0),\qquad p\in\mathbb{H}$$

$$||\dot\gamma(t)||_{\mathbb{H}}^2=\langle\dot\gamma(t),\dot\gamma(t)\rangle=(\dot x(t))^2+(\dot y(t))^2=2K(\gamma(t))$$

$$||\dot\gamma||_{\mathcal{L}^2}^2=\sqrt{\int_0^T||\dot\gamma(t)||_{\mathbb{H}}^2dt.}$$

Proposition 20 (Coercivity) *The squared length of $\tilde{\gamma}$ is bounded above by twice the action of γ.*

Proof Let $\tilde{\gamma}$ denote the projection of $\gamma(t) = (r(t), \theta(t), z(t))$ to the r, θ plane. Then we compute

$$
\begin{aligned}
A(\gamma) &= \int_0^T L(t, \gamma, \dot{\gamma}) dt \\
&= \int_0^T \left(\tfrac{1}{2}\dot{r}^2 + \tfrac{1}{2}r^2\dot{\theta}^2 + \alpha(r^4 + \tfrac{1}{16}z^2)^{-1/2}\right) dt \\
&\geq \int_0^T \left(\tfrac{1}{2}\dot{r}^2 + \tfrac{1}{2}r^2\dot{\theta}^2\right) dt \\
&= \tfrac{1}{2} \int_0^T (\dot{x}^2 + \dot{y}^2) dt \\
&\geq \tfrac{1}{2} \left(\int_0^T \sqrt{\dot{x}^2 + \dot{y}^2} dt \right)^2 \\
&= \tfrac{1}{2} \big(l(\gamma)\big)^2,
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality in $\mathcal{L}^2(\mathbb{R})$. $\square$

Now suppose $\{\gamma_n\}_{n \in \mathbb{N}}$ is a minimizing sequence in $\mathcal{F}_k$. That is, suppose

$$\lim_{n \to \infty} A(\gamma_n) = \inf_{\gamma \in \mathcal{F}_k} A(\gamma).$$

We may discard finitely many terms of the sequence and assume that there exists some large $M > 0$ such that

$$A(\gamma_n) \leq M$$

for all n. Note that the previous Proposition implies that the lengths $l(\gamma_n)$ are bounded; specifically, $l(\gamma_n) \leq \sqrt{2M}$ for all n.

3.3.2 The Potential Solution

We will denote the usual Euclidean norm by $||\cdot||_E$ and the corresponding Euclidean distance function by $d(\cdot,\cdot)_E$. Note that $d(\cdot,\cdot)_E \leq d(\cdot,\cdot)$ in general, but that both the Euclidean and sub-Riemannian distances (and consequently, norms) agree when restricted to the xy-plane.

Proposition 21 *The projections $\tilde{\gamma}_n$ are uniformly bounded.*

Proof Since $\gamma_n \in \mathcal{F}_k$, it is horizontal, so the length of γ_n is equal to that of $\tilde{\gamma}_n$. Also, condition (S1) implies that

$$z(0) = z(T/k) = z(2T/k) = \cdots = z((k-1)T/k) = z(T)$$

and thus that the k points

$$\gamma_n(0) = \gamma_n(T),\ \gamma_n(T/k),\ \gamma_n(2T/k), \ldots,\ \gamma_n((k-1)T/k)$$

form a regular k-gon in the plane $z = z(0)$ centered at the point $(0, 0, z(0))$. See Figure 3.6 for a rendering of the $k = 5$ case. Then we have:

$$\begin{aligned}
l(\gamma_n) &\geq d(\gamma_n(0), \gamma_n(\tfrac{T}{k})) + d(\gamma_n(\tfrac{T}{k}), \gamma_n(\tfrac{2T}{k})) + \cdots + d(\gamma_n(\tfrac{(k-1)T}{k}), \gamma_n(T)) \\
&\geq d_E(\gamma_n(0), \gamma_n(\tfrac{T}{k})) + d_E(\gamma_n(\tfrac{T}{k}), \gamma_n(\tfrac{2T}{k})) + \cdots + d_E(\gamma_n(\tfrac{(k-1)T}{k}), \gamma_n(T)) \\
&= d_E(\tilde{\gamma}_n(0), \tilde{\gamma}_n(\tfrac{T}{k})) + d_E(\tilde{\gamma}_n(\tfrac{T}{k}), \tilde{\gamma}_n(\tfrac{2T}{k})) + \cdots + d_E(\tilde{\gamma}_n(\tfrac{(k-1)T}{k}), \tilde{\gamma}_n(T)) \\
&= C||\tilde{\gamma}_n(0)||_E \\
&= C||\tilde{\gamma}_n(0)||_{sR},
\end{aligned}$$

where $C = 2k\sin(2\pi/k)$ and the penultimate equality is given by the usual perimeter of a regular k-gon inscribed in a circle of radius $||\tilde{\gamma}_n(0)||_E$.

Then since $l(\gamma_n) = l(\tilde{\gamma}_n)$, we find

$$\begin{aligned}
||\tilde{\gamma}_n(t)||_{sR} &\leq ||\tilde{\gamma}_n(0)||_{sR} + l(\tilde{\gamma}_n) \\
&\leq (\tfrac{1}{C} + 1) l(\gamma_n) \\
&\leq (\tfrac{1}{C} + 1)\sqrt{2M}.
\end{aligned}$$

□

Lemma 22 *Suppose $\gamma\colon [0, S] \to \mathbb{H}$ is horizontal. Suppose c, the projection of γ to the xy-plane, satisfies $||c(t)||_E \leq R$ for some $R > 0$ and all $t \in [0, S]$. Then*

$$|z(S) - z(0)| \leq \tfrac{1}{2} R l(\gamma).$$

Proof Let $I(v_1, v_2) = (v_2, -v_1)$. Note that $||I(\dot{c})||_{\mathbb{H}} = ||\dot{c}||_{\mathbb{H}}$. Without loss of generality, we may assume that c has constant speed v. Then by the horizontal

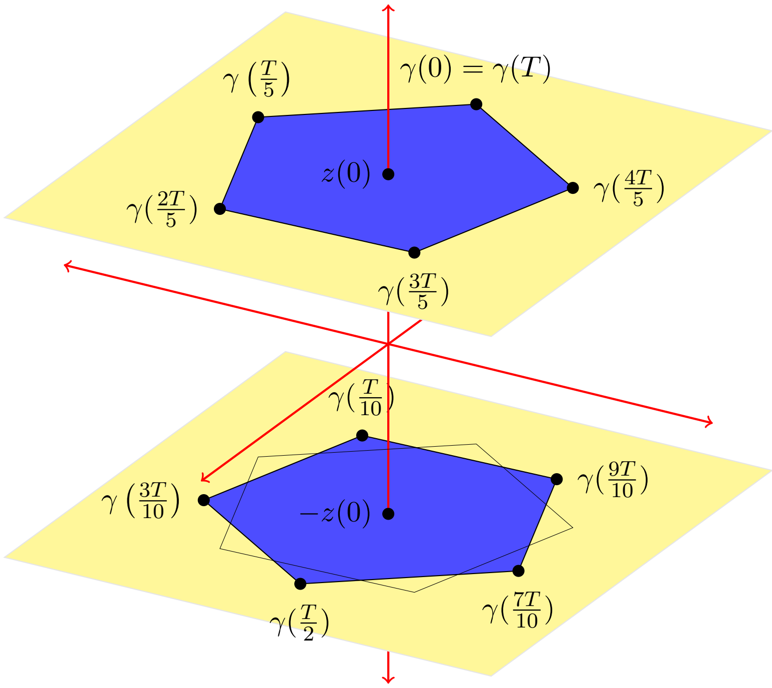

Figure 3.6: Symmetries of a path in $\mathcal{F}_5$.

condition $\frac{1}{2}x\dot{y} - \frac{1}{2}y\dot{x} = \dot{z}$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}
|z(S) - z(0)| &= |\int_0^S \dot{z}dt| \\
&= \tfrac{1}{2}|\int_0^S (x\dot{y} - y\dot{x})dt| \\
&= \tfrac{1}{2}|\int_0^S \langle c, I(\dot{c})\rangle_E dt| \\
&\leq \tfrac{1}{2}\sqrt{\int_0^S ||c(t)||_E^2 dt}\sqrt{\int_0^S ||I(\dot{c}(t))||_E^2 dt} \\
&= \tfrac{1}{2}\sqrt{\int_0^S ||c(t)||_E^2 dt}\sqrt{\int_0^S ||\dot{c}(t)||_E^2 dt} \\
&\leq \tfrac{1}{2}\sqrt{R^2 S}\sqrt{v^2 S} \\
&= \tfrac{1}{2}RvS \\
&= \tfrac{1}{2}Rl(\gamma).
\end{aligned}$$

$\square$

Proposition 23 *The set $\{\gamma_n\}$ is uniformly bounded.*

Proof By Proposition 21, the curves γ_n satisfy the hypothesis of Lemma 22: we can take $R = (\frac{1}{C} + 1)\sqrt{2M}$. Take $S = T/2$ and denote the z-components of the curves γ_n by z_n. Then the Lemma, with the fact that $z_n(0) = -z_n(T/2)$, implies

$$|2z_n(0)| \leq \tfrac{1}{2}Rl(\gamma_n|_{[0,T/2]}) \leq \tfrac{1}{2}Rl(\gamma_n).$$

Then we find

$$\begin{aligned} |z_n(t)| &\leq |z_n(0)| + l(\gamma_n) \\ &= (\tfrac{1}{4}R + 1)l(\gamma_n) \\ &\leq (\tfrac{1}{4}R + 1)\sqrt{2M}. \end{aligned}$$

Thus, the family $z_n(t)$ is uniformly bounded. Since the projections $\tilde{\gamma}_n$ and the z-coordinates z_n are uniformly bounded, so are the curves γ_n. □

Lemma 24 *If $\gamma \in \mathcal{F}_k$ then γ is Hölder continuous with Hölder exponent $\frac{1}{2}$.*

