APPLICABLE NONLINEAR ANALYSIS
Volume 3 (2026), No. 1, 8–44
https://doi.org/10.69829/apna-026-0301-ta02
Tulipa Opera Scholarum

BETA FUNCTION FORMALISM FOR GRUSHIN OPERATORS (1)

DER-CHEN CHANG[1,2] AND JONATHAN RIESS[3,*]

[1] *Department of Mathematics and Statistics, Georgetown University, Washington D.C. 20057, USA*
[2] *Graduate Institute of Business Administration, College of Management, Fu Jen Catholic University, New Taipei City 242, Taiwan, ROC*
[3] *Department of Mathematics and Statistics and Department of Physics, Georgetown University, Washington DC, 20057, USA*

Abstract. This is the first part of a series of articles. Here we we develop the Beta function formalism and apply it to the Grushin operator $\Delta_G = \frac{1}{2}(\partial_x^2 + x^{2k}\partial_y^2)$ with $k \geq 1$. We want to construct the heat kernel for the operator Δ_G. In other words, we want to construct the fundamental solution for the operator $\frac{\partial}{\partial t} - \Delta_G$ for $t > 0$. We apply Hamiltonian mechanics to handle this problem and first need to solve the Hamiltonian system defined by the principal symbol $H(x, y, \xi, \eta) = \frac{1}{2}(\xi^2 + x^{2k}\eta^2)$ and given by Hamilton's equations:

$$\dot{x} = H_\xi = \xi, \;\; \dot{y} = H_\eta = x^{2k}\eta, \qquad \dot{\xi} = -H_x = -kx^{2k-1}\eta^2, \;\; \dot{\eta} = -H_y = 0$$

Hence, from the Hamilton's equations, we know that $\ddot{x} = -k\eta^2 x^{2k-1}$ and $\dot{y} = \eta x^{2k}$. Set $-k\eta^2 x^{2k-1} = -V'(x)$, therefore $\frac{1}{2}\dot{x}^2 + V(x) = E$. Here $V(x) = \frac{1}{2}\eta^2 x^{2k}$ is the potenial and E is the total energy. This implies that

$$s = \int_{x_0}^{x_1} \frac{dx}{\sqrt{2(E - \frac{1}{2}\eta^2 x^{2k})}} \quad \text{and} \quad y(s) = \eta \int_0^s x^{2k}(\tau)d\tau.$$

(1). When $k = 1$, the equation is a linear ODE and we can write down its solutions in the form of

$$x(s) = C_1 \sin(\eta s + C_2),$$

where C_1 and C_2 are two constants. So this case can be studied by a careful analysis of the trigonometric functions.
(2). The integrand in the above integral is not integrable for $k \neq 1$ We can express this Abelian integral as a (Jacobi's) Elliptic functions: but this *only really works for* $k = 2$:

$$\int R(t, \sqrt{P(t)})dt = \text{Elementary Functions} + c_1 I_0 + c_2 I_1 + c_3 J_1,$$

where $P(t)$ is degree 3 or 4.
(3). When $k \geq 3$, this allows us to find the parametrisations in terms of the incomplete Beta function of the (normal) geodesics solutions given by the Hamiltonian system for the associated Hamiltonian of the Grushin operator, *i.e.,* its principal symbol. The Beta functions arise as a result of expressing the standard energy integral $t = \int \frac{dx}{\sqrt{2E - \theta^2 x^{2k}}}$, which is elliptic for $k > 1$, in terms of Beta functions.

Given these parametrisations, we can then count the number of geodesics connecting a given starting and ending point, which is done by finding the intersections of the x- and y-parametrisations. The existence of such curves joining any given two points on the underlying sub-Riemannian manifold, which is defined by the principal symbol of the Grushin operator, we justify by the Chow-Rashevskii theorem.

In this article, we develop the motivation and mathematical background necessary to understand how and why we are interested in studying subelliptic and hypoelliptic operators such as Grushin and Kohn's operator. We also introduce the basics of the elliptic function theory and elliptic integrals, which we will use throughout the entire article.

Keywords. Geometric analysis, Grushin operator, Elliptic function theory, Hamilton-Jacobi mechanics, Heat kernel, Path integrals.

1. Introduction

This is the first part of a series articles. In this series, we develop the Beta function formalism and apply it to the Grushin operator $\Delta_G = \frac{1}{2}(\partial_x^2 + x^{2k}\partial_y^2) = \frac{1}{2}(X_1^2 + X_2)$ whee $X_1 = \frac{\partial}{\partial x}$ and $X_2 = x^k \frac{\partial}{\partial y}$ with $k \geq 1$. This allows us to find the parametrisations in terms of the incomplete Beta function of the (normal) geodesics solutions given by the Hamiltonian system for the associated Hamiltonian of the Grushin operator, *i.e.,* its principal symbol. The Beta functions arise as a result of expressing the standard energy integral $t = \int \frac{dx}{\sqrt{2E - \theta^2 x^{2k}}}$, which is elliptic for $k > 1$, in terms of Beta functions.

Given these parametrisations, we can then count the number of geodesics connecting a given starting and ending point, which is done by finding the intersections of the x- and y-parametrisations. The existence of such curves joining any given two points on the underlying sub-Riemannian manifold, which is defined by the principal symbol of the Grushin operator, we justify by the Chow-Rashevskii theorem, [9] which we review in chapter one.

In this article, we also develop the motivation and mathematical background necessary to understand how and why we are interested in studying subelliptic and hypoelliptic operators such as Grushin and Kohn's operator. We also introduce the basics of the elliptic function theory and elliptic integrals, which we will use later in the discussion.

In the second part of the series, we focus on studying the Grushin operator when $k = 1$. Here we show how the geodesics given in Beta function form reduce to previously obtained results by Chang and his collaborators, who expressed the parametrisations in terms of trigonometric functions. Then we generalize to the case of arbitrary $k \geq 1$.

Not only can we find the geodesics using the Beta function formalism, but we also can find the heat kernel of the Grushin operator, which is the fundamental solution for the associated heat equation $\partial_t u = \Delta_G u$. We do this by finding the propagator of associated $2k$-Hermite operator $\mathbb{L} = \frac{1}{2}(\partial_x^2 - x^{2k}\eta^2)$ (obtained by partial Fourier transforming the heat equation in y) using the path integral method and van Vleck's formula. This requires us to find the geodesics and classical action of the associated $2k$-Hermite operator, which we again can express in terms of incomplete Beta functions using the Hamilton-Jacobi formalism. Lastly, we comment on alternative methods of expressing the energy integral using other special functions such as Gauss's hypergeometric function or Jacobi's elliptic functions/elliptic integral functions. We determine that these tend to be more inconvenient than using the Beta function formalism.

In the final part of this series, we also introduce the necessary background in several complex variables relevant for understanding the Heisenberg group and the Kohn Laplacian. We then apply Hamiltonian mechanics to Kohn's operator $P = LL^* + (\bar{z}^k L)^*(\bar{z}^k L)$ where $L = \partial_z + i\bar{z}\partial_t$ and $L^* = \partial_{\bar{z}} - iz\partial_t =: -\bar{L}$ to find the Hamiltonian system in complex coordinates $z, \bar{z}, t$ and real coordinates x, y, t. In addition, we show how Kohn's operator may be viewed as a Grushin operator after evaluating the commutator between $\bar{L}$ and $|z|^{2k}$.

*Corresponding author.
E-mail address: chang@georgetown.edu (D.-C. Chang), jdr143@terpmail.umd.edu (J. Riess)
2020 Mathematics Subject Classification: 35H05, 35H20, 33C05, 58I99, 22E25.
Accepted: August 19, 2025.

2. Principal Symbols

Partial differential equations (PDEs) describe many important phenomena. For our purposes, we can define a kth-order partial differential equation as an equation involving an unknown function u of $n \geq 2$ variables, its partial derivatives ∂^i up to a combined k-order, and the coordinates x:

$$F[\partial^k u, \partial^{k-1} u, \ldots, \partial u, u, x] = 0$$

where $x = (x_1, \ldots, x_n)$ are the local coordinates for the domain U on which the unknown function u is defined and ∂ is the partial derivative operator.

If we restrict ourselves to linear PDEs, we can write a generic PDE in terms of a linear differential operator L:

$$L[u] = f$$

where a linear operator satisfies for two functions u, v and numbers β, γ:

$$L[\beta u + \gamma v] = \beta L[u] + \gamma L[v]$$

and can be written with the help of multi-index notation, introduced by L. Schwartz, as shown below. Therefore, to study and classify the behaviour and solutions of PDEs, we can simply study and classify differential operators.

A multi-index is an n-tuple $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}_0^n$ where each entry is a non-negative integer $\alpha_i \in \mathbb{N}_0$. For the multi-index, we define the following conventions in terms of powers:

$$\partial^\alpha := \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d}\right)^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$$

and absolute value:

$$|\alpha| := |\alpha_1| + \cdots + |\alpha_d|$$

Hence, we can write a general linear partial differential operator L of order m as:

$$L(x, \partial) = \sum_{|\alpha| \leq m} a_\alpha(x) \partial^\alpha$$

We can now associate to the operator $L(x, \partial)$ a function $P(x, i\xi)$ of x and ξ, by replacing $(\partial_1, \ldots, \partial_d)$ by $(i\xi_1, \ldots, i\xi_d)$:

$$P(x, i\xi) = \sum_{|\alpha| \leq m} a_\alpha(x)(i\xi)^\alpha$$

Note here that ξ is simply a formal variable at this point, it does not necessarily live in any important underlying space. The function $P(x, i\xi)$ is called the **complete symbol** (also called symbol or total symbol) of the differential operator $L(x, \partial)$ and is a polynomial in ξ of degree m whose coefficients depend on x. Due to the imaginary unit, we will often simply write differential operators directly using the operators $D_j := \frac{1}{i}\partial_j$ and $D := \frac{1}{i}(\partial_1, \ldots, \partial_d)$, giving us $L(x, D) = \sum a_\alpha(x) D^\alpha$ and symbols $P(x, \xi) = \sum a_\alpha(x)\xi^\alpha$. Therefore, much like we mapped the problem of studying PDEs to that of studying differential operators, we can again map the problem of studying differential operators to an algebraic problem of studying certain types of polynomials:

$$\text{PDEs} \longrightarrow \text{Operator Theory} \longrightarrow \text{“Algebra”}$$

It turns out that the complete symbol, *i.e.,* including all lower powers, is usually not necessary for characterising the behaviour and solutions of the associated PDEs. Instead, it suffices to look at only the highest order powers. This motivates the definition of the **principal symbol** or principal part.

Definition 2.1. The principal symbol of a linear partial differential operator $L(x, D) = \sum_{|\alpha|\leq m} a_\alpha(x)D^\alpha$ is the function:

$$P_m(x, \xi) := \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha$$

Hence, the principal symbol consists of only terms of order exactly m making P_m a homogeneous polynomial of degree m in ξ.

It turns out that unlike the complete symbol, the principal symbol is intrinsically defined, *i.e.,* $P_m(x, \xi)$ is a function on the cotangent bundle. By this we mean that while the variables ξ are simply formal variables in the case of the complete symbol for the principal symbol $\xi \in T^*_x M$, where $x \in M$. We summarize this result in the following theorem.

Theorem 2.2. *If $L = L(x, D)$ is a linear partial differential operator defined on some subset $\Omega \subset \mathbb{R}^d$, then its principal symbol P_m is a well-defined function on the cotangent bundle, $T^*\Omega$. The full symbol $P(x, \xi)$ is however not necessarily a function on the cotangent bundle $T^*\Omega$.*

For a proof and a counterexample for the full symbol case, see Theorem 2 (page 29) and Problem 4 (page 30) in Chapter 1 of [29].

The cotangent bundle can be understood from the following construction. We begin by visualising a tangent vector v at a point $x \in M$ as the tangent vector to a curve passing through x in M. The set of all such tangent vectors for all different curves at the point x is called the tangent space at x and is denoted $T_x M$. This tangent space forms a vector space and we can therefore consider the dual space of the tangent space which we call the cotangent vector space $T^*_x M$. A dual vector ξ in $T^*_x M$ can be visualised as the levels sets of ξ, *i.e.,* as hypersurfaces formed by $\xi(v) =$ const. in the tangent space. One can then think of $\xi(v)$ as counting the number of level surfaces pierced by the vector v. Since however we have multiple points in a manifold, we can consider the collection of all tangent spaces or all cotangent spaces. Thus, we call the space formed by all pairs $(x \in M, v \in T_x M)$ the tangent bundle TM and the space of all pairs $(x \in M, \xi \in T^*_x M)$ as the cotangent bundle T^*M.

There are generalised definitions of the principal symbol involving defining differential operators as maps on smooth sections of vector bundles, however, this definition is sufficient for our purposes.

Note: For the interested reader, the generalised definition goes as follows. Let $E \to M$ and $F \to M$ be smooth vector bundles over the underlying smooth base manifold M. We can think of a kth-order linear differential operator as a map between the space of smooth section:

$$L : \Gamma(E) \to \Gamma(F)$$

In local coordinates $\{x^i\}$ on $U \subset M$, we can write this operator as $L = \sum_{|\alpha|\leq k} \ell_\alpha \partial^\alpha = \sum_{|\alpha|\leq k} \ell_\alpha (\partial/\partial x)^\alpha$ and its action on local sections $u \in \Gamma(E|_U)$, which can be written as $u = \sum_i f_i x^i$ for some scalar values unctions f_i on U, as:

$$u(x) \mapsto L[u(x)] = \sum_{|\alpha|\leq k} \ell_\alpha(x)\partial^\alpha u = \sum_{|\alpha|\leq k} \sum_i \partial^\alpha(f_i)\ell_\alpha(x^i)$$

where for each multi-index $\alpha = (\alpha_1, \ldots, \alpha_m)$, $\ell_\alpha(x) : E \to F$ is a bundle homomorphism/bundle morphism/bundle map $E \to F$, and hence $\ell_\alpha(x^i)$ is a local section of F, implying $L[u(x)] \in \Gamma(F|_U)$. Since the coordinate system $\{x^i\}$ on M gives us a local trivilisation of T^*M by the coordinate differentials $\{dx^i\}$, which give us the fibre coordinates $\{\xi^i\}$, we can write the symbol (of L in the direction of the covector field ξ) with the help of the covector field/1-forms $\xi = \xi_i dx^i \in \Gamma(T^*M)$ as:

$$\sigma_L(\xi) = \sum_{|\alpha|\leq k} \xi^\alpha \ell_\alpha : E \to F$$

Thus, if we have a chosen direction ξ, we can think of $\sigma_L : E \to F$ as a bundle morphism. However, more generally, if we don't want to a priori pick a direction ξ, we can write the symbol as a map with its domain being the tensor product of vector bundles of the cotangent bundle of M for each ξ^i and E:

$$\sigma_L : T^*M \otimes E \to F$$

3. Elliptic, Hypoellitic and Subelliptic Operators

We are interested in classifying differential operators in order to classify different types of PDEs. A familiar classification from an undergraduate course on partial differential equations might go as follows for second-order linear PDEs of the form:

$$A\partial_x^2 u + B\partial_x\partial_t u + C\partial_t^2 u + D\partial_x u + E\partial_t u + Fu + H = 0$$

where the coefficients are functions of x and t only. We can classify these equation using their determinant $\Delta := B^2 - 4AC$ as:

Type	Condition	Prototypical Example	Geometry
Elliptic	$\Delta < 0$	Laplace Equation: $\partial_x^2 u + \partial_t^2 u = 0$	Riemannian
Hyperbolic	$\Delta > 0$	Wave Equation: $\partial_x^2 u - \partial_t^2 u = 0$	Lorentzian
Parabolic	$\Delta = 0$	Heat Equation: $\partial_x^2 u - \partial_t u = 0$	…

The naming convention stems from the naming used for the conic sections derived from the quadratic forms resulting from the principal symbols of each equation type. We can carry such a classification over to the more general operator theoretic approach by defining an elliptic PDE as one in which the operator is elliptic, and so worth.

Definition 3.1. A linear partial differential operator $L(x, D)$ is called elliptic at x if and only if:

$$P_m(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \neq 0 \tag{3.1}$$

for all $\xi \in \mathbb{R}^d \backslash 0$. It is elliptic on $\Omega \subset \mathbb{R}^d$ if it is elliptic at each point of Ω.

Often, it is said that condition (3.1) means that P_m is invertible. Hence, we say that an operator is elliptic if its principle symbol is invertible. It turns out that if we assume the coefficients $a_\alpha(x)$ of L to be continuous, then the set of points x where L is elliptic is open. We should note that we can also phrase ellipticity in terms of the characteristic variety of an operator.

Definition 3.2. Given an operator $L(x, D)$, the characteristic variety of L at a fixed point $\underline{x}$ is the set of $\xi \in \mathbb{R}^d \backslash 0$ such that:

$$P_m(\underline{x}, \xi) = 0 \tag{3.2}$$

The principle symbol $\hat{\sigma}_L$ picks out only the highest-order partial derivative terms in the symbol. Given a particular direction ξ, we again can think of $\hat{\sigma}_L : E \to F$ as a bundle morphism:

$$\hat{\sigma}_L(\xi) = \sum_{|\alpha|=k} \xi^\alpha \ell_\alpha : E \to F \quad \text{with} \quad e \in E \mapsto \hat{\sigma}_L(\xi)e = \sum_{|\alpha|=k} \xi^\alpha \ell_\alpha e \in F$$

Or, if we don't want to a priori pick a direction, we can use the fact that the bundle maps $\ell^\alpha(x) : E \to F$ are symmetric on the indices α and hence the kth order coefficients of L transform as symmetric tensors. We can represent the kth symmetric power of the cotangent bundle as $S^k(T^*M)$, this giving us the principle symbol are the symmetric tensor:

$$\hat{\sigma}_L : S^k(T^*M) \otimes E \to F$$

For a fixed point $\underline{x} \in M$, the principle symbol is thus a function on $T^*_{\underline{x}}M$ with values in $\mathrm{Hom}(E_{\underline{x}}, F_{\underline{x}})$.

Note: We should note that the 'undergraduate' classification of parabolic, elliptic, and hyperbolic is essentially impossible to generalise to a broader consideration of PDEs. While we give insight into how this is done for elliptic PDEs, this can only really be done also for hyperbolic ones as well. It turns out that parabolic PDEs not only do not physically make such sense, e.g. the heat equation assumes and an unphysical infinite propagation speed for diffusion purposes and thus needs to be made hyperbolic through the relativistic heat equation, but also don't have a clear mathematical characterisation either.

Note: More generally, if we consider $\xi \in \Gamma(T^*M)$ covector fields, then the operator $L : \Gamma(E) \to \Gamma(F)$ is elliptic at a point $x \in M$ if $\hat{\sigma}_L(\xi) : E_x \to F_x$ is a linear-space isomorphism for all nonzero covector fields $\xi \neq 0$. Note that the invertibility of the principal symbol is precisely what makes $\hat{\sigma}_L$ an isomorphism.

If L is defined on a open set Ω, then the characteristic variety of L is the set:

$$\text{char}(L) := \{(x,\xi) \in T^*\Omega \backslash \{\xi = 0\} : P_m(x,\xi) = 0\} \tag{3.3}$$

We can therefore define an operator as elliptic in Ω if and only if its characteristic variety is the empty set. As a result, characteristic varieties and characteristic directions (the null directions of the principal symbol) of operators play a larger role in hyperbolic operators than elliptic operators. Elliptic operators are thus nicer to study since, due to the fact that there are no real characteristic direction, the wave front set of the solution u, which is the set of non-smooth or singular points, is the same as that of $L[u] = f$. This allows us to prove *elliptic regularity theorems* which tell us that if f is of some regularity class, e.g. smooth, then u is of the same regularity class, e.g. smooth. And vice versa, if u is not smooth at a given point, then we expect f not to be either. However, sometimes ellipticity is not strong enough to achieve a certain regularity condition. As a result, ellipticity is sometimes referred to as *weak* ellipticity to contrast it with a stronger condition, often referred to as *strong* or *uniform* ellipticity:

Definition 3.3. A linear partial differential operator $L(x,D)$ is called strongly elliptic at a point $x \in \Omega$ if there exists a constant $c > 0$ such that:

$$|P_m(x,\xi)| = \left| \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \right| \geq c|\xi|^m$$

for all $\xi \in \mathbb{R}^d$. We say L is strongly elliptic on $\Omega \subset \mathbb{R}^d$ if it is strongly elliptic for all $x \in \Omega$.