Proof This is a version of the Sobolev embedding theorem. Using the Cauchy-Schwarz inequality with $f = \sqrt{\langle\dot{\gamma}(t), \dot{\gamma}(t)\rangle} = ||\dot{\gamma}(t)||_{\mathbb{H}}$ and $g = 1$, one has

$$\begin{aligned} d(\gamma(r), \gamma(s)) &\leq l(\gamma|_{[r,s]}) \\ &= \int_r^s ||\dot{\gamma}(t)||_{\mathbb{H}} dt \\ &\leq \sqrt{|r-s| \int_r^s ||\dot{\gamma}(t)||^2_{\mathbb{H}} dt} \\ &\leq \sqrt{|r-s| \int_0^T ||\dot{\gamma}(t)||^2_{\mathbb{H}} dt} \\ &= |r-s|^{1/2}\, ||\dot{\gamma}||_{\mathcal{L}^2}. \end{aligned}$$

This shows that γ satisfies the Hölder condition with exponent 1/2 and coefficient $||\dot{\gamma}||_{\mathcal{L}^2}$. □

Lemma 25 *The norms $||\dot{\gamma}_n||_{\mathcal{L}^2}$ are bounded.*

Proof We have

$$\begin{aligned}||\dot{\gamma}_n||^2_{\mathcal{L}^2} &= \int_0^T ||\dot{\gamma}_n||^2_{\mathbb{H}} dt \\ &= \int_0^T 2K(\gamma_n) dt \\ &\leq 2\int_0^T (K(\gamma_n) + U(\gamma_n)) dt \\ &= 2A(\gamma_n).\end{aligned}$$

Thus,

$$||\dot{\gamma}_n||_{\mathcal{L}^2} \leq \sqrt{2M}.$$

□

Proposition 26 *The family $\{\gamma_n\}$ is equicontinuous.*

Proof By Lemma 24, we know

$$d(\gamma_n(r), \gamma_n(s)) \leq \sqrt{|r-s|}\ ||\dot{\gamma}_n||_{\mathcal{L}^2}.$$

Now choose some $\epsilon > 0$ and let $\delta = (\epsilon/\sqrt{2M})^2$. Then if $|r - s| < \delta$, by Lemma 25, for all n we have

$$\begin{aligned}d(\gamma_n(r), \gamma_n(s)) &\leq \sqrt{|r-s|}\ ||\dot{\gamma}_n||_{\mathcal{L}^2} \\ &< \frac{\epsilon}{\sqrt{2M}}\ ||\dot{\gamma}_n||_{\mathcal{L}^2} \\ &\leq \epsilon.\end{aligned}$$

□

Proposition 27 *There is a subsequence $\{\gamma_{n_j}\}$ converging uniformly to some γ_*.*

Proof By Propositions 23 and 26, the sequence $\{\gamma_n\}$ is uniformly bounded and equicontinuous. The Arzela-Ascoli theorem guarantees the existence of such a subsequence, which implies that γ_* is continuous. □

Proposition 28 *There is a subsequence $\{\gamma_{n_{j_i}}\}$ which converges weakly in $\mathcal{F}_k$ to $\gamma_* \in \mathcal{F}_k$.*

Proof From Proposition 23 we know that the norms $||\gamma_n(0)||_{sR}$ are bounded; also, by Lemma 25, the norms $||\dot{\gamma}_n||_{\mathcal{L}^2}$ are bounded. Thus, the norms

$$||\gamma_n||^2_{\mathcal{F}_k} = ||\dot{\gamma}_n||^2_{\mathcal{L}^2} + T||\gamma_n(0)||^2_{sR}$$

are bounded as well. Then the Banach-Alaoglu theorem guarantees the existence of another subsequence which converges weakly to γ_* so that this limiting curve is absolutely continuous. It is clear that this curve must satisfy the horizontal and symmetry conditions. □

Remark 3 *In the following we will re-index so that the sequence* $\{\gamma_n\}$ *converges to* γ_* *both weakly and uniformly.*

3.3.3 Minimization

Lemma 29 *One has*

$$\int_0^T \langle \dot{\gamma}_n, \dot{\gamma}_* \rangle dt \to \int_0^T ||\gamma_*||^2_{\mathbb{H}} dt.$$

Proof From Proposition 28, we know that $\{\gamma_n\}$ converges weakly to γ_*:

$$\gamma_n \rightharpoonup \gamma_*.$$

By definition, we have that

$$\phi(\gamma_n) \to \phi(\gamma_*)$$

for any $\phi \in \mathcal{F}_k^*$. In particular, the functional

$$\phi_c(\gamma) := \int_0^T \langle \dot{\gamma}(t), \dot{c}(t) \rangle dt$$

is indeed an element of $\mathcal{F}_k^*$ for any $c \in \mathcal{F}_k$. To see that ϕ_c is a bounded operator, note that the Cauchy-Schwarz inequality in $\mathcal{L}^2$ gives

$$\phi_c(\gamma) = \langle \dot{c}, \dot{\gamma} \rangle_{\mathcal{L}^2} \leq ||\dot{c}||_{\mathcal{L}^2} ||\dot{\gamma}||_{\mathcal{L}^2} = C||\dot{\gamma}||_{\mathcal{L}^2}.$$

We may therefore choose $c = \gamma_*$, and the Lemma follows.

Alternatively, one can see that ϕ_{γ_*} is bounded on the sequence $\{\gamma_n\}$ as follows. Write

$$\phi_{\gamma_*}(\gamma_n) = \langle \dot{\gamma}_n, \dot{\gamma}_* \rangle_{\mathcal{F}_k} - \langle \gamma_n(0), \gamma_*(0) \rangle_E, \tag{3.9}$$

where $\langle \cdot, \cdot \rangle_E$ denotes the Euclidean inner product on $\mathbb{R}^3$. Then, considering the right hand side of (3.9), note that one can control the first term by weak convergence, and the second by uniform convergence. $\square$

Proposition 30 *Our limiting curve γ_* realizes the infimum of the action:*

$$A(\gamma_*) = \inf_{\gamma \in \mathcal{F}_k} A(\gamma).$$

Proof Since

$$0 \leq \int_0^T ||\dot{\gamma}_n - \dot{\gamma}_*||_{\mathbb{H}}^2 dt = \int_0^T (||\dot{\gamma}_n||_{\mathbb{H}}^2 + ||\dot{\gamma}_*||_{\mathbb{H}}^2 - 2\langle \dot{\gamma}_n, \dot{\gamma}_* \rangle) dt,$$

we have

$$\int_0^T 2\langle \dot{\gamma}_n, \dot{\gamma}_* \rangle dt \leq \int_0^T (||\dot{\gamma}_n||_{\mathbb{H}}^2 + ||\dot{\gamma}_*||_{\mathbb{H}}^2) dt.$$

Taking the limit inferior of both sides and using the previous Lemma yields

$$\int_0^T ||\dot{\gamma}_*||_{\mathbb{H}}^2 dt \leq \liminf \int_0^T ||\dot{\gamma}_n||_{\mathbb{H}}^2 dt.$$

We may rewrite this last inequality as

$$\int_0^T K(\gamma_*(t)) dt \leq \liminf \int_0^T K(\gamma_n(t)) dt. \tag{3.10}$$

Now, we know $\gamma_n(t) \to \gamma_*(t)$ uniformly. Since our potential U is continuous (except at the origin), we have that $U(\gamma_n(t)) \to U(\gamma_*(t))$ uniformly almost everywhere. Then Fatou's Lemma implies

$$\int_0^T U(\gamma_*(t)) dt \leq \liminf \int_0^T U(\gamma_n(t)) dt. \tag{3.11}$$

Adding (3.10) and (3.11) gives

$$A(\gamma_*) = \int_0^T L(\gamma_*(t)) dt \leq \liminf \int_0^T L(\gamma_n(t)) dt = \liminf A(\gamma_n).$$

But $\{\gamma_n\}$ is a minimizing sequence, and $\gamma_* \in \mathcal{F}_k$ so

$$A(\gamma_*) = \liminf A(\gamma_n) = \inf_{\gamma \in \mathcal{F}_k} A(\gamma).$$

$\square$

3.3.4 Avoiding Collision

We will show that a curve in $\mathcal{F}_k$ suffering a collision necessarily has infinite action. Without loss of generality, we may assume the collision occurs at time $t = 0$.

Let $H_g\colon T^*\mathbb{H} \to \mathbb{R}$ denote the Hamiltonian generating geodesic flow on the Heisenberg group (which is also our kinetic energy, K). In other words, let

$$H_g(q,p) = \frac{1}{2}(P_X^2 + P_Y^2) = \frac{1}{2}(p,p),$$

where $(\cdot,\cdot)$ is the cometric induced by the inner product $\langle \cdot,\cdot \rangle$. Then let

$$S(q,t) = \inf \int_0^t L_g(\gamma(t), \dot{\gamma}(t))dt = \inf \int_0^t \tfrac{1}{2}||\dot{\gamma}(s)||_{\mathbb{H}}^2 ds,$$

where the infimum is taken over all paths γ connecting 0 to q in time t. Here, the Lagrangian L_g is related to the Hamiltonian H_g by the Legendre transform. Note that the function S is known as *Hamilton's generating function* or the *action* in mechanics, and the *value function* in optimal control. Then the Hamilton-Jacobi equation (see [1] or [3]) says

$$H_g(q, dS) = -\frac{\partial S}{\partial t} \tag{3.12}$$

at all points (q,t) where S is differentiable. As the next Lemma shows, S is differentiable at points (q,t) where $t \neq 0$ and the function $f(q) = d(q,0)$ is smooth at q. The latter holds if, for the minimizing geodesic γ connecting q to the origin, γ contains no conjugate or cut points. In the Heisenberg group, there are no cut points, and the locus of points conjugate to the origin consists of the z-axis (see [8]). Thus, the Hamilton-Jacobi equation (3.12) holds almost everywhere: for all points (q,t) such that $t \neq 0$ and q does not lie on the z-axis.