Strong ellipticity being a stronger condition can be seen from the fact that (weak) ellipticity is equivalent to the following condition.

Theorem 3.4. *A linear partial differential operator $P(x,D)$ is elliptic at a fixed $\underline{x}$ if and only if there exist two constants $c_1 > 0$ and c_2 (not necessarily positive) such that:*

$$|P(\underline{x},\xi)| \geq c_1|\xi|^m - c_2|\xi|^{m-1}$$

If however L is an elliptic operator on a bounded subset $\Omega \subset \mathbb{R}^d$ with continuous coefficients $a_\alpha(x) \in C^0(\overline{\Omega})$, then L is also strongly elliptic. There is much more to be said about ellipticity, however, rather than strengthening ellipticity conditions, we will be interested in weakening or generalising them to the cases of so-called sub-elliptic and hypo-elliptic operators. The idea behind these generalisations goes as follows. We know from elliptic regularity theorems that under certain conditions, smoothness of f can guarantee smoothness of the solution u in $L[u] = f$. However, ellipticity is sufficient but not necessary to guarantee smoothness of u in many cases. For example, the heat equation $(\partial_x^2 - \partial_t)u = 0$, which is not elliptic, has smooth solutions. This motivates the definition of hypoellipticity.

Definition 3.5. If all solutions to a PDE are smooth, then the PDE is called hypoelliptic. Accordingly, we define a linear partial differential operator L to be hypoelliptic on an open subset $\Omega \subset \mathbb{R}^d$ if for every neighbourhood (open and bounded subset) $U \subset \Omega$ and for every distribution $u \in \mathcal{D}(U)$, it holds that:

$$L[u] \in C^\infty(U) \implies u \in C^\infty(U)$$

We thus can also say whether an operator is hypoelliptic at a point $\underline{x}$, namely L is hypoelliptic at $\underline{x}$ if there exists a neighbourhood $U \ni \underline{x}$ such that:

$$L[u] \in C^\infty(U) \implies u \in C^\infty(U)$$

Note: For proofs of these statements, see Corollary 8.3.2 in [19] which stats that if L is elliptic, then $WF(u) = WF(L[u])$. This is related to Hörmander's propagation of singularities theorem, Theorem 8.3.1 in [19]. Note that these results for elliptic regularity are in general not extendable to hyperbolic PDEs.

Thus, we can think of hypoellipticity as telling us that if the image of the operator L is C^∞, then the preimage is also C^∞ in the sense of distributions. Any distributional solution must be smooth if the data are. And to localise this idea to a point, we can think of an operator being hypoelliptic at a point $\underline{x}$ if $L[u]$ is smooth near $\underline{x}$ implies u is smooth near $\underline{x}$, *i.e.,* in a neighborhood $U \ni \underline{x}$. According to this definition we have some very clear examples and counterexamples:

Example 3.6. We list the following examples.

(1) The derivative operator $L = d/dx$ on $\mathbb{R}$ is hypoelliptic since if the derivative is smooth, then the primitive function is also smooth by the fundamental theorem of calculus: $u(x) = \int_0^x L[u(t)]dt + c$.
(2) The partial derivative operator $L = \partial_x$ is not hypoelliptic at any point in $\mathbb{R}^2$ with coorindates (x, y) since $L[u] = \partial_x u(y) = 0$ but $u(y)$ need not be smooth.
(3) Every elliptic operator $L = \sum_{|\alpha|\leq m} a_\alpha(x)\partial^\alpha$ with smooth coefficients $a_\alpha(x) \in C^\infty$ is hypoelliptic, e.g. the Laplacian. In fact, the Laplacian is **analytic hypoelliptic**, meaning if f is analytic, so is u.
(4) The heat operator $L = \partial_t - k\Delta_x$ with $k > 0$ is hypoellitpic but not elliptic.
(5) The wave or d'Alembertian operator $L = \Box = \partial_t^2 - c^2\Delta_x$ is neither hypoelliptic nor elliptic.
(6) On $\mathbb{R}^2$, the Grushin operator $L = \partial_x^2 + x^2\partial_y^2$ is (analytic and C^∞) hypoelliptic but not always elliptic. Namely, L is elliptic everywhere except on the y-axis where $x = 0$. We call this **degenerate elliptic**.

This should give us a good intuition of what hypoellipticity is. We should note that there are many other characterisations or equivalent definitions of hypoellipticity which make the above definition more precise. Furthermore, as we will see, instead of thinking of hypoellipticity as something only relating to smoothness, *i.e.,* of class C^∞, we can loosen or strengthen the requirement on hypoellipticity to other regularity classes, e.g. involving Sobolev spaces.

Let us now turn to a specific notion of hypoellipticity, called subellipticity. To understand the motivation for the concept of subellipticity, we will need to review the basics of Sobolev spaces, in particular the Sobolev space $H^2(\Omega)$, and some standard elliptic regularity theorems.

Definition 3.8. If $\Omega \subset \mathbb{R}^d$ is open and $s \in \mathbb{N}$, then:

$$H^s(\Omega) := \{u \in L^2(\Omega) : \forall |\alpha| \leq s, \partial^\alpha u \in L^2(\Omega)\} \tag{3.4}$$

$H^s(\Omega)$ is thus the space of square-integrable functions whose distributional derivatives $\partial^\alpha u$ are also square integrable up to order s. If $s = 0$, then the space $H^0(\Omega) = L^2(\Omega)$. The function space $H^s(\Omega)$

Note: For example, Hörmander defines hypoellipticity as any operator P which satisfies Theorem 4.4.1 on page 110 in [19]. He then gives a list of equivalent definitions in Theorem 11.1.1 on page 61 in [20], defining hypoellipticity accordingly in Definition 11.1.2 in [20]. He then gives further necessary and sufficient conditions for hypoellipticity in Theorem 11.1.3. One of these conditions, specifically IIb, is often also used to define hypoellipticity as:

Definition 3.7. A linear operator $P(D) = \sum_{|\alpha|\leq m} a_\alpha D^\alpha$ with constant coefficients $a_\alpha \in \mathbb{C}$ is called hypoelliptic if there exists $C > 0$ such that for all $\xi \in \mathbb{R}^n$ with $|\xi| \geq C$ and all $\alpha \in \mathbb{N}^n$ such that:

$$P(\xi) \neq 0 \quad \text{and} \quad |D^\alpha P(\xi)| \leq C|P(\xi)|\xi|^{-|\alpha|}$$

Note: We should note that there are other equivalent definitions of the space H^s. For example, Hörmander defines in Definition 7.9.1 in [19] the space $H^s(\mathbb{R}^n)$ as the space of all tempered distributions $u \in \mathscr{S}'$ (for a definition of $\mathscr{S}'$ see Definition 2.4.1 on page 74 in [29]) such that their Fourier transform $\hat{u}$ is in L^2_s, *i.e.,* $\hat{u} \in L^2_s$, where L^2_s is the L^2 space of square integrable functions with respect to the measure $(1 + |\xi|^2)^s d\xi/(2\pi)^n$. Accordingly, the norm for H^s he gives is $||u||_{H^s} = [(2\pi)^{-n} \int |\hat{u}(\xi)|^2(1 + |\xi|^2)^s d\xi]^{1/2}$. When s is a positive integer, this is equivalent to our definition which follows Definition 1 on page 225 in section 5.9 of [29]. The only reason for using the other characterisation of H^s is to make the fact that the Fourier transform is an isomorphism $H^s \to L^2_s$ clear, since $L^2_s \subset \mathscr{S}'$.

forms a Hilbert space with the norm:

$$||u||^2_{H^2(\Omega)} := \sum_{|\alpha|\leq s} ||\partial^\alpha u||^2_{L^2(\Omega)}$$

For the purposes of localising results, we can define that a distribution $u \in \mathcal{D}(\Omega \subset \mathbb{R}^d)$ belongs to $H^2(\Omega)$ at the point $x \in \overline{\Omega}$, denoted $u \in H^s(x)$, if $\exists r > 0$ such that $u \in H^2(\Omega \cap B_r(x))$, *i.e.,* there exists an open ball around x such that in that neighbourhood, u belong to $H^2(\Omega)$. This is in general how we will define localised results. For example, we will also write $C^\infty(\underline{x})$ to mean C^∞ in a ball centered at $\underline{x}$, *i.e.,* $\exists r > 0 : C^\infty(\mathbb{R} \cap B_r(\underline{x}))$.

Notice that we are only considering functions of $L^2(\Omega)$. We could however also consider the space of functions $L^p(\Omega)$ for $p \in [1, \infty]$. Thus, if we replace $L^2 \to L^p$ in (3.4), we denote the Sobolev space as $W^{s,p}$.

Using this Sobolev space $H^s(\Omega)$, we can state two standard elliptic regularity theorems which are particularly important for motivating subellipticity.

Theorem 3.9. *Let $L(D)$ be a constant coefficient elliptic operator of order m and $s \in \mathbb{R}$, then:*

$$\boxed{\text{If } u \in H^s(\mathbb{R}^d) \text{ and } L[u] \in H^s(\mathbb{R}^d), \text{ then } u \in H^{s+m}(\mathbb{R}^d)} \tag{3.5}$$

Vice versa, if $L(D)$ is an mth order operator such that the a (3.5) holds for some $s \in \mathbb{R}$, then L must be elliptic. This is a global result, we can localise it as follows. An mth order $P(x, D)$ operator with variable smooth coefficients is elliptic if and only if $u \in H^{s+m}(x)$ whenever $u, L[u] \in H^s(x)$. Generalising even more, for $p \in [1, \infty]$ we have $L[u] \in W^{s,p} \implies u \in W^{s+m,p}$, where the case $p = \infty$ holds by Sobolev embedding theorem.

For a proof, see Proposition 1 (5.9.1) on page 226 in section 5.9 in [29]. The other important elliptic regularity theorem is called the **Interior Elliptic Regularity theorem**.