Lemma 31 *We can express the generating function S as*

$$S(q,t) = \frac{||q||_{sR}^2}{2t}.$$

Proof Suppose $\gamma\colon [0,t] \to \mathbb{H}$, with $\gamma(0) = 0$ and $\gamma(t) = q$. Then the Cauchy-Schwarz inequality gives

$$\int_0^t ||\dot{\gamma}(s)||_{\mathbb{H}} ds \leq \sqrt{\int_0^t ||\dot{\gamma}(s)||_{\mathbb{H}}^2 ds}\sqrt{\int_0^t ds},$$

with equality if and only if the speed $||\dot{\gamma}(s)||_{\mathbb{H}}$ is constant. We recognize the left-hand side as the length $l(\gamma)$. Now suppose γ has constant speed, so we may rewrite this (in)equality as

$$l(\gamma) = \sqrt{t}\,\sqrt{\int_0^t ||\dot{\gamma}(s)||_{\mathbb{H}}^2 ds},$$

or,

$$\int_0^t \tfrac{1}{2}||\dot{\gamma}(s)||_{\mathbb{H}}^2 ds = \frac{l(\gamma)^2}{2t}.$$

Finally, taking the infimum (over all such γ) of both sides yields the desired result.

□

Lemma 32 *Suppose* $\gamma \colon [0,T] \to \mathbb{H}$ *is horizontal and* $\gamma(0) = 0$*. Then*

$$\frac{d}{dt}||\gamma(t)||_{sR} \leq ||\dot{\gamma}(t)||_{\mathbb{H}}$$

almost everywhere.

Proof Let $R(q) = ||q||_{sR}$ for sake of notation. Then, by Lemma 31, we have

$$S(q,t) = \frac{R^2(q)}{2t},$$

so that

$$dS = \frac{R(q)dR}{t}$$

and

$$\frac{\partial S}{\partial t} = -\frac{R^2(q)}{2t^2}.$$

Choosing (q,t) such that $R(q) = t$ implies $dS = dR$ and

$$\frac{\partial S}{\partial t} = -\frac{1}{2}.$$

Thus, the Hamilton-Jacobi equation reads[3]

$$\frac{1}{2}(dR, dR) = \frac{1}{2},$$

which is equivalent to

$$||\nabla_{sR} R||_{\mathbb{H}} = 1,$$

[3] An optimal control theoretic version of this result can be found in Chapter 1, Section 9, of [28].

and this holds almost everywhere. Finally, we can employ the chain rule and the Cauchy-Schwarz inequality in the Heisenberg group to obtain:

$$\begin{aligned}\frac{d}{dt}||\gamma(t)||_{sR} &= \frac{d}{dt}R(\gamma(t))\\ &\leq |\frac{d}{dt}R(\gamma(t))|\\ &= |\langle \nabla_{sR}R(\gamma(t)), \dot{\gamma}(t)\rangle|\\ &\leq ||\nabla_{sR}R(\gamma(t))||_{\mathbb{H}}||\dot{\gamma}(t)||_{\mathbb{H}}\\ &= ||\dot{\gamma}(t)||_{\mathbb{H}}\end{aligned}$$

almost everywhere. □

Proposition 33 *Suppose $\gamma \in \mathcal{F}_k$ and $\gamma(0) = 0$. Then $A(\gamma) = \infty$.*

Proof Recall that $\rho = ((x^2+y^2)^2 + \frac{1}{16}z^2)^{1/4}$, so that $U = \rho^{-2}$. Since the Heisenberg sphere is homeomorphic to the Euclidean sphere, the standard argument which shows that any two norms on $\mathbb{R}^n$ are Lipshitz equivalent shows that ρ and $||\cdot||_{sR}$ are Lipshitz equivalent: there exist positive constants c and C such that

$$c\rho(x, y, z) < ||(x, y, z)||_{sR} < C\rho(x, y, z)$$

for $(x, y, z) \neq 0$. Then, using this fact and the general fact that $a^2 + b^2 \geq 2ab$, we compute:

$$\begin{aligned}A(\gamma) &= \int_\gamma L\\ &= \int_\gamma \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) + \rho^{-2}\\ &\geq \int_0^T \left(\tfrac{1}{2}||\dot{\gamma}||_{\mathbb{H}}^2 + \frac{c_1}{||\gamma||_{sR}^2}\right)dt\\ &\geq c_2 \int_0^T \frac{||\dot{\gamma}||_{\mathbb{H}}}{||\gamma||_{sR}}dt.\end{aligned}$$

But this last integrand is non-negative, so the value of its integral decreases when taken over a sub-interval of $[0, T]$. In particular, the value of the integral is smaller over the interval $[0, \epsilon]$ for small ϵ. This fact, along with Lemma 32, implies

$$\begin{aligned}
A(\gamma) &\geq c_2 \int_0^T \frac{||\dot{\gamma}||_{\mathbb{H}}}{||\gamma||_{sR}} dt \\
&\geq c_2 \int_0^\epsilon \frac{||\dot{\gamma}||_{\mathbb{H}}}{||\gamma||_{sR}} dt \\
&\geq c_3 \int_0^\epsilon \frac{\frac{d}{dt}||\gamma||_{sR}}{||\gamma||_{sR}} dt \\
&= c_3 \int_0^{u(\epsilon)} \frac{du}{u} \\
&= \lim_{a \to 0} c_3 \log u|_a^{u(\epsilon)} \\
&= \infty.
\end{aligned}$$

Here, we have made the substitution $u(t) = ||\gamma(t)||_{sR}$ with $u(0) = 0$. □

Corollary 34 *Our curve γ_* does not suffer a collision.*

Proof From Proposition 30, we know $A(\gamma_*)$ equals the infimum of the action restricted to $\mathcal{F}_k$, which is finite. □

3.3.5 A Critical Point of the Action

Lemma 35 *The action functional A is differentiable at any curve which avoids collision.*

Proof We first consider A as a functional on the space of all (sufficiently smooth) paths in $\mathbb{R}^3 \cong \mathbb{H}$. If A is differentiable there, it then follows that A is differentiable on the (infinite codimension) subspace $\mathcal{F}_k$. To this end, let $c, e \in H^1([0,T], \mathbb{R}^3)$, and let $h \in \mathbb{R}$. Recall that $\langle \cdot, \cdot \rangle$ denotes the sub-Riemannian inner product, which obeys the so-called dot product rule

$$\frac{d}{d\epsilon}\langle c_1(\epsilon), c_2(\epsilon)\rangle = \langle \frac{d}{d\epsilon} c_1(\epsilon), c_2(\epsilon)\rangle + \langle c_1(\epsilon), \frac{d}{d\epsilon} c_2(\epsilon)\rangle.$$

(The proof is just as in $\mathbb{R}^2$.) Also let $\langle \cdot, \cdot \rangle_E$ denote the Euclidean ($\mathbb{R}^3$) dot product, and ∇ denote the Euclidean gradient operator (not to be confused with ∇_{sR}, the sub-Riemannian gradient). Finally, we may compute

$$\begin{aligned} dA(c)(e) &= \frac{d}{dh}A(c+he)|_{h=0} \\ &= \frac{d}{dh}\Big|_{h=0}\int_0^T \left(\tfrac{1}{2}\langle \dot{c}(t)+h\dot{e}(t), \dot{c}(t)+h\dot{e}(t)\rangle + U(c(t)+he(t)\right)dt \\ &= \int_0^T \langle \dot{c}(t), \dot{e}(t)\rangle + \langle \nabla U(c(t)), e(t)\rangle_E dt, \end{aligned} \tag{3.13}$$

where we used the chain rule for the composition

$$[0,T] \xrightarrow{e} \mathbb{R}^3 \xrightarrow{U} \mathbb{R}.$$

The expression (3.13) shows that $dA(c)$ exists and is indeed continuous, so long as c does not pass through the origin, where U has a singularity. □

Proposition 36 *We have that $dA(\gamma_*)(e) = 0$ for any $e \in \mathcal{F}_k$.*

Proof Computing the derivative of the functional $A\colon \mathcal{F}_k \to \mathbb{R}$, for any $c \in \mathcal{F}_k$ we again have

$$dA(c)(e) = \frac{d}{dh}A(c+he)|_{h=0}$$

Now we shall consider this derivative taken at the point $\gamma_* \in \mathcal{F}_k$. Recall that, according to Corollary 34, γ_* does not pass through the origin, so by the previous Lemma, $dA(\gamma_*)$ exists and is continuous. Suppose for sake of contradiction that $\exists e \in \mathcal{F}_k$ such that $dA(\gamma_*)(e) \neq 0$. Replacing e with $-e$ if necessary, we can assume $dA(\gamma_*)(e) < 0$. As this is the derivative of the real-valued function $f(h) = A(\gamma_* + he)$ at $h = 0$, this function is initially decreasing. That is, for sufficiently small $h > 0$ we have $A(\gamma_* + he) < A(\gamma_*)$. As $\gamma_* + he$ is in our space $\mathcal{F}_k$, this contradicts the fact that γ_* is a local minimum of $A|_{\mathcal{F}_k}$. □

To combine the results obtained thus far, we need the following lemma, due to R. Palais ([27]).

Lemma 37 (Principle of Symmetric Criticality) *Let Γ be a finite group acting on Hilbert space V and let V^Γ denote the fixed points of this action. Suppose $f\colon V \to \mathbb{R}$ is Γ-invariant, and that $f|_{V^\Gamma}$ has a critical point at $x_0 \in V$. Then x_0 is also a critical point for f.*

Proof We know V^Γ is an invariant subspace of V, and we can choose a metric on V so that the group Γ acts orthogonally. Now, by Maschke's theorem, V^Γ admits a Γ-invariant orthogonal complement W, so that $V = V^\Gamma \oplus W$. Then we can decompose

$$\nabla f(x_0) = u + w,$$

with $u \in V^\Gamma$ and $w \in W$. But x_0 is a critical point of $f|_{V^\Gamma}$, so $u = 0$. Then for any $g \in \Gamma$, since x_0 is fixed and f is invariant, we have

$$w = \nabla f(x_0) = \nabla f(g \cdot x_0) = g \cdot \nabla f(x_0) = g \cdot w.$$

This shows w must be fixed, but as w was in the direct sum complement of the fixed points, w must be zero. Thus, $\nabla f(x_0) = 0$. □

Proposition 38 *We have $dA(\gamma_*)(e) = 0$ for any horizontal $e \in H^1([0,T], \mathbb{H})$.*

Proof To apply this Lemma to our situation, we take V to be the space of all horizontal paths $e \in H^1([0,T], \mathbb{H})$, with the Hilbert structure given in Section 3.3.1; f is the action functional and x_0 is the curve γ_*. We take Γ to be the group $\mathbb{Z}/\mathbb{Z}_2 \times \mathbb{Z}/\mathbb{Z}_k$, $k \geq 3$ an odd integer, whose action is given as follows[4]:

$$\begin{aligned}
\mathbb{Z}/\mathbb{Z}_2 \times \mathbb{Z}/\mathbb{Z}_k &= \langle (\sigma, \tau) \mid (\sigma^2, \tau^k) = e \rangle \\
\sigma \cdot (x(t), y(t), z(t)) &= (x(t - T/2), y(t - T/2), -z(t - T/2)) \\
\tau \cdot (x(t), y(t), z(t)) &= R_{2\pi/k}(x(t - T/k), y(t - T/k), z(t - T/k))
\end{aligned}$$

where again

$$R_{2\pi/k} = \begin{bmatrix} \cos(2\pi/k) & -\sin(2\pi/k) & 0 \\ \sin(2\pi/k) & \cos(2\pi/k) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then V^Γ is precisely the function space $\mathcal{F}_k$. In other words, to say that $x_0 \in V^\Gamma$ is to say that γ_* satisfies the symmetry assumptions (S1) and (S2). As we have shown that the action, restricted to $\mathcal{F}_k$, has a critical point at γ_*, the Principle of Symmetric Criticality implies that γ_* is indeed a critical point of the action functional where admissible variations need not satisfy (S1) nor (S2). □

[4]An additional application of this argument allows us to restrict our attention to periodic orbits, so that the domains of these curves are well-defined.

3.3.6 Satisfaction of the Equations of Motion

As shown in the previous section, γ_* is a critical point of the action functional restricted to horizontal variations. According to the Principal of Least Action, γ_* should satisfy the equations of motion. More precisely, we will employ the standard argument from the calculus of variations (see [6, 7, 13, 16, 33]): invoke the method of Lagrange multipliers, integrate by parts, then apply the fundamental lemma of the calculus of variations. This shows that γ_* satisfies the Euler-Lagrange equations from Section 2.2.3, which were apparently equivalent to Hamilton's equations.

Recall that, as in Section 2.2.3, our horizontal constraint is precisely the zero set of the function

$$G(t, q, \dot{q}) = \tfrac{1}{2}x\dot{y} - \tfrac{1}{2}y\dot{x} - \dot{z}$$

and our modified action functional is

$$A_\lambda(\gamma) = \int_0^T L_\lambda(t, \gamma, \dot{\gamma})dt,$$

where $\lambda = \lambda(t)$ is a scalar and we have written $L_\lambda(t, q, \dot{q}) = L(t, q, \dot{q}) - \lambda(t)G(t, q, \dot{q})$.

Lemma 39 (Lagrange multipliers) *If c is a critical point of the action A restricted to horizontal curves, then there exists some $\lambda = \lambda(t)$ such that c is a critical point of A_λ.*

Proof This is a classical result whose various proofs may be found in, for example, [6], Section 39 of [7], Section 12 of [13], Section IV(e) of [16], or Volume II of [33].

□

Proposition 40 *Our γ_* satisfies the equation*

$$\frac{\partial L_\lambda}{\partial \gamma_*} - \frac{d}{dt}\left(\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\right) = 0$$

for some λ.

Proof According to Lemma 39, γ_* is a critical point of A_λ for some λ which we now fix. Let $q_\epsilon(t) = \gamma_*(t) + \epsilon\eta(t)$, for some η which is twice differentiable and satisfies the periodicity condition $\eta(0) = \eta(T)$. Let

$$A_{\epsilon,\lambda} := \int_0^T L_{\epsilon,\lambda}dt := \int_0^T L_\lambda(t, q_\epsilon, \dot{q}_\epsilon)dt.$$

By the chain rule

$$\begin{aligned}\frac{d}{d\epsilon}L_{\epsilon,\lambda} &= \frac{\partial L_{\epsilon,\lambda}}{\partial t}\frac{dt}{d\epsilon} + \frac{\partial L_{\epsilon,\lambda}}{\partial q_\epsilon}\frac{dq_\epsilon}{d\epsilon} + \frac{\partial L_{\epsilon,\lambda}}{\partial \dot{q}_\epsilon}\frac{d\dot{q}_\epsilon}{d\epsilon} \\ &= \frac{\partial L_{\epsilon,\lambda}}{\partial q_\epsilon}\eta(t) + \frac{\partial L_{\epsilon,\lambda}}{\partial \dot{q}_\epsilon}\dot{\eta}(t),\end{aligned}$$

so that

$$\frac{d}{d\epsilon}L_{\epsilon,\lambda}\bigg|_{\epsilon=0} = \frac{\partial L_\lambda}{\partial \gamma_*}\eta(t) + \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\dot{\eta}(t).$$

Since γ_* is a critical point of A_λ, then $A_{\epsilon,\lambda}$, considered a function of ϵ, will have a critical point when $\epsilon = 0$. So we have

$$0 = \frac{d}{d\epsilon}A_{\epsilon,\lambda}\bigg|_{\epsilon=0} = \int_0^T \left(\frac{\partial L_\lambda}{\partial \gamma_*}\eta(t) + \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\dot{\eta}(t)\right) dt.$$

Temporarily assuming that γ_* is twice differentiable, we can integrate the second term by parts:

$$\begin{aligned}\int_0^T \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\dot{\eta}(t)dt &= \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\eta(t)\bigg|_0^T - \int_0^T \frac{d}{dt}\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\eta(t)dt \\ &= \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(T)\eta(0) - \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(0)\eta(0) - \int_0^T \frac{d}{dt}\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\eta(t)dt,\end{aligned}$$

where we used the periodicity assumption on η. Thus, we have

$$0 = \int_0^T \left(\frac{\partial L_\lambda}{\partial \gamma_*}\eta(t) - \frac{d}{dt}\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\eta(t)\right) dt + \left(\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(T) - \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(0)\right)\eta(0).$$

Then the fundamental lemma of the calculus of variations (see Section 7 of [33] for a version which suits our needs) implies that our smoothness assumption on γ_* was justified, and that, choosing suitable test functions η, we must have

$$\frac{\partial L_\lambda}{\partial \gamma_*} - \frac{d}{dt}\left(\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}\right) = 0 \tag{3.14}$$

and

$$\frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(T) = \frac{\partial L_\lambda}{\partial \dot{\gamma}_*}(0). \tag{3.15}$$

□

Then (3.14) yields the version of the Euler-Lagrange equations given in (2.4) which were seen to be equivalent to Hamilton's equations. Thus, γ_* satisfies the equations of motion. Also, (3.15) simplifies to the three equations

$$p_x(0) = p_x(T)$$
$$p_y(0) = p_y(T)$$
$$p_z(0) = p_z(T),$$

where $\gamma_*(t) = (x(t), y(t), z(t), p_x(t), p_y(t), p_z(t))$. This guarantees that our curve γ_* is periodic in all of phase space, not just in configuration space.

3.4 Open Problems

In this final section, we collect a list of future problems deserving of investigation, and questions we have failed to answer.

(i) Is the Kepler-Heisenberg problem an integrable system? More specifically, are the dynamics integrable for the non-restricted ($H \neq 0$) system?

(ii) Can we explicitly integrate the equations of motion for the restricted ($H = 0$) system? Can we parametrize the family of curves presented in Section 3.2.2?

(iii) According to the proof of Theorem 2, there exist periodic solutions with k-fold rotational symmetry for any odd integer $k \geq 3$. Can we find them numerically? What do they look like?

(iv) We know that orbits with negative energy are bounded but not periodic. Do they always tend towards collision? Similarly, we know that orbits with positive energy are unbounded in phase space. Are they always asymptotic to Heisenberg geodesics? Is there an open set of initial conditions whose orbits are asymptotic to Heisenberg geodesics? Numerical experiment suggests that the answers to these questions are affirmative.

(v) Is there a sub-Riemannian "Newton's equation" analogous to the famous equation $\ddot{\gamma} = -\nabla U(\gamma)$? Presumably the Euclidean gradient would be replaced by the sub-Riemannian gradient. Is there a natural connection whereby the left hand side could be replaced by $\nabla_{\dot{\gamma}}\dot{\gamma}$ or something similar?

Part II

The Puiseux Characteristic of a Goursat Germ

Chapter 4

Introduction

4.1 History and Motivation

In this part we give a formula for the Puiseux characteristic of an analytic plane curve germ which represents a Goursat distribution germ with prescribed small growth vector. It has been shown earlier that the Puiseux characteristic and the small growth vector are equivalent data, as both are equivalent to a geometric stratification of Goursat germs known as RVT coding. (Puiseux characteristics do not cover the entire RVT stratification, but the difference is secondary.) However, the existing algorithms for calculating one from another are cumbersome and recursive. Here, we effectively compose two known algorithms and derive a tidy formula which greatly simplifies the existing methods. In addition, the Puiseux characteristic is a much more compact labelling than the small growth vector or the RVT code, and our formula allows for a quick conversion to this more convenient labelling. The problem solved herein was first proposed in [22] as Question 9.19, part 3, and was asked again in [26] in the Afterword. An earlier version of this work can be found in [30].

Goursat distributions are located in the antipodes of integrable distributions, as they are *bracket-generating*. Cartan ([9]) studied the model of the canonical contact distribution on the jet space $J^k(\mathbb{R}, \mathbb{R})$. All Goursat distributions were believed to be equivalent to Cartan's until Giaro, Kumpera, and Ruiz discovered the first singularity in 1978 ([14]).