Theorem 3.10. *(Interior Elliptic Regularity theorem). Consider an elliptic operator L with smooth coefficients $a_\alpha(\underline{x}) \in C^\infty(\underline{x} : \mathbb{R})$, i.e., L is uniformly/strongly elliptic. Consider further a distributional solution $u \in \mathcal{D}(\Omega)$ belonging also to $u \in H^1(\underline{x})$. Then, if for some $0 \leq s \in \mathbb{N}$, $L[u] \in H^s(\underline{x})$, then $u \in H^{s+2}(\underline{x})$, i.e.,:*

$$\boxed{L[u] \in H^s(\underline{x}) \implies u \in H^{s+2}(\underline{x})}$$

For a proof, see Theorem 2 on page 227 in section 5.9 of [29]. Note that there is a related theorem called the boundary elliptic regularity theorem. From the interior result, it clearly follows by the Sobolev embedding theorem that $L[u] \in C^\infty(\underline{x}) \implies u \in C^\infty(\underline{x})$ as stated multiple times above. Note however that it does not follow $L[u] \in C^k(\underline{x}) \implies u \in C^{k+2}(\underline{x})$. This is often known as **Weyl's Lemma**, although this name is sometimes only exclusively used to refer to this property in the case of the Laplace operator/equation.

How should we interpret these theorems and why are they important? The standard interpretation is that not only are we guaranteed by the corollaries of these theorems that we have smooth C^∞ solutions for elliptic PDEs, but also, in a sense, we gain derivatives. If the RHS of the equation $L[u] = f$ has s derivatives, *i.e.,* is in H^s meaning $\partial^\alpha f \in L^2$, for all $\alpha| \leq s$, then the solution u has $s + m$ or $s + 2$ derivatives, *i.e.,* is in H^{s+m} or H^{s+2} respectively, according to the two theorems above. Hence, we say that for elliptic operators/PDEs, the solution gains m or gains 2 derivatives. Said more loosely, u is smoother than $L[u]$ by m or by 2 derivatives.

In a sense, this is the strongest possible hypoellipticity result one could hope for. Namely, in the spirit of these theorems, in the context of Sobolev spaces we can interpret L being hypoelliptic as meaning $L[u] \in H^s \implies u \in H^{r(s)}$, where u is thought of as gaining (or loosing) derivatives according to the relation $r(s)$, (see page 944 in [22]). Thus, in theorem (3.9), we have the strongest possible

hypoellipticity result since we assume the operator is of order m and we gain all m of the derivatives for our solution u. We gain the maximal possible smoothness in a sense.

But let's consider when this is not the case. Let's weaken this ellipticity result to the scenario in which we don't gain all m derivatives but instead only gain $m-\epsilon$, where ϵ is some small number. We say that we *loose* ϵ derivatives. Or if we consider the interior elliptic regularity theorem, we would say that we don't gain 2 derivatives or improve regularity by 2, we gain a little less than 2 derivatives or improve regularity a little less than 2. This is what we call **subellipticity** and is a specific case of hypoellipticity.

Definition 3.11. We call an operator L subelliptic if $L[u] \in H^s \implies u \in H^{s+m-\epsilon}$ for some fixed $\epsilon \in (0,1)$. Furthermore, we say L looses ϵ derivatives. The constraint that $\epsilon < 1$ is there to ensure that subellipticity is only a condition on the principal symbol, *i.e.,* independent of the lower terms in L. (see [18]).

Therefore, we can think of elliptic operators as those which loose no derivatives. We should note that there are other characterisations of, conditions for, and definitions of subellipticity. For the interested reader, see Hörmander's chapter in [19] and [20]. These mostly all generalize the notion of elliptic estimates to subelliptic estimates. However, the notion of subellipticity above suffices for us. We will only mention one other major characterizations of subelliptic operators according to so-called *a priori* subelliptic estimates.

Theorem 3.12. *(A priori subelliptic estimate). An operator L is subelliptic if for any relatively compact open subset $V \subset\subset U$, there exists an $\epsilon > 0$ and a constant $C < \infty$ such that for all $u \in C_0^2(V)$, it holds:*

$$||u||_{H^\epsilon} \leq C(||Lu||_{H^0} + ||u||_{H^0}) \tag{3.6}$$

This is in fact often used as a definition for subellipticity, as is done in the case of Kohn's paper [22] and Christ's paper [11]. Using analogous hypoelliptic estimates then, this subelliptic estimate is used to state that subellipticity implies hypoellipticity with a gain of 2ϵ derivatives in Sobolev norms.

Let us now turn to an important class of examples which is closely related to sub-Riemannian geometry and the Kohn's operator which we will study below.

Example 3.13. Consider r number of *real* smooth vector fields $\{X_1, \dots, X_r\}$ on $\mathbb{R}^d$. Let X_i^* be the formal L^2 adjoint of X_i. Define the operator (often called the sum of squares):

$$L = X_1^* X_1 + \cdots + X_r^* X_r = \sum_{i=1}^{r} X_i^* X_i \tag{3.7}$$

Notice that when $X_i = \partial_{x_i}$, then L is simply the Laplacian. We can now study two cases:

(1) If $\{X_i\}$ at a point $\underline{x} \in \mathbb{R}^d$ are tangent to a submanifold of $\mathbb{R}^d$, then L is not hypoelliptic. For example, let $d = 2, r = 1$ and $X_1 = \partial_x$, then L is not hypoelliptic as seen in the examples above.

Note: Standard elliptic estimates give results such as the Remark on page 228 in section 5.9 in [29]: If L is elliptic with smooth coefficients, then there exists a constant $c = c(s, \Omega_1 \subset \mathbb{R}^n, \Omega_2 \subset \mathbb{R}^n, L)$ such that for all u it holds:

$$||u||_{H^{s+2}(\Omega_1)} \leq c\Big[||L[u]||_{H^s(\Omega_2)} + ||u||_{H^1(\Omega_2)}\Big]$$

Such estimates can also be extended to the case of elliptic operators of order r over compact manifolds M. See Porpositions 6 (and 5) in [17]. This then motivates four different definitions of subellipticity based on what type of weakening in the estimate we assume. One of these will will mention below in Theorem (3.12). The other three can be found in Definitions 3.7-3.10 on page 11 in [32]

In the case of the sum of squares operator (3.7), this condition can equivalently be reformulated as:

$$||u||_{H^\epsilon}^2 \leq C(|Q(u,u)| + ||u||_{H^0}^2)$$

where $Q(u,u) = \sum_j ||X_j u||_{H^0}^2$. For a proof and discussion, see [23]

(2) If $\{X_i\}$ satisfy **Hörmander's condition** or the **bracket generating condition** at a point $\underline{x}$, that is the Lie algebra generated by $\{X_i\}$ (which includes Lie-brackets of $\{X_i\}$ up to some rank $\mathfrak{r}$) spans $T_{\underline{x}}\mathbb{R}^d$, then L is subelliptic (and therefore hypoelliptic) at $\underline{x}$ and L is called **Hörmander sub-Laplacian**. As a example, consider again Gurshin's operator on $\mathbb{R}^2$ with $X_1 = \partial_x$ and $X_2 = x\partial_y$. We can show that $L = X_1^*X_1 + X_2^*X_2 = -\partial_x^2 - x^2\partial_y^2$ satisfies the bracket generating condition at every $(\underline{x}, \underline{y}) \in \mathbb{R}^2$ since $[X_1, X_2] = [\partial_x, x\partial_y] = \partial_x(x\partial_y) - x\partial_y\partial_x = \partial_y + x\partial_x\partial_y - x\partial_x\partial_y = \partial_y$. Hence, L is subellipic at every $(\underline{x}, \underline{y}) \in \mathbb{R}^2$ (even though it is not elliptic on the y-axis).

The final examples we will consider are operators of the form (3.7) but with *complex* vector fields rather than *real* vector fields. This will give us an examples of operators which are subelliptic, hypoellitpic but not subelliptic, and not hypoelliptic at all. A key result proven by Kohn as Theorem A in [22] is that if we have complex-valued vector fields $\{X_i\}$ in $\mathbb{R}^n$ and they satisfy the bracket generating condition of order one, *i.e.,* only $[X_i, X_j]$ are needed to span the tangent space, meaning $\{X_i, [X_i, X_j]\}$ span the entire tangent space together, then $L = \sum_i X_i^*X_i$ is still subellipticity with $\epsilon = 1/2$. However, if we need higher order brackets such as $[X_i, [X_j, X_k]]$ to span the tangent space, then $L = \sum_i X_i^*X_i$ is no longer subelliptic.

Example 3.14. (Kohn's Operator [22].) Consider on $\mathbb{R}^2$ Kohn's operator:

$$P = LL^* + (\bar{z}^k L)^*(\bar{z}^k L)$$

where $L = \partial_z + i\bar{z}\partial_t$ and $L^* = \partial_{\bar{z}} - iz\partial_t =: -\bar{L}$. Kohn's operator can therefore also be written as:

$$P = -L\bar{L} - \bar{L}|z|^{2k}L = \bar{L}^*\bar{L} - (\bar{L}\bar{z}^k)z^kL = \bar{L}^*\bar{L} + (z^kL)^*z^kL$$

As a result, we can again see that this is a sum of squares operator. The complex Lie algebra generated by $\bar{L}$ and x^kL spans the complexified tangent space at every point, *i.e.,* we essentially satisfy Hörmander's condition in the complex vector field case. However, P turns out to be hypoelliptic but not subelliptic for $k \geq 1$ as shown by Kohn in [22]. At the origin $(0, 0)$, there is a large loss of derivatives, namely:

$$P[u] \in H^s \implies u \in H^{s-k+1}$$

Example 3.15. (Christ's operator [11]). Consider $\mathbb{R}^3$ with coordinates (x, t, s) and the operators:

$$\bar{L} = \partial_x - ix\partial_t, \quad L = \partial_x + ix\partial_t$$

Now choose a positive integer $k \geq 1$ and let the complex three vector fields $\{X_1, X_2, X_3\}$ be:

$$X_1 = \bar{L}, \quad X_2 = x^kL, \quad X_3 = \partial_s$$

We can show $\{X_1, X_2, X_3\}$ satisfy the bracket generating condition at each point in $\mathbb{R}^3$. However, the operator:

$$\mathcal{L} = \sum_{j=1}^{3} X_j^*X_j$$

is not hypoelliptic in any neighbourhood of the origin. Note that if we consider just $\{X_1, X_2\}$ as vector fields in $\mathbb{R}^2$, then the operator $P = X_1^*X_1 + X_2^*X_2$ is a version o Kohn's operator and is hypoelliptic.