Jean ([18]) studied the kinematic model of a car pulling N trailers, a system which is locally universal for Goursat distributions of corank $N+1$. He developed a geometric stratification given by regions in the configuration space of the model in terms of critical angles. He also derived recurrence relations enabling one to compute the small growth vector of a Goursat germ from these geometric strata. Montgomery and Zhitomirskii ([23]) introduced the Monster tower, a sequence of manifolds with distributions in which every Goursat germ occurs, as well as the Sandwich Lemma, allowing for Jean's strata to be recast in terms of positions of members of a canonical subflag of the Goursat flag. Mormul ([25]) labelled the strata from [23] by words in the letters GST, which became the RVT code in [22]. He solved Jean's relations in terms of the derived vector of the small growth vector, allowing for the calculation of the RVT code from the small growth vector. In [22], Montgomery and Zhitomirskii showed that Goursat germs correspond to finite jets of Legendrian curve germs, and that the RVT coding corresponds to the classical invariant in the singularity theory of planar curves: the Puiseux characteristic (see Section 4.2.3 below). They gave a recursive algorithm for computing the Puiseux characteristic from the RVT code.

In short, the present contribution provides the dashed arrow in the following diagram:

$$\begin{tikzcd} \{\mathrm{SGV}\} \arrow[r, shift left] \arrow[dr, dashed] & \{\mathrm{RVT}\} \arrow[l, shift left] \arrow[d, leftrightarrow] \\ & \{\mathrm{PC}\} \end{tikzcd}$$

Here, the arrow $\{\mathrm{RVT}\} \longleftrightarrow \{\mathrm{PC}\}$ was given in [22], and the arrow $\{\mathrm{SGV}\} \longrightarrow \{\mathrm{RVT}\}$ was given in [25]. The arrow $\{\mathrm{RVT}\} \longrightarrow \{\mathrm{SGV}\}$ was given recursively in [18], and later explicitly in [25].

It is worth noting that now, with both mappings $\{\mathrm{RVT}\} \longrightarrow \{\mathrm{SGV}\}$ and $\{\mathrm{SGV}\} \dashrightarrow \{\mathrm{PC}\}$ having explicit presentations, the recursive mapping Pc of [22] (see Section 4.2.4) has become explicit as well.

4.2 Background

4.2.1 Goursat distributions and small growth vectors

Let M be a smooth manifold and let $D \subset TM$ be any smooth distribution (subbundle). Let $[D, D] + D$ be called the *Lie square* of D. Iterate this squaring to obtain a chain, which we write as

$$D^s \subseteq D^{s-1} \subseteq \cdots \subseteq D^i \subseteq D^{i-1} \subseteq \cdots$$

where $D^s = D$ and D^{i-1} is the Lie square of D^i. Note that D^i may not, in general, have constant rank, and thus fail to be a distribution on M.

A rank 2 distribution D on a manifold M of real dimension $n = s + 2 \geq 4$ is called *Goursat* if corank $D^i = i$ for $i = 0, \ldots, s$. In this case, one has a *Goursat flag* $\mathcal{F}$:

$$D^s \subset D^{s-1} \subset \cdots \subset D^1 \subset D^0 = TM.$$

Note that when D is Goursat, each member D^i of the flag is itself a distribution, and a hyperplane in D^{i-1}.

Given a Goursat distribution D, one can alternatively form the sequence $D_i = [D, D_{i-1}] + D_{i-1}$, where $D_1 = D$ and $i \geq 1$. It is not hard to show that this sequence will also eventually terminate. That is, there exists an r such that $D_r = TM$. Thus, Goursat distributions are *completely nonholonomic.* For any $p \in M$, the least r such that $D_r(p) = T_pM$ is called the *degree of nonholonomy* of D at p. Note that for a general completely nonholonomic distribution, the degree of nonholonomy depends on the base point p, and so it happens for the Goursat distributions. For each $p \in M$, we define the *small growth vector at p* to be the integer valued vector

$$sgv(p) = \big(\mathrm{dim}D_1(p),\ \mathrm{dim}D_2(p), \ldots,\ \mathrm{dim}D_{r(p)}(p) = n\big).$$

In the following we shall only be concerned with germs of Goursat distributions.

While the small growth vector is the traditional object of interest, for completely nonholonomic distributions it is equivalent to the *derived vector*, which will be more convenient for us to work with.

Definition 4 ([25]) *The* derived vector *of a completely nonholonomic distribution germ consists of the multiplicities of the entries in the small growth vector at the reference point.*

For a Goursat distribution, the dimensions of the sequence D_i grow by at most one at a time, so the multiplicities (the entries in the derived vector) are nonzero. By convention, we omit the last multiplicity 1 from the derived vector. For example, if we are given a small growth vector (2, 3, 4, 4, 5), then the corresponding derived vector is (1, 1, 2). Similarly, given a derived vector (1, 1, 1, 3, 3), the corresponding small growth vector is (2, 3, 4, 5, 5, 5, 6, 6, 6, 7). While it is not obvious, for Goursat distributions the derived vector is always non-decreasing (see Section 2 of [25]).

4.2.2 Construction of the RVT code

An RVT code is a word in the letters R,V,T satisfying one spelling rule: the letter T cannot follow the letter R. An RVT code represents an equivalence class of Goursat germs. The construction of a Goursat germ's RVT code was implicit in [23], and made explicit in [25], where the letters G,S,T were used instead of R,V,T. Beginning with a Goursat germ, one forms the Goursat flag, which possesses a canonical integrable subflag called the *characteristic foliation* or *Cauchy characteristic.* The geometric relationship between the members of the two flags can be characterized as Regular, Vertical, or Tangent (or, alternatively, Generic, Singular, or Tangent) and one encodes this information into a word called the RVT code. See Section 1.2 of [25] for details.

In [22], a parallel definition of the RVT codes for Goursat germs was proposed using a tower of manifolds called the "Monster Tower." This tower is Goursat universal: every Goursat germ occurs somewhere within the Monster. Each point in the Monster Tower is assigned an RVT code, and the code of a Goursat germ at a reference point p is that of p itself.

The tower is constructed through a series of *Cartan prolongations.* Begin with the manifold $M^0 = \mathbb{R}^2$ and the distribution $\Delta^0 = T\mathbb{R}^2$. The first prolongation is the fiber bundle

$$M^1 = \bigcup_{p \in \mathbb{R}^2} \mathbb{P}\Delta^0_p,$$

whose elements have the form (p, l), where p is a point in $\mathbb{R}^2$ and l is a line in the tangent space $T_p\mathbb{R}^2$. The distribution on M^1 is given by

$$\Delta^1_{(p,l)} = (d\pi^1_0)^{-1}(l)$$

where $\pi^1_0 \colon M^1 \to M^0$ is the bundle projection.

Iterating the prolongation procedure gives a sequence of manifolds

$$M^i = \bigcup_{p \in M^{i-1}} \mathbb{P}\Delta^{i-1}_p.$$

Every point in M^i has the form (p, l), where p is a point in M^{i-1} and l is a line in the distribution plane Δ^{i-1}_p. The dimension of M^i is thus $i + 2$. The bundle projection map $\pi^i_{i-1} \colon M^i \to M^{i-1}$ has fibers diffeomorphic to $\mathbb{P}\Delta^{i-1}_p \cong \mathbb{RP}^1 \cong S^1$. The distribution on M^i is given by

$$\Delta^i_{(p,l)} = (d\pi^i_{i-1})^{-1}(l).$$

One verifies that each distribution Δ^i is rank 2 and Goursat. The *Monster Tower* is thus the sequence of circle bundles

$$\cdots \to M^i \to M^{i-1} \to \cdots \to M^1 \to M^0 = \mathbb{R}^2$$

equipped with a Goursat distribution at each level.

By composing the projection maps π^k_{k-1}, $\pi^{k-1}_{k-2}, \ldots, \pi^{i+1}_i$ we obtain projections $\pi^k_i : M^k \to M^i$, $i < k$. The horizontal curves at level i naturally prolong (i.e., lift) to horizontal curves at level k. However, the curves coinciding with fibers of π^{i+1}_i, $i \geq 1$, are special – they project down to points and are not prolongations of curves from below. They are called *vertical* and can themselves be prolonged to (first order) tangency curves, then prolonged again to (second order) tangency curves, and so on. Vertical curves and their prolongations are called *critical.* Thus, at each level $i \geq 2$ there are vertical directions, and, in addition, at each level $i \geq 3$ there are tangency directions different from the vertical direction. At any level, all the remaining (non-critical) horizontal directions are called *regular.*

Recall that a point p at level i has the form (q, l) where $q \in M^{i-1}$ and l is a line in Δ^{i-1}. We call p a *regular, vertical,* or *tangency* point if the direction of l is regular, vertical, or tangency, respectively. Points which are vertical or tangency are called *critical.* Therefore, at each level $i \geq 3$ there are regular and vertical points, and at each level $i \geq 4$ there are regular, vertical, and tangency points.

Now, the RVT code of a point p at level $k \geq 3$ is a word $(\omega_3\omega_4 \dots \omega_k)$ in the letters R, V, T satisfying

$$\omega_i = \begin{cases} R & \text{if } \pi_i^k(p) \text{ is a regular point,} \\ V & \text{if } \pi_i^k(p) \text{ is a vertical point,} \\ T & \text{if } \pi_i^k(p) \text{ is a tangency point.} \end{cases}$$

It follows that the only spelling rule for the codes of points is the absence of the sequence 'RT' in the codes; all other codes are realizable. Thus, the general form of an RVT code of a point is R^{k-2} (not dealt with in the present paper), or else

$$R^{r_{v+1}} V T^{t_v} R^{r_v} \cdots V T^{t_2} R^{r_2} V T^{t_1} R^{r_1}. \tag{4.1}$$

Let us pause momentarily to explain the notation in (4.1). Here, v is the number of letters V in the RVT code. We have thus partitioned our code into $v + 1$ pieces separated by the letters V. We write that the *last* letter V in the code is followed by t_1 many letters T, and then r_1 many letters R. We continue for $1 \leq i \leq v$ letting t_i denote the number of letters T following the ith V *from the right*, and r_i denote the number of letters R following those letters T. Let r_{v+1} denote the number of letters R preceding[1] the first letter V. Finally, superscripts in (4.1) denote multiplicities.

The germ of a horizontal curve passing through a reference point p is called *regular* when it is immersed and tangent neither to the vertical nor tangency direction at p. It is a central and deep fact in [22] that each germ of a non-constant well-parametrized analytic plane curve becomes regular after finitely many prolongations and stays regular in subsequent prolongations. The least number of prolongations needed to regularize the curve is called the *regularization level* k, and the k-fold prolongation of the original curve γ is called the *regularization* of γ. This result bridges two seemingly distant areas: Goursat geometry and the singularity theory of plane curves. In [22], the first of two proofs of this deep fact is based on the Puiseux characteristic of a singular plane curve. It is worth noting that the reference point at level k, hit by the regularized curve, is still critical. Only its 'daughter' at level

[1]Note that here, as in [22], we allow $r_{v+1} \geq 0$. In [25], where the letters R,V,T are replaced by G,S,T, respectively, the code always begins with two letters G. Thus, the GST codes in [25] are two letters longer than the RVT codes here and in [22].

$k+1$ (hit by the prolongation of the newly regularized curve) and her daughters will be regular.

Now the germ of a plane curve γ with regularization level $k \geq 3$ is assigned an RVT code as well: it is the code of the (critical) reference point hit by the regularization of γ. Therefore, the codes of plane curves always end with V or T; such codes are called *critical* (codes consisting solely of letters V and T are called *entirely critical*). That is, in the notation from (4.1), the equation $r_1 = 0$ holds for the codes of curves. The mapping Pc of [22], one of the key players in the present work, acts precisely on the critical codes.

Lastly, note that if the original curve germ, or its first prolongation, is already immersed, then the curve's code is undefined. The further prolongations of such curves hit (and completely exhaust) the simplest points of the Monster: the so-called Cartan points, or jet-like points (see Section 4). At each level $k \geq 2$ these points populate the only open dense stratum R^{k-2} (the entire M^2 when $k = 2$) which remains outside the field of interest of the present paper.

We now recall the construction of the RVT code from the derived vector (see [25]). Suppose we are given a Goursat germ whose derived vector (see Definition 4) is

$$der = (\underbrace{M_1,\ M_1, \ldots, M_1}_{m_1},\ \underbrace{M_2,\ M_2, \ldots, M_2}_{m_2}, \ldots, \underbrace{M_{v+1},\ M_{v+1}, \ldots, M_{v+1}}_{m_{v+1}}),$$

with $M_1 < M_2 < \cdots < M_v < M_{v+1}$.

Then v turns out to be the number of letters V in the RVT code of this germ, which, therefore, has the form of (4.1):

$$R^{r_{v+1}} V T^{t_v} R^{r_v} \cdots V T^{t_2} R^{r_2} V T^{t_1} R^{r_1}.$$

Mormul derived the following relations for ascertaining the multiplicities r_i and t_i. See Theorem 3.5 in [25]. One has:

$r_{v+1} = m_{v+1} - 1$

$t_1 = M_2 - 2$

$r_1 = m_1 - M_2 \geq 0.$

For $2 \leq i \leq v$ we have:

Case 1: M_i divides M_{i+1}. Then

$$t_i = \frac{M_{i+1}}{M_i} - 2 \tag{4.2}$$

$$r_i = m_i - t_i - 1 > 0. \tag{4.3}$$

Case 2: M_i does not divide M_{i+1}. Then

$$t_i = m_i - 1 \tag{4.4}$$

$$r_i = 0. \tag{4.5}$$

4.2.3 The Puiseux characteristic

Suppose $\gamma\colon (\mathbb{R}, 0) \to \mathbb{R}^2$ is the parametrization of an analytic plane curve germ. We say γ is *badly-parametrized* if there exist analytic germs $\mu\colon (\mathbb{R}, 0) \to \mathbb{R}^2$ and $\phi\colon (\mathbb{R}, 0) \to (\mathbb{R}, 0)$ such that $d\phi/dt(0) = 0$ and $\gamma = \mu \circ \phi$. Otherwise, γ is called *well-parametrized*. If γ is well-parameterized and not immersed then we may define its Puiseux characteristic, an invariant with respect to the RL-equivalence of curve germs. Up to RL-equivalence, γ has the form

$$\gamma(t) = \Big(t^m, \sum_{k \geq m} a_k t^k\Big)$$

where $m \geq 2$.

The definition of the Puiseux characteristic is the following. Let $\lambda_0 = e_0 = m$. Then define inductively for $j \geq 0$

$$\lambda_{j+1} = \min\{k \mid a_k \neq 0,\ e_j \nmid k\}, \qquad e_{j+1} = \gcd(e_j, \lambda_{j+1})$$

until we first obtain a g with $e_g = 1$. Then the vector $[\lambda_0; \lambda_1, \ldots, \lambda_g]$ is called the *Puiseux characteristic* of γ.

The Puiseux characteristic is the fundamental invariant in the singularity theory of plane curves. In [32], Proposition 4.3.8 shows that it is equivalent to at least seven other classical invariants.

Here, as in [22], we restrict our attention to Puiseux characteristics satisfying

$$\lambda_1 > 2\lambda_0. \tag{4.6}$$

This is a normalization condition, and no equivalence classes of Legendrian curves are excluded by its imposition. The Puiseux characteristic is an invariant with respect to RL-equivalence, but not a complete invariant. For example, (t^2, t^5) and (t^3, t^5) have different Puiseux characteristics, but their first prolongations are equivalent Legendrian curves. The restriction (4.6) resolves this ambiguity.

4.2.4 The map Pc

Here we recall the definition of the map Pc constructed in Section 3.8.4 of [22]. Given a critical RVT code (α), this map yields a Puiseux characteristic $\mathrm{Pc}(\alpha)$ satisfying (4.6). The relationship between (α) and $\mathrm{Pc}(\alpha)$ is given in Theorem 3.23 of [22]. The map is constructed recursively as follows.

First define the two maps

$$\begin{aligned}\mathbb{E}_T\colon (n_1, n_2) &\mapsto (n_1, n_1 + n_2)\\ \mathbb{E}_V\colon (n_1, n_2) &\mapsto (n_2, n_1 + n_2).\end{aligned}$$

Then, for an entirely critical code

$$(\omega) = (\omega_1, \ldots, \omega_m), \quad \omega_i \in \{V, T\}$$

define $\mathbb{E}_\omega$ to be the composition

$$\mathbb{E}_\omega = \mathbb{E}_{\omega_1} \circ \mathbb{E}_{\omega_2} \circ \cdots \circ \mathbb{E}_{\omega_m}.$$

Next, we note that any critical RVT code (α) has one of the two following forms:

A. $(\alpha) = (R^s\omega)$, where $s \geq 0$ and (ω) is an entirely critical RVT code;

B. $(\alpha) = (\beta R^s \omega)$, where $s \geq 1$, (β) is a critical RVT code, and (ω) is an entirely critical RVT code.

In case A, let $(a, b) = \mathbb{E}_\omega(1, 2)$. Then

$$\mathrm{Pc}(\alpha) = [\lambda_0; \lambda_1], \quad \lambda_0 = a,\ \lambda_1 = sa + a + b.$$

In case B, let $(a, b) = \mathbb{E}_\omega(1, 2)$ and $\mathrm{Pc}(\beta) = [\tilde{\lambda}_0; \tilde{\lambda}_1, \ldots, \tilde{\lambda}_{g-1}]$. Then

$$\mathrm{Pc}(\alpha) = [\lambda_0; \lambda_1, \ldots, \lambda_g],$$

where

$$\lambda_i = a\tilde{\lambda}_i \quad \text{for} \quad 0 \leq i \leq g-1$$
$$\lambda_g = a(\tilde{\lambda}_{g-1} + s - 1) + b - a.$$

Equipped with the tools developed in this section, we are now prepared to state and prove our theorem.

Chapter 5

Main Result

As explained in Section 1, here we compose the formulas presented in [25] and the algorithm in [22], yielding a formula for the Puiseux characteristic of the plane curve corresponding to a Goursat germ with given small growth vector. This formula turns out to be simpler than either of the two from which it was derived, suggesting a deeper geometric link between singularities of plane curves and singular Goursat distributions. The problem solved herein was first proposed in [22] as Question 9.19, part 3, and was asked again in [26] in the Afterword.

5.1 Main Theorem

Suppose we are given a Goursat germ whose derived vector (see Definition 4) is

$$der = (\underbrace{M_1,\ M_1, \ldots, M_1}_{m_1},\ \underbrace{M_2,\ M_2, \ldots, M_2}_{m_2}, \ldots, \underbrace{M_{v+1},\ M_{v+1}, \ldots, M_{v+1}}_{m_{v+1}}),$$

with $M_1 < M_2 < \cdots < M_v < M_{v+1}$. Consider the set $S = \{M_i |\ M_{i-1} \text{ divides } M_i\}$. Let $g = |S|$. For $1 \leq j \leq g$, let $N_1, N_2, \ldots, N_g$ denote the elements of S in decreasing order. We always have $N_g = M_2$, since $M_1 = 1$. For $1 \leq j \leq g$ let $M_{k_j} = N_j$.

Theorem 3 *The corresponding Puiseux characteristic is*

$$[\lambda_0; \lambda_1, \ldots, \lambda_g]$$

where

$$\lambda_0 = M_{v+1} \tag{5.1}$$

$$\lambda_j = \sum_{i \geq k_j} m_i M_i + M_{k_j} + M_{k_j - 1} \tag{5.2}$$

for $1 \leq j \leq g$.