Note: We should note that a key step in the proof of this result is using the fact that the bracket generating condition implies that there exists a neighbourhood U of the point for which there exists an $\epsilon > 0$ and constant $C = C(\epsilon, U)$ such that we can have the estimate:

$$||u||_\epsilon^2 \leq C\left(\sum_j ||X_j u||_0^2 + ||u||_2^2\right)$$

for all $u \in C_0^\infty(U)$. This gives us a so-called **subelliptic estimate** for the operator. Note that the Sobolev norm $||u||_\epsilon$ is simply the norm given above for $||\cdot||_{H^\epsilon}$, with $||\cdot||_0$ being simply the L^2 norm. The ϵ is related to the rank $\mathfrak{r}$ of the brackets needed to generate the whole tangent space through $\epsilon = 1/\mathfrak{r}$. For more, see page 944 in [22].

Thus, adding the term $-\partial_s^2$ to P destroys hypoellipticity for $\mathcal{L}$ due to singularities propagating along curves such as $\{(0,0,s)\}$.

Let us now summarise and give an overview of the different notions of ellipticity we have introduced. We have the following hierarchy of ellipticity:

$$\text{Uniformly/Strongly Elliptic} \implies \text{Elliptic} \implies \text{Subelliptic} \implies \text{Hypoelliptic}$$

The inverse directions do not hold. As mentioned, a famous counterexample for hypoellipticity not implying subellipticity is Kohn's operator. We have also tried to describe some important conditions which are sufficient for hypoellipticity and subellipticity. It turns out that the bracket generating condition or Hörmander's condition for real vector fields can guarantee for us hypo- and subellipticity for sum of squares operators (3.7). However, if we consider complex vector fields, the bracket generating condition can only guarantee subellipticity as long as only rank one commutators $[X_i, X_j]$ are needed to span the tangent space. Otherwise, we can construct examples of operators which are hypoelliptic but not subelliptic (Kohn's operator) or not hypoelliptic at all (Christ's operator). We should also note that satisfying the bracekt condition is only a sufficient condition and not a necessary condition to capture hypoellipticity even in the real vector field case. Namely, Christ also showed that the sum of squares operator is sometimes hypoelliptic without satisfying the bracket generating condition in [10].

Lastly, let us mention one important aspect of elliptic and subelliptic operators which lays at the foundation at our analysis of them throughout this thesis. Namely, it turns out that an elliptic operator can be associated in a natural way to a geometric Riemannian structure, much like how the standard Laplacian can be associated with Euclidean structure. Furthermore, a subelliptic operator can be associated to a generalisation of Riemannian geometry, namely a sub-Riemannian structure. Concretely we mean with this that the principal symbol of the operator gives us either a Riemmanian metric or a sub-Riemannian metric. This is analogous in the case for (normally) hyperbolic operators whose principal symbols give us semi-Riemannian or Lorentzian metrics.

Before we turn to the study of differential geometry however, let us introduce one last important concept in PDE theory, that of the fundamental solution or heat kernel.

4. Heat Kernels and Partial Fourier Transforms

A central object for solving partial differential equations is called the fundamental solution. They allow us to find any solution for some boundary or initial data given the solution for the delta function case. We will focus on a specific type of fundamental solution called heat kernels, defined as follows.

Definition 4.1. Let $\mathbb{L}$ be a partial differential operator in variables (x, y) on some manifold. We define the **heat kernel** as the function $K(x_0, y_0, x, y; t) = \mathcal{P}(x, y, t)$ which satisfies:

$$\begin{cases} \partial_t \mathcal{P} = \mathbb{L}\,\mathcal{P}, \quad t > 0, \\ \lim_{t\to 0} \mathcal{P}(x, y; t) = \delta_{x_0}(x) \otimes \delta_{y_0}(y) \end{cases} \tag{4.1}$$

where δ_a is the Dirac delta distribution centered at a.

Often, solving for the heat kernel directly can be difficult. Therefore, it is often convenient to apply the partial Fourier transform to (4.1) to reduce the number of variables in the operator one has to deal with. Namely, if we assume $\mathbb{L}$ is a partial differential operator with variables $(x, y) \in \mathbb{R}^n \times \mathbb{R}$, then we can define the partial Fourier transform in y as follows.

Note: Interested readers may also find [17], [28], [30], [31], [33] helpful.

Definition 4.2. The partial Fourier transform with respect to y of a function $\mathcal{P}(x, y; t)$ with $y \in \mathbb{R}$ is given by:

$$(\mathcal{F}_y \mathcal{P})(x, \xi; t) := \int_{\mathbb{R}} e^{-iy\xi}\, \mathcal{P}(x, y; t)\, dy \tag{4.2}$$

As a result, if we set $u = \mathcal{F}_y(\mathcal{P})$, then (4.1) becomes and u satisfies:

$$\begin{cases} \partial_t u = \mathbb{P}_x\, u, \quad t > 0, \\ u(x, \xi; 0) = \delta_{x_0}(x) \otimes I_\eta \end{cases} \tag{4.3}$$

where $\mathbb{P}_x$ is obtained the usual way in which Fourier transforms relate to their Fourier polynomials, by replacing essentially $\mathcal{F}(\partial_y^n) \to (i\xi)^n$.

We can then usually solve for the heat kernel $K(x_0, x; t)$ of $\mathbb{P}_x$ more easily by using the Hamilton-Jacobi and path integral formalism reviewed in a later section below. This comes in handy since the solution to (4.3) can usually be solved by the exponential Ansatz:

$$u(x, \xi; t) = e^{t\, \mathbb{P}_x}(\delta_{x_0}(x) \otimes I_\eta) \tag{4.4}$$

Applying the inverse partial Fourier transform with respect to ξ to $u(x, \xi; t)$ recovers $\mathcal{P}$ with the help of the heat kernel $K(x_0, x; t)$ via:

$$\mathcal{P}(x, y, ; t) = (\mathcal{F}_\xi^{-1} u)(x, y; t) = \frac{1}{2\pi} \int e^{iy\xi}\, u(x, \xi; t)\, d\xi = \frac{1}{2\pi} \int e^{iy\xi}\, K(x_0, x; t)\, d\xi$$

5. Differential Geometry

5.1. **(Semi-)Riemannian geometry.** Given a smooth manifold M, we can define a (semi-)Riemannian metric on M which allows us to study not just the topological properties of manifolds but also geometric ones. Geometric properties we can simply think of as lengths of curves or angles between two tangent vectors. The way we can do this is by equipping each tangent space of the manifold with an inner product which allows us to find angles and lengths of tangent vectors. We can do this abstractly using the idea of tensor fields. Recall that the set of all smooth vector fields over M forms a module over the ring $C^\infty(M)$, where $C^\infty(M)$ is the set of all smooth real-valued functions on M.

Definition 5.1. A tensor of type (r, s) over a V, which is a module over a ring K is a K-multilinear map (linear in each slot):

$$A : (V^*)^r \times V^s \to K$$

where $r \geq 0$ and $s \geq 0$ are integers, both not zero. The set of all tensors of type (r, s) over V is a module over K denoted $\mathfrak{T}_s^r(V)$.

Definition 5.2. A tensor field A on a smooth manifold M is a tensor over the C^∞-module $\mathfrak{X}(M)$, *i.e.,* A is a $C^\infty(M)$-multilinear map of type (r, s):

$$A : \mathfrak{X}^*(M)^r \times \mathfrak{X}(M) \to C^\infty(M)$$

Using these defintion, we can now define a semi-Riemannian metric as follows.

Definition 5.3. A semi-Riemannian metric tensor g on a smooth manifold M is a symmetric, nondegenerate $(0, 2)$ tensor field on M of constant index.

A $(0, 2)$ tensor field defines a bilinear form which we call symmetric i $g(V, W) = g(W, V)$ for all vector fields $V, W \in \mathfrak{X}(M)$. In addition, we call this bilinear form nondegenerate if $g(V, W) = 0$ for all $W \in \mathfrak{X}(M)$ implies $V = 0$. Using this metric tensor, we can then define a semi-Riemannian manifold as follows.

Definition 5.4. A semi-Riemannian manifold is a smooth manifold M equipped with a metric tensor g, *i.e.,* it is a duplet (M, g).

The common index value ν of each g_p is called the index of the manifold M. It must take values between $0 \leq \nu \leq \dim(M)$. If $\nu = 0$, then (M, g) is called a **Riemannian manifold**. The resulting $g_p : T_pM \times T_PM \to \mathbb{R}$ are then positive definite inner products on T_pM. If $\nu = 1$ and $n \geq 2$, then (M, g) is called a **Lorentzian manifold**. If M is a connected manifold, then the condition that g must be of constant index turns out to be superfluous.

This lays the foundation for (semi-)Riemannian geometry. However, in the next section we will see how to generalise this notion of geometry to so-called sub-Riemmanian geometry.

5.2. **Sub-Riemannian geometry.** Sub-Riemannian geometry is a natural generalisation of Riemannian geometry to the case where you are constrained to measure distances only along distinguished directions. Before we introduce the definition of what a sub-Riemannian manifold is however, we need to define the concept of a **distribution** D on a manifold M which will give us a way to prescribe a set of distinguished directions on a manifold.

Definition 5.5. A distribution D of rank k on a manifold M assigns to each point $p \in M$ a k-dimensional subspace $D_p \in T_pM$.

Hence, we can think of distributions as maps $D : p \in M \mapsto D_p \in T_pM$ that designate to each point p a restricted subset D_p of tangent vectors from the set of all tangent vectors T_pM at that point. This *restricted* or *distinguished* subset of tangent vectors D_p is called the set of **horizontal vectors** at p, with $v \in D_p$ being the horizontal vectors. As a result, we usually call the distribution D a **horizontal distribution**. The point of this distribution therefore is to distinguish the directions on the manifold, *i.e.,* tangent vectors at a given point, into either a horizontal direction, which are completely normal and familiar from the Riemannian case, or a *missing/forbidden* direction. These missing directions are then directions on the manifold which we cannot measure in the sense that the sub-Riemannian metric will be restricted only to D_p, not all of T_pM, allowing us to find lengths and angles only for horizontal vectors. Such restrictions might arise due to the system we are studying being constrained or from a fundamental consideration of underlying theory governing the dynamics of the system.

In the spirit of studying manifolds with differential structure, we call a distribution D differentiable if every point $p \in M$ has a neighbourhood V which has k differentiable vector fields $X_1, \ldots, X_k$ on it which forms a basis of D_q for all $q \in V$. This is important since then we can write any **differentiable distribution** D as the span:

$$D = \text{span}\{X_1, \ldots, X_k\}$$

locally on V.

The last property of a distribution we will need is so-called **involutive**.

Definition 5.6. A distribution D is called involutive if $[X, Y] \in D$ for any $X, Y \in D$.

Thus, we find that if any of the vector fields $X_i \in D$ do not commute, then, we call the distribution **noninvolutive**. A theorem by Frobenius shows that the property of being noninvolutive is equivalent to being nonintegrable, which has a very technical definition that we will ignore. Know however that in the definition of a sub-Riemannian manifold that follows, nonintegrable can simply be interpreted as having vector fields in the distribution that do not commute.

Note: For those familiar with the langauge of bundles, we can think of distributions as subbundles of the tangent bundle, $D \subset TM$, with the rank of the distribution being the dimension of the fibres.

Note: For the interested reader, a distribution D is called integrable if it admits a unique maximal integrable manifold through each point. An integral manifold N of a distribution D is a connected submanifold $N \subset M$ such that $f_*(T_pN) = D_p$ for all $p \in N$, where $f : N \to M$ is an imbedding. N is called maximal if there is no other integral manifold of D that contains N as a submanifold.

Definition 5.7. An n dimensional sub-Riemannian manifold is a real manifold M of dimensions n with a nonintegrable distribution D of rank k ($k < n$) endowed with a sub-Riemannian metric g, *i.e.,* a positive definite, nondegerenate, inner product $g_p : D_p \times D_p \to \mathbb{R}$ for all $p \in M$. Thus, a sub-Riemannian manifold is a triplet (M, D, g).

We note here that the distribution being nonintegrable is essential. Without it, it would imply that at least locally there existed a submanifold $S \subset M$ such that (S, g) would be a Riemannian manifold. Thus, the theory of sub-Riemannian manifolds would simply reduce to the Riemannian case. In addition, it is important that $k \neq n$ because otherwise D would be tangent to the manifold M and $D_p = T_pM$, reducing again the sub-Riemannian theory to the Riemannian case.

Much like in the case of Riemannian geometry, we can define the length of tangent vectors. However, because the sub-Riemannian metric g_p is only defined on the distributions D_p, we can only define the length for horizontal vectors $v \in D_p$:

$$||v||^2 = g_p(v, v) = \langle v, v \rangle$$

Similarly, we define the length of horizontal curves, i,e, curves $c : [0, t] \to M$ in M such that their tangent velocity vector $\dot{c}(s)$s in the horizontal distribution $\dot{c}(s) \in D_{c(s)}$ for all $s \in [0, t]$, as:

$$\ell(c) := \int_0^t g(\dot{c}(s), \dot{c}(s))^{1/2}\, ds$$

Using this length function, we can define the **subRiemannian distance** d_{cc}, called the **Carnot-Caratheodory distance** between two point $p, q \in M$ if there is a horizontal curve joining them.

Definition 5.8. Let $p, q \in M$ be two points joined by at least one horizontal curve $c : [0, t] \to M$ parametrised such that $c(0) = p$ and $c(t) = q$. Let the set of all such horizontal curves be denoted C. The Carnot-Caratheodory distance between p and q is defined as:

$$d_{cc}(p, q) := \inf\{\ell(\gamma) : \gamma \in C\}$$

Note that there is no guarantee that any two given points are connected by a horizontal curve. To examine such cases, we will need to introduce theorems which guarantee for us **horizontal connectivity** or other **global connectivity properties**. The most important of these is called the **Chow-Rashevskii theorem**. In order to introduce it, we will need define the **bracket generating condition**, also referred to as **Hörmander's condition** or completely non-integrability. .

Definition 5.9. A distribution D is bracket generating at a point $p \in M$ if there exists an integer $r \geq 0$ such that $D_p^r = T_pM$, where:

$$D_p^r = D_p^{r-1} + [D_p, D_p^{r-1}] \tag{5.3}$$

'**Note**: Here'Real" meaning locally M is homeomorphic to $\mathbb{R}^n$.

Note: Assume that $D = \{X_1, \ldots, X_m\}$, $m \leq n$, on a n-dimensional compact manifold M_n with smooth measure μ. Consider the sum of square of vector fields $\mathcal{L} = \sum_{j=1}^m X_j^2$. Obviously, $\mathcal{L}$ is not necessary elliptic.

Let $B_{\mathcal{L}}(x, \rho) = \big\{y \in M_n : \ d_{cc}(x, y) < \rho\big\}$ be a "ball" consists of all $y \in M_n$ that can be joined to x by a horizontal curve γ with $d_{cc}(x, y) < \rho$. Let $B_E(x, \rho)$ be an ordinary Euclidean ball of radius ρ about x. In [12], C. Fefferman and Phong showed that if D satisfies bracket generating condition of step Q, then there exists a constant $c_{\mathcal{L}} > 0$ such that

$$B_E(x, \rho) \subseteq B_{\mathcal{L}}\big(x, c_{\mathcal{L}}\rho^{\frac{1}{Q}}\big) \quad \forall\ x \in \mathcal{M}_n, \ \ 0 < \rho < 1. \tag{5.1}$$

In fact, (5.1) is equivalent to $\mathcal{L}$ satisfies the sub-elliptic estimate (3.6):

$$\big\||\nabla|^{\frac{2}{Q}} u\big\|_{L^2} \leq c_Q\Big\{\|\mathcal{L}u\|_{L^2} + \widetilde{c}_Q\|u\|_{L^2}\Big\}, \quad \forall\ u \in C^\infty(\mathcal{M}_n) \tag{5.2}$$

where $c_Q > 0$ and $\widetilde{c}_Q \geq 0$. Here $|\nabla|^{\frac{2}{Q}}$ is a pseudodifferential operator with symbol $|\xi|^{\frac{2}{Q}}$. In particular, if $Q = 1$, (3.6) and (5.2) reduce to elliptic estimates. .

with $D_p^1 = D_p$ and $[D_p, D_p^{r-1}] = \{[X, Y] : X \in D_p \wedge Y \in D_p^{r-1}\}$. The integer r is called the step of the sub-Riemannian manifold (M,D,g) at the point p. Since the step is independent of the metric g, we also say that r is the step of the distribution, which just tells us that at point p, $r-1$ number of iterated brackets are needed to span the tangent space T_pM.

The idea behind this is that we can form a sequence of ascendant linear subspaces of the tangent space T_pM of a given point, starting with simply the distribution at that point D_p. We then add to D_p all the vectors generated by the Lie bracket of vector fields evaluated at that point. This gives us the sequence:

$$\begin{aligned} D_p^1 &= D_p \\ D_p^2 &= D_p^1 + [D_p, D_p^1] \\ D_p^3 &= D_p^2 + [D_p, D_p^2] \\ &\vdots \\ D_p^r &= D_p^{r-1} + [D_p, D_p^{r-1}] = D_p + \sum_{i=1}^{r-1} [D_p, D_p^i] \end{aligned}$$

We are thus, in a sense, adding more and more vectors into the space D_p^i until we have all tangent vectors at the point p.

We could in principle formulate the bracket generating property in the language of vector fields on the entire manifold directly. In this setting, the sequence of linear spaces forms a so-called **associated Lie flag** of $C^\infty(M)$-Modules of the distribution $D \subset TM$:

$$\Gamma^1(D) \subset \Gamma^2(D) \subset \ldots \subset \mathfrak{X}(M)$$

where $\mathfrak{X}(M)$ is the set of all smooth vector fields on M which forms a Lie algebra when equipped with the Lie bracket $[\cdot, \cdot]$. The space $\Gamma^1(D) = \Gamma(D) \subset \mathfrak{X}(M)$ is the $C^\infty(M)$ module of vector fields on M tangent to D, *i.e.* it is the set of all vector fields whose vectors are always horizontal vectors. We then get the other spaces $\Gamma^i(D)$ by bracketing in more vector fields:

$$\Gamma^i(D) := \operatorname*{span}_{C^\infty(M)} \left\{ A_i(X_1, \ldots X_i) : X_1, \ldots, X_i \in \Gamma(D) \right\}$$

where $A_i(\cdot)$ ranges over all brackets of length at most i. This flag then gives us the *growth* of the set of vector fields tangent to a distribution with respect to the Lie bracket. We can then say that a distribution D is bracket generating if there is an integer r such that $\Gamma^r(D) = \mathfrak{X}(M)$. Thus, if D is bracket generating, $\Gamma(D)$ generates $\mathfrak{X}(M)$ as an algebra .

In short, a distribution is bracket generating if for any point $p \in M$, all tangent vectors can be represented as a linear combination of the nested Lie brackets of horizontal vector fields $X, Y, Z, R, \ldots$ evaluated at that point, *i.e.* , linear combinations of vectors:

$$X_p,\ [X, Y]_p,\ [X, [Y, Z]]_p,\ [X, [Y, [Z, R]]]_p, \ldots$$

Having met the bracket generating property, we can now introduce Chow's theorem or Chow-Rashevskii theorem.

Theorem 5.10. *(Chow–Rashevskii theorem). If D is a bracket generating distribution on a connected manifold M, then any two points can be joined by a horizontal piecewise curve.*

Note: It turns out that being a bracket generating distribution is one extreme of the growth behaviour of the set of vector fields tangent to the distribution. The other extreme is the distribution being involute in which $\Gamma(D)$ is said to form a Lie subalgebra of all vector fields on M.