Example 1 *Suppose* $der = (1, 1, 2, 2, 2, 2, 2, 2, 4, 6, 6, 6, 18, 24, 24)$. *The associated RVT code is* $RVVTRVVRRRRRV$. *Note that* $\lambda_0 = M_{v+1} = M_6 = 24$. *We also have* $S = \{18, 4, 2\}$, *and therefore* $g = 3$. *Then write* $S = \{18, 4, 2\} = \{N_1, N_2, N_3\} = \{M_5, M_3, M_2\}$ *so that* $k_1 = 5$, $k_2 = 3$, *and* $k_3 = 2$. *Finally, we compute*

$$\lambda_1 = \sum_{i \geq 5} m_i M_i + M_5 + M_4 = 90$$

$$\lambda_2 = \sum_{i \geq 3} m_i M_i + M_3 + M_2 = 94$$

$$\lambda_3 = \sum_{i \geq 2} m_i M_i + M_2 + M_1 = 103.$$

The Puiseux characteristic is thus

$$[24; 90, 94, 103].$$

Example 2 *This example is very similar to the previous, and the subtle differences should provide room for comparison. It appears in [22] as Example 3.28.*

Suppose $der = (1, 1, 2, 2, 2, 2, 2, 2, 4, 6, 6, 6, 6, 18, 24, 24)$. *Then the associated RVT code is* $RVVTRRVVRRRRRV$. *Note that we have the same values of* g, N_j, k_j, *and* $M_1, M_2, \ldots, M_{v+1}$ *as in the previous example. Thus, we compute*

$$\lambda_1 = \sum_{i \geq 5} m_i M_i + M_5 + M_4 = 90$$

$$\lambda_2 = \sum_{i \geq 3} m_i M_i + M_3 + M_2 = 100$$

$$\lambda_3 = \sum_{i \geq 2} m_i M_i + M_2 + M_1 = 109.$$

The Puiseux characteristic is thus

$$[24; 90, 100, 109].$$

Remark 4 *We are implicitly assuming that the underlying RVT class is critical (ends with V or T). Then the associated planar curve germ is necessarily non-immersed. We only discuss the Puiseux characteristic for these singular planar curve germs, since any immersed planar curve germ has normal form $(t, 0)$. This restriction agrees with the domain of the map Pc given in Section 4.2.4. In terms of the derived vector, according to Section 4.2.2, we must assume that $m_1 = M_2$.*

5.2 Proof of Theorem

The theorem is proved by induction on g. For readability, we break the proof into three subsections. In the first, we prove a useful lemma. In the second, we verify the base case $g = 1$. In the third, we complete the inductive step.

5.2.1 Lemma

The following lemma makes use of the basic bricks A_i constructed in Section 3.1 of [25]. These are integers from which the entries M_i in the derived vector are built, and the two actually coincide in some cases – for the precise relationship, see Theorems 3.3 and 3.4 in [25]. The A_i (and subsequently the M_i) depend only on the parameters $t_1, t_2, \ldots, t_v$; the multiplicity m_i depends on both t_i and r_i. The bricks are constructed as follows:

$$\begin{aligned} A_1 &= 1 \\ A_2 &= 2 + t_1 \\ A_i &= A_{i-2} + A_{i-1}(1 + t_{i-1}) \quad \text{for} \quad 3 \leq i \leq v + 1. \end{aligned}$$

Lemma 41 *Let $(\omega) = VT^{t_N} \cdots VT^{t_2}VT^{t_1}$ for $N \geq 2$. Let $(a, b) = \mathbb{E}_\omega(1, 2)$. Then*

$$\begin{aligned} a &= A_{N+1} \\ b &= A_1 + A_2 + \sum_{i=2}^{N} (1 + t_i) A_i. \end{aligned}$$

Proof The lemma follows from the following two observations:

$$A_1 + A_2 + \sum_{i=2}^{N}(1+t_i)A_i = A_N + A_{N+1} \tag{5.3}$$

and

$$\mathbb{E}_V\mathbb{E}_T^{t_N}\cdots\mathbb{E}_V\mathbb{E}_T^{t_2}\mathbb{E}_V\mathbb{E}_T^{t_1}(1,2) = (A_{N+1}, A_N + A_{N+1}). \tag{5.4}$$

Both observations can be easily verified for $N = 2$. Assuming Equation (5.3) holds for N, we find

$$\begin{aligned} A_1 + A_2 + \sum_{i=2}^{N+1}(1+t_i)A_i &= A_1 + A_2 + \sum_{i=2}^{N}(1+t_i)A_i + (1+t_{N+1})A_{N+1} \\ &= A_N + A_{N+1} + (1+t_{N+1})A_{N+1} \\ &= A_{N+1} + A_{N+2}, \end{aligned}$$

so (5.3) is proved by induction. Similarly, assuming Equation (5.4) holds for N, we find

$$\begin{aligned} \mathbb{E}_V\mathbb{E}_T^{t_{N+1}}\mathbb{E}_V\mathbb{E}_T^{t_N}\cdots\mathbb{E}_V\mathbb{E}_T^{t_2}\mathbb{E}_V\mathbb{E}_T^{t_1}(1,2) &= \mathbb{E}_V\mathbb{E}_T^{t_{N+1}}(A_{N+1}, A_N + A_{N+1}) \\ &= \mathbb{E}_V(A_{N+1}, t_{N+1}A_{N+1} + A_N + A_{N+1}) \\ &= \mathbb{E}_V(A_{N+1}, A_{N+2}) \\ &= (A_{N+2}, A_{N+1} + A_{N+2}), \end{aligned}$$

so (5.4) is proved by induction. This completes the proof of the lemma. $\square$

5.2.2 Base case of the induction

For the base case, assume $g = 1$. Note that $k_1 = k_g = 2$. Now, according to Section 4.2.2 we have $r_i = 0$ for $i = 2, \ldots, v$, so the associated RVT code is of the form $R^s\omega$, with (ω) entirely critical (containing no letters R). We also have $m_{v+1} = s + 1$ and $m_i = 1 + t_i$ for $i = 2, \ldots, v$. Then Lemma 41, with $N = v$, gives

$$\begin{aligned} a &= A_{v+1} \\ b &= A_1 + A_2 + \sum_{i=2}^{v}(1+t_i)A_i \\ &= A_1 + A_2 + \sum_{i=2}^{v} m_iA_i. \end{aligned}$$

Then, according to Section 4.2.4, we have that the Puiseux characteristic is $[\lambda_0; \lambda_1]$, with

$$\begin{aligned}
\lambda_0 &= a = A_{v+1} \\
\lambda_1 &= sa + a + b \\
&= (m_{v+1} - 1)A_{v+1} + A_{v+1} + A_1 + A_2 + \sum_{i=2}^{v} m_i A_i \\
&= A_1 + A_2 + \sum_{i=2}^{v+1} m_i A_i.
\end{aligned}$$

However, according to Theorem 3.3 from [25], in the case $g = 1$ we have that $A_i = M_i$ for all i. Thus, the Puiseux characteristic is $[\lambda_0; \lambda_1]$ with

$$\begin{aligned}
\lambda_0 &= M_{v+1} \\
\lambda_1 &= M_1 + M_2 + \sum_{i=2}^{v+1} m_i M_i
\end{aligned}$$

as desired.

5.2.3 Inductive step

We now assume Theorem 3 for derived vectors satisfying $|S| < g$, and prove that the result must hold for derived vectors with $|S| = g$. We may assume $g > 1$.

Begin with an arbitrary derived vector

$$der = (\underbrace{M_1,\ M_1, \dots, M_1}_{m_1},\ \underbrace{M_2,\ M_2, \dots, M_2}_{m_2}, \dots, \underbrace{M_{v+1},\ M_{v+1}, \dots, M_{v+1}}_{m_{v+1}})$$

for which $|S| = g$. The idea is to truncate the associated RVT code (α) after the last occurring letter R. Our inductive hypothesis will then apply to the derived vector associated to the truncated code, and we can reconstruct the Puiseux characteristic of the original derived vector from here.

To this end, we must give special attention to the entry $N_{g-1} = M_{k_{g-1}}$ in der. For notational purposes, we set $q = k_{g-1}$. Then by assumption we have that M_{q-1} divides M_q, and M_{q-1} is the smallest such entry (besides $M_1 = 1$).

The relations in Section 4.2.2 imply that we can write

$$(\alpha) = (\beta R^s \omega)$$

with

$$\begin{aligned}
s &> 0 \\
\omega &= \text{enitrely critical RVT string} \\
\beta &= \text{critical RVT code.}
\end{aligned}$$

More explicitly, we write

$$(\alpha) = \underbrace{R^{r_v+1}VT^{t_v}R^{r_v}\cdots VT^{t_{q-1}}}_{\beta} R^{r_{q-1}} \underbrace{VT^{t_{q-2}}VT^{t_{q-3}}\cdots VT^{t_1}}_{\omega}$$

and note that $s = r_{q-1} = m_{q-1} - \frac{M_q}{M_{q-1}} + 1$, by Equations (4.2) and (4.3).

Now, β is our truncated code and we will adorn all data concerning (β) with tildes to distinguish them from those of (α). In particular, we write $[\tilde{\lambda}_0; \tilde{\lambda}_1, \ldots, \tilde{\lambda}_{g-1}]$ for the Puiseux characteristic of (β) and

$$(\tilde{M}_1^{\tilde{m}_1}, \tilde{M}_2^{\tilde{m}_2}, \ldots, \tilde{M}_{\tilde{v}+1}^{\tilde{m}_{\tilde{v}+1}})$$

for the derived vector. Note that $\tilde{g}$ is indeed equal to $g-1$ by virtue of the recursive description of the mapping Pc in [22] (see Section 4.2.4).