For a proof, see [9] or Theorem 3.5.2 on pages 80-82 in [4]. These definitions and theorems lay the basics of the theory of sub-Riemannian geometry, which we can now apply to the special case of Grushin manifolds. The interested reader may find, besides the reference [4], also the references [25] and [2] useful.

5.3. **Grushin manifolds.** Grushin manifolds will play a key role in our understanding and analysis of the Grushin operator in the subsequent chapters. Grushin manifolds also played a key role in the understanding and development of sub-Riemannian geometry and its relation to Riemannian geometry. This is because it was once believed that sub-Riemannian geometry can be obtained from Riemannian geometry in some sort of limiting procedure, e.g. sub-Riemannian geodiescs can be obtained from the limit of Riemannian geodesics. It turns out however that some sub-Riemannian geodesics cannot be obtained from a limit procedure. However, Grushin manifolds are interested as they can indeed be viewed as a limit of Riemannian manifolds, see section 11.6 in [4]. In fact, while Grushin manifolds are strictly speaking sub-Riemannian manifolds not Riemannian ones, they can viewed as Riemannian manifolds equipped with a singular Riemmanian metric, making them a class of manifolds that lie between Riemannian and sub-Riemannian geometry.

Let us now give a precise definition of Grushin manifolds. Before we do that however, we will define the concept of a **regular point**. It is precisely because the distribution coincides with the tangent bundle on regular points that the horizontal distribution is largely unimportant, giving us a Riemannian flavour rather than a sub-Riemannian one.

Definition 5.11. Given an n-dimensional manifold M with n vector fields $\{X_1, \dots, X_n\}$ on it, we can define the set of singular point as the set of points $p \in M$ such that the vector given by the vector fields at p do not span the tangent space at p:

$$S = \{p \in M : \text{Span}\{X_1, \dots, X_n\}_p \neq T_pM\}$$

A point $p \in S$ is called singular and a point $p \in M \backslash S$ is called regular.

Recall from definition (5.9) the notion of the step of a distribution at a point $p \in M$, which 1 plus the number of iterated Lie brackets required to span the tangent space T_pM. Since for a regular point, the vector fields $\{X_i\}$ span the tangent space, *i.e.,* $\text{Span}\{X_1, \dots, X_n\}_p = T_pM$, the step is 1 as we do not need any iterated Lie brackets $[X_i[X_j[\dots[X_k, X_\ell]\dots]]]$ to span the entire tangent space. In contrast, for a singular point, the step is at least 2, since we will need at least $[X_i, X_j]$ to span the tangent space, if not more iterated brackets.

Definition 5.12. A Grushin manifold is the triplet $(M, \mathcal{D}, g)$, *i.e.,* a sub-Riemannian manifold, where $\mathcal{D} = \{X_i\}$ is the set of vector fields which give us a singular set S. The sub-Riemmanian metric is then defined as the Riemannian metric on the set of regular points $M \backslash S$ as:

$$g(X_i, X_j) = \delta_{ij}$$

Let us give some examples of Grushin manifolds before turning to study the Grushin operator, which gives us the associated Grushin geometry.

Example 5.13. Consider $\mathbb{R}^2$ with the vector fields $X_1 = \partial_x$ and $X_2 = x\partial_y$. This gives us a Grushin manifold with the metric:

$$g = \begin{pmatrix} 1 & 0 \\ 0 & 1/x^2 \end{pmatrix}$$

which is defined on the set of regular points $\mathbb{R}^2 \backslash S$, where $S = \{(0, y) : y \in \mathbb{R}\}$ is the y-axis. Readers can consult the paper [6] for detailed discussion of this subject. The step on S is 2 since we only need one bracket $[X_1, X_2] = [\partial_x, x\partial_y] = \partial_y$ to span the tangent space. For a complete discussion of step 2 Grushin operator, see [8].

Generalising the previous example, we can consider the following two vector fields instead.

Example 5.14. Consider $\mathbb{R}^2$ instead with $X_1 = \partial_x$ and $X_2 = x^k\partial_y$. This again gives us a Grushin manifold but with the step of the singular points S, again the y-axis, being $k+1$.

For more examples, see Examples (11.1.3) and (11.1.4) in [4] on page 272. These two examples above are however important as they can be associated with the (higher step) Grushin operators:

$$\Delta_G = \frac{1}{2}(X_1^2 + X_2^2) = \frac{1}{2}(\partial_x^2 + x^{2k}\partial_y^2)$$

This is a sum of square operator of the form:

$$\Delta_X = \frac{1}{2}\sum_{i=1}^{n} X_i^2$$

We can write their principal symbols in the form of:

$$H(p,x) = \frac{1}{2}\sum_{i=1}^{n}\langle X_i(x), p\rangle^2 = \frac{1}{2}\sum_{i,j=1}^{n} g^{ij}p_ip_j$$

where $\sum_k X_i^k(x)\partial_{x_k}$ and $\sum_k p^k\partial_{x_k}$, with $\langle\cdot,\cdot\rangle$ being the standard inner product on $\mathbb{R}^n$. For the second form, we have used $g(\partial_{x_i}, \partial_{x_j}) = g_{ij}$ and $g(X_i, X_j) = \delta_{ij}$. We have written the principal symbol is this second form to suggest the close analogy of Grushin manifolds and Riemannian manifolds. Applying these considerations to Δ_G, we find

$$\begin{aligned}
H(p,x) &= \frac{1}{2}\sum_{i=1}^{n}\langle X_i(x), p\rangle^2 = \frac{1}{2}\sum_{i=1}^{n}\left\langle \sum_k X_i^k(x)\partial_{x_k}, \sum_k p^k\partial_{x_k}\right\rangle^2 \\
&= \frac{1}{2}\sum_{i=1}^{2}\left\langle X_i^1(x)\partial_x + X_i^2(x)\partial_y, p^1\partial_x + p^2\partial_y\right\rangle^2 \\
&= \frac{1}{2}\left(\left\langle 1\cdot\partial_x + 0\cdot\partial_y, p^1\partial_x + p^2\partial_y\right\rangle^2 + \left\langle 0\cdot\partial_x + x^k\partial_y, p^1\partial_x + p^2\partial_y\right\rangle^2\right) \\
&= \frac{1}{2}\left((p^1)^2 + x^{2k}(p^2)^2\right)
\end{aligned}$$

which is precisely what we expected for the principal symbol for Δ_G.

Lastly, let us comment on why we have denoted the principal symbol as $H(p,x)$. This is suggestive of the fact that we will call this the Hamiltonian and use Hamiltonian mechanics to study and find the geodesics associated with the Grushin manifolds underlying the Grushin operator. To understand this, let us review some of the basic undergraduate analytical mechanics formalism. However, to give us a glimpse of where we are going with this formalism, let us comment on the fact that the *normal* geodesics on a Gurshin manifold can be found according to the following definition.

Definition 5.15. A normal geodesics connecting two points $P(x_0)$ and $Q(x_0)$ on a Gurshin manifold is the projection onto the x-space (as opposed to including also the p-space) of a solution to the Hamiltonian system defined by the principal symbol $H(p,x)$ and given by Hamilton's equations:

$$\dot{x} = H_p, \quad \dot{p} = -H_x$$

Note: We recommend readers to read a book [7] for another important examples which are related to geometric analysis on a family of higher-step hypersurfaces in $\mathbb{C}^2$.

Notice that we have referred to these geodesics as *normal*. This is because while in the Riemannian case, the characteristic curves of $H(p,x)$ are simply the locally length-minimising geodesics. However, in the sub-Riemannian case, there can exist *abnormal* geodesics, which are locally length-minimising geodesics which are not solutions to the associated Hamiltonian system. For an example, see the paper [26] by Liu and Sussmann or section 10.6 in [4].

We are justified in searching for normal geodesics since we know it is possible to find curves connecting any two points on this Grushin manifold. This is because the vector fields given in examples 5.13 and 5.14 satisfy the Chow–Rashevskii theorem, or Chow's condition, since the bracket generating condition holds for them. Thus, every two points on the (x,y)-plane can be joined by a piecewise curve whose tangent vector is a linear combination of X_1, X_2, giving us geodesics.

6. Analytical Mechanics and Hamilton-Jacobi Formalism

Before we jump into the Hamiltonian and Hamilton-Jacobi formalism which we will use throughout this thesis, we first need to give a general overview of the Lagrangian formalism. This is because the Hamiltonian is generally defined as the Legendre transform of the Lagrangian and the classical action used in the Hamilton-Jacobi formalism is defined as the time integral of the Lagrangian. For a deeper dive into these topics, We recommend the references [1], [3], and chapter 1 of [4].

6.1. Lagrangian formalism.

The state of a finite-dimensional classical mechanical system is in general described by only the positions $q(t)$ and velocities $\dot{q}(t)$ of its constituents according to Newton-Laplace's determinacy principle and other phenomena such as Ostrograskii instability. The space of possible/allowed positions is usually assumed to be a (Riemannian) manifold M, called the coordinate space or **configuration space**, with the Riemannian metric related to the kinetic energy expression of the system. The coordinates on M are called the **generalised coordinates** $q(t)$ and the number of coordinate functions needed, *i.e.,* $\dim M$, is called the number of **degrees of freedom**. The evolution of the system is then described by curves $\gamma(t)$ called **classical trajectories** in this configuration space. The velocities live in the tangent space of M, *i.e.,* $\dot{q} \in T_qM$ and hence the **space of states** $(q,\dot{q})$ is described by the tangent bundle TM of the configuration space.

On the tangent bundle, we can define a smooth, real-valued function called the **Lagrangian** $L(q,\dot{q},t)$:

$$L : TM \times \mathbb{R} \to \mathbb{R}$$

which for our purposes will always be expressed as the kinetic energy $K(\dot{q})$ minus the potential energy $U(q)$, which will simply be polynomial functions in the coordinates:

$$L(q,\dot{q}) = K(\dot{q}) - U(q)$$

The kinetic energy will be of the form $K(\dot{q}) = \frac{1}{2}\sum_{i,j=1}^{n=\dim M} g_{ij}(q)\dot{q}_i\dot{q}_j = \frac{1}{2}||\dot{q}||^2 = \frac{1}{2}\langle \dot{q},\dot{q}\rangle$, where the inner product $\langle\cdot,\cdot\rangle$ in fibres of TM is given by the Riemannian metric $g(q) : T_qM \times T_qM \to \mathbb{R}$, a positive, definite, symmetric, bilinear form.