With this setup, we begin the calculations. Applying Lemma 41 with $N = q-2$ we find that $(a, b) = \mathbb{E}_\omega$ is given by

$$\begin{aligned}
a &= A_{q-1} \\
b &= A_1 + A_2 + \sum_{i=2}^{q-2} (1 + t_i) A_i.
\end{aligned}$$

But according to Theorem 3.4 in [25], we have[1] $A_i = M_i$ for $1 \le i \le q-1$. Also, by Equations (4.3)-(4.5), we have $m_i = 1 + t_i + r_i$ for all $i \ge 2$. But $r_1 = r_2 = \cdots = r_{q-2} = 0$, so for $2 \le i \le q-2$ we have $1 + t_i = m_i$. Thus,

$$\begin{aligned}
a &= M_{q-1} \\
b &= M_1 + M_2 + \sum_{i=2}^{q-2} m_i M_i.
\end{aligned}$$

[1]In [25], the quantity n_1 tells us the position of the last occurring letter R in a critical RVT code. Here, we have $n_1 = q-2$, which means $r_{q-1} \neq 0$ and $r_i = 0$ for $i < q-1$; in other words, ω is the largest entirely critical string at the tail of (α).

Next, we need to relate the derived vector of (α) to that of (β). Proposition 1 from [26] gives the following relations:

$$\begin{aligned}
\tilde{m}_1 &= \frac{M_q}{M_{q-1}} \\
\tilde{m}_i &= m_{q-2+i} \quad \text{for} \quad 2 \leq i \leq \tilde{v}+1 \\
\tilde{M}_i &= \frac{M_{q-2+i}}{M_{q-1}} \quad \text{for} \quad 1 \leq i \leq \tilde{v}+1 \\
\tilde{v} &= v - q + 2.
\end{aligned}$$

Also, according to Section 4.2.4, we know how to compute the Puiseux characteristic of (α) from that of (β):

$$\begin{aligned}
\lambda_j &= a\tilde{\lambda}_j \quad \text{for} \quad 0 \leq j \leq g-1 \\
\lambda_g &= a(\tilde{\lambda}_{g-1} + s - 1) + b - a.
\end{aligned}$$

But by induction we may assume

$$\begin{aligned}
\tilde{\lambda}_0 &= \tilde{M}_{\tilde{v}+1} \\
\tilde{\lambda}_j &= \sum_{i \geq \tilde{k}_j} \tilde{m}_i \tilde{M}_i + \tilde{M}_{\tilde{k}_j} + \tilde{M}_{\tilde{k}_j - 1} \quad \text{for} \quad 1 \leq j \leq g-1.
\end{aligned}$$

Putting this all together, we now compute the desired Puiseux characteristic of $der = der(\alpha)$ in three stages. First we compute λ_0, then λ_j for $1 \leq j \leq g-1$, then finally λ_g.

First, we easily find

$$\begin{aligned}
\lambda_0 &= a\tilde{\lambda}_0 \\
&= M_{q-1}\frac{M_{q+\tilde{v}-1}}{M_{q-1}} \\
&= M_{v+1}.
\end{aligned}$$

Next, for $1 \le j \le g-1$ we have

$$\begin{aligned}
\lambda_j &= a\tilde{\lambda}_j \\
&= M_{q-1}\Big(\sum_{i \ge \tilde{k}_j} \tilde{m}_i \tilde{M}_i + \tilde{M}_{\tilde{k}_j} + \tilde{M}_{\tilde{k}_j-1}\Big) \\
&= \sum_{i \ge \tilde{k}_j} m_{q-2+i} M_{q-2+i} + M_{q+\tilde{k}_j-2} + M_{q+\tilde{k}_j-3} \\
&= \sum_{i \ge q+\tilde{k}_j-2} m_i M_i + M_{q+\tilde{k}_j-2} + M_{q+\tilde{k}_j-3} \\
&= \sum_{i \ge k_j} m_i M_i + M_{k_j} + M_{k_j-1}.
\end{aligned}$$

The last equality comes from the general fact that for $1 \le j \le g-1$ we have

$$q + \tilde{k}_j - 2 = k_j. \tag{5.5}$$

To see this, recall that k_j is defined so that $N_j = M_{k_j}$ is the jth smallest entry in *der* which is divisible by the preceding entry. Since this observation applies to the derived vectors of both (α) and (β), we find that

$$\begin{aligned}
\tilde{M}_{\tilde{k}_j-1} \text{ divides } \tilde{M}_{\tilde{k}_j} \Leftrightarrow \frac{M_{q+\tilde{k}_j-3}}{M_{q-1}} &\text{ divides } \frac{M_{q+\tilde{k}_j-2}}{M_{q-1}} \\
&\Rightarrow M_{q+\tilde{k}_j-3} \text{ divides } M_{q+\tilde{k}_j-2} \\
&\Rightarrow M_{q+\tilde{k}_j-2} = M_{k_j} \\
&\Rightarrow q + \tilde{k}_j - 2 = k_j.
\end{aligned}$$

Lastly, we compute λ_g. From above, we know

$$\begin{aligned}
a\tilde{\lambda}_{g-1} = \lambda_{g-1} &= \sum_{i \ge k_{g-1}} m_i M_i + M_{k_{g-1}} + M_{k_{g-1}-1} \\
&= \sum_{i \ge q} m_i M_i + M_q + M_{q-1}.
\end{aligned}$$

Whence

$$\begin{aligned}
\lambda_g &= a(\tilde{\lambda}_{g-1} + s - 1) + b - a \\
&= \Big(\sum_{i \geq q} m_i M_i + M_q + M_{q-1} + m_{q-1} M_{q-1} - M_q \Big) \\
&\quad + \Big(M_1 + M_2 + \sum_{i=2}^{q-2} m_i M_i \Big) - M_{q-1} \\
&= M_1 + M_2 + \sum_{i \geq 2} m_i M_i.
\end{aligned}$$

Since we always have $k_g = 2$, this is the desired result.

$\blacksquare$

Bibliography

[1] R. Abraham and J.E. Marsden, *Foundations of Mechanics*, Benjamin-Cummings, (1978).

[2] A. Albouy, Projective dynamics and classical gravitation, arXiv:math-ph/0501026v2, (2005).

[3] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, (1978).

[4] V. I. Arnold, V.V. Kozlov, and A.I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, 3rd Ed., Springer-Verlag, (2010).

[5] A. Bellaiche and J. J. Risler, editors, *Sub-Riemannian Geometry*, Birkhäuser, (1996).

[6] G. Bliss, The problem of Lagrange in the calculus of variations, *American J. Math.*, **52**, (1930), 673-744.

[7] O. Bolza, *Calculus of Variations*, 2nd Ed., Chelsea, (1960).

[8] R. W. Brockett, Control theory and singular Riemannian geometry, *New Directions in Appl. Math.*, P. J. Hilton and G. S. Young, Eds., Springer-Verlag, (1981), 11-27.

[9] E. Cartan, Sur l'équivalence absolue de certains systèmes d'équations différentielles et sur certaines familles de courbes, *Bull. Soc. Math. France*, **XLII**, (1914), 12-48.

[10] A. Chenciner and R. Montgomery, A remarkable periodic solution of the three body problem in the case of equal masses, *Annals of Math.*, **152**, (2000), 881-901.

[11] F. Diacu, E. Perez-Chavela, and M. Santoprete, The n-body problem in spaces of constant curvature, arXiv:0807.1747v6 [math.DS], (2008).

[12] G. Folland, A fundamental solution for a subelliptic operator, *Bulletin of the AMS*, **79**, (1973).

[13] I. M. Gelfand, and S. V. Fomin, *Calculus of Variations*, Dover, (1963).

[14] A. Giaro, A. Kumpera, C. Ruiz, Sur la lecture correcte d'un résultat d'Élie Cartan, *C. R. Acad. Sci. Paris*, **113**, (1978).

[15] W. B. Gordon, A minimizing property of Keplerian orbits, *American J. Math.*, **99**, 5, (1970), 961-971.

[16] P. Griffiths, *Exterior Differential Systems and the Calculus of Variations*, Birkhäuser, (1983).

[17] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces,* Birkhäuser, 3rd printing, (2007).

[18] F. Jean, The car with N trailers: characterisation of the singular configurations, *ESAIM: Control, Optimisation, and Calculus of Variations*, **1**, (1996), 241-266.

[19] N. I. Lobachevsky, The new foundations of geometry with full theory of parallels [in Russian], 1835-1838, In Collected Works, V. 2, GITTL, Moscow, (1949), p. 159.

[20] F. Luca and J.J. Risler, The maximum of the degree of nonholonomy for the car with N trailers, *Proceedings of the 4th IFAC Symposium on Robot Control*, Capri, (1994), 165-170.

[21] R. Montgomery, *A Tour of Subriemannian Geometries*, AMS, (2002).

[22] R. Montgomery and M. Zhitomirskii, *Points and Curves in the Monster Tower*, Memoirs of the AMS, **956**, (2010).

[23] R. Montgomery and M. Zhitomirskii, Geometric approach to Goursat flags, *Ann. Inst. H. Poincaré - AN*, **18**, (2001), 459-493.

[24] R. Montgomery and C. Shanbrom, Keplerian dynamics on the Heisenberg group and elsewhere, arXiv:1212.2713 [math.DS], (2012). Submitted to Geometry, Mechanics and Dynamics: the Legacy of Jerry Marsden, Fields Institute Communications series.

[25] P. Mormul, Geometric classes of Goursat flags and their encoding by small growth vectors, *Central European J. Math.*, **2**, (2004), 859-883.

[26] P. Mormul, Small growth vectors of the compactifications of the contact systems on $J^r(1,1)$, *Contemporary Mathematics*, **569**, (2012), 123-141.

[27] R. Palais, The principle of symmetric criticality, *Commun. Math. Phys*, **69**, (1979), 19-30.

[28] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Wiley, (1962).

[29] P.J. Serret, *Théorie nouvelle géométrique et mécanique des lignes a double courbure*, Librave de Mallet-Bachelier, Paris, (1860).

[30] C. Shanbrom, The Puiseux characteristic of a Goursat germ, arXiv:1103.2999v3 [math.DG], (2013). To appear in *J. Dynamical and Control Systems.*

[31] R. Strichartz, Sub-Riemannian geometry, *J. Differential Geometry*, **24**, (1986), 221-263.

[32] C. Wall, *Singular Points of Plane Curves*, London Mathematical Society Student Texts, **63**, (2004).

[33] L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*, 2nd Ed., Chelsea, (1980).