We call a configuration manifold M together with a specified Lagrangian L a **Lagrangian system** (M,L). The Lagrangian describes the time evolution or dynamics of the system through the **action integral**:

$$S(q,t) = \int_0^t L(q(\tau),\dot{q}(\tau),\tau)\,d\tau$$

More specifically, the action can be defined as follows. Let $P(M) := \{\gamma : [t_0,t_1] \subset \mathbb{R} \to M \mid \gamma(t_0) = q_0, \gamma(t_1) = q_1\}$ be the space of all (smooth and parametrised) paths connecting q_0 and q_1 in M. $P(M)$ is an infinite dimensional real Frechet manifold. On $P(M)$ we define the **action functional** $S : P(M) \to$

$\mathbb{R}$ of a Lagrangian system (M, L) by:

$$S(\gamma) = \int_{t_0}^{t_1} \mathcal{L}(\gamma(t), \gamma'(t), t)\, dt$$

The time evolution comes into play through the postulate called the principle of stationary action, or Hamilton's principle, which states: A path $\gamma \in P(M)$ describes the motion of (M, L) between $q_0 \in M$ at t_0 and $q_1 \in M$ at t_1 iff γ is a critical/extremum of S, *i.e.,*

$$\delta_V S = \underbrace{\frac{d}{d\epsilon}\bigg|_{\epsilon=0}}_{\text{Gato der.}} S(\gamma_\epsilon) = 0, \quad \forall \text{ fixed end point variations } \gamma_\epsilon(t) \text{ of } \gamma(t) \tag{6.1}$$

where a fixed end point variation is the map $\gamma_\epsilon(t) \in P(M)$ for a given ϵ belonging the the family Γ of curves, *i.e.,* $\gamma_\epsilon(t) = \Gamma(t, \epsilon)$, satisfying:

$$\Gamma : [t_0, t_1] \times [-\epsilon_0, \epsilon_0] \to M$$

such that $\gamma_0(t) = \Gamma(t, 0) = \gamma(t)$ for $t_0 \le t \le t_1$ and $\gamma_\epsilon(t_0) = \Gamma(t_0, \epsilon) = q_0,\ \gamma_\epsilon(t_1) = \Gamma(t_1, \epsilon) = q_1$ for $-\epsilon_0 \le \epsilon \le \epsilon_0$. Note that the Gato derivative δ_V is simply the the directional derivative of a functional with respect to a tangent vector $V \in T_\gamma P(M)$, which in this case is the tangent vector of $\gamma_\epsilon(t)$ at $\gamma_0 = \gamma$ and is called the infinitesimal variation.

If a path $\gamma(t)$ satisfies condition (6.1), we call it an extremal point of the action and we say $\gamma(t)$ is the **motion** or classical path of the system. These extremals turn out to be characterised by a set of second order different equations called the **Euler-Lagrange equations**. We can express these equations in standard coordinates as follows. Let (U, φ) be a coordinate chart on M with local coordinates given by $q = (q^1, \dots, q^n)$, then the **standard coordinates** on a chart TU on TM are:

$$(q, v) = (q^1, \dots, q^n, v^1, \dots, v^n)$$

where $v = (v^1, \dots, v^n)$ are the coordinates in the fibre corresponding to the basis $\{\partial_{q^i}\}_{i=1}^n$ for T_qM. In these standard coordinates on TU, the Euler-Lagrange equations are:

$$frac\partial L\partial q(q(t), \dot{q}(t), t) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t), t)\right) = 0$$

6.2. **Hamiltonian formalism.** Rather than working in the tangent bundle to study the dynamics of a system for a given (M, L), we can move to the cotangent bundle T^*M, called the **phase space** instead. Assuming the correct (non-degeneracy/convexity) conditions, this is done using the **Legendre transform**, which can be thought of as a (fibre-wise) map $\tau_L : TM \to T^*M$ from the tangent bundle to the cotangent bundle by replacing your velocity variables $\dot{q}$ with momenta $p = \frac{\partial L}{\partial \dot{x}}$. This results in an alternative formulation of dynamics called **Hamiltonian mechanics**, named after the Legendre transform of the Lagrangian, *i.e.,* the Hamiltonian:

$$H : T^*M \times \mathbb{R} \to \mathbb{R}$$

which for our purposes will be:

$$H(q, p, t) = p\dot{q} - L(q, \dot{q}, t) = K + U$$

Note: To be precise, we in general want the Legendre transform $\tau_L : TM \to T^*M$ of a Lagrangian L to be a diffeomorphism. This occurs iff the L is non-degenerate, *i.e.,* for every chart U of M the $n \times n$ Hessian matrix $H_L(q, \dot{q} := \{\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\}_{i,j=1}^n$ is invertible on TU.

As mentioned above, given a differential operator, we can jump right into the Hamiltonian formalism by simply defining the Hamiltonian associated to a differential operator as the principle symbol. The Lagrangian associated to the operator is then found by the inverse Legendre transform:

$$L(q, \dot{q}, t) = p\dot{q} - H(q, p, t)$$

It turns out that if the Legendre transform τ_L is a diffeomorphism, then the Euler-Lagrange equations in standard coordinates on TU are equivalent to the so-called **Hamilton's equations**, a first-order system of ordinary differential equation written in standard coordinates on T^*U as:

$$\begin{cases} \dot{p}_i = -\frac{\partial H}{\partial q^i} \\ \dot{q}^i = \frac{\partial H}{\partial p^i} \end{cases}$$

for $i = 1, \ldots, \dim M$. A solution to $(q(t), p(t))$ to Hamilton's equations is called a **bicharacteristic** curve. Accordingly, any projection of the bichracteristic curve in the cotangent space to a curve $q(t) = \gamma(t) \subset M$ in the configuration space gives a solution to the Euler-Lagrange equations. Conversely, given curve $\gamma(t)$ which is a solution to the Euler-Lagrange equations, then $(\gamma(t) = q(t), \frac{\partial L}{\partial \dot{q}(t)}(t))$ gives a bicharacteristic curve called the **lift** fo $\gamma(t)$, since we are lifting a curve from the underlying manifold to the cotangent space.

Often we will refer to the solutions of Hamilton's equations and solutions of the Euler-Lagrange equations as **geodesics**. This is a bit of an abuse of vocabulary, but we will ignore that .

6.2.1. **Conservative systems and operators with potentials.** Lastly, we note that for **conservative systems**, *i.e.,* when H is not explicitly dependent on t, then H is constant along bicharacteristic curves, *i.e.,* evaluated on $(q(t), p(t))$. We define H evaluated on $(q(t), p(t))$ as the total energy E of the system and say it is a first integral of motion. This is in fact the case for Hamilton's of the form:

$$H(x, p) = \frac{1}{2}p^2 + U(x) \tag{6.2}$$

These Hamiltons arise from the principle symbols of **operators with potentials**, *i.e.,* operators of the form $\mathbb{L} = \frac{1}{2}\partial_x^2 + U(x)$. Since these Hamiltons are conservative, we can use the fact that the energy is constant along the solutions of Hamilton's equations to solve Hamilon's equations. This is called the **energy method**. It is in fact just a fancy form of separation of variables and goes as follows.

Note: There are many technical definitions of what a geodesics which are mostly equivalent. In the context of sub-Riemannian manifolds and hence Riemannian geometry as well, geodesics are simply projections of horizontal curves in the bundle onto your manifold. If we use the free Hamiltonian $H = 1/2 g^{ij} p_i p_j$, then Hamilton's equations define a cogeodesic flow (*co-* for the flow being in the *co*tangent bundle) and the corresponding Euler-Lagrange equations define a geodesic flow in the tangent bundle. The *geodesic lines* are then projections of these geodesic flows in the bundle onto your manifold. There is another definition for geodesics which is often used which defines geodesics as critical points of an energy functional. This definition is equivalent to ours above. Geodesics can also be defined as solutions $\gamma(t)$ to the geodesic equation $\nabla'_\gamma(t)\gamma'(t) = 0$, where $\gamma'(t)$ is the velocity vector of the curve (hence since $\nabla'_\gamma(t)\gamma'(t)$ is interpreted as the acceleration, we can say that geodesics are curves with zero acceleration). This geodesics equation is however just the Euler-Lagrange equations of the energy functional, *i.e.,* its solutions are the critical points of the functional, and hence this definition is also equivalent to ours. However, we should never define our geodesics are curves which are length minimisers, not even in the Riemannian seeting. This is because even on a sphere, both directions around a great circle connecting two points are geodesics according to the definitions above but both are not length (global) minimisers. One could instead say that geodesics are only *critical points* of a length functional $L(\gamma) = \int \sqrt{g(\gamma'(t), \gamma'(t)}dt$. This works in the Riemannian case since here critical points of the length functional are critical points of your energy functional (assuming we restrict ourselves to only constant speed curves, otherwise this is not true). However, for the Semi-Riemannian case, while we can define a 'length functional' analogous to the Riemannian case, it's not a 'length' or distance since this functional satisfies the reverse triangle inequality and not the triangle inequality.

The Hamilton equations associated with (6.2) are:

$$\begin{cases} \dot{x} = H_p = p \\ \dot{p} = -H_x = -U'(x) \end{cases}$$

implying that $\ddot{x} = \dot{p} = -U'(x)$. As a result, classical paths between two points $x(0) = x_0$ and $x(t) = x_1$ are solutions to the ODEs problem:

$$\begin{cases} \ddot{x} = -U'(x) \\ x(0) = x_0 \\ x(t) = x_1 \end{cases} \tag{6.3}$$

Given standard ODE theory, we can savely assume there is a unique solution to (6.3) which can be solved by setting:

$$E = \frac{1}{2}\left(\frac{dx}{ds}\right)^2 + U(x)$$

and rearranging to get:

$$s = \int_{x_0}^{x_1} \frac{dx}{\sqrt{2E - 2U(x)}} \tag{6.4}$$

Solving this integral and inverting it to solve for $x(s)$ amounts to finding the geodesics and solving (6.3).

6.3. **Hamilton-Jacobi formalism.** Another alternative formulation of classical mechanics is known as the Hamilton-Jacobi formalism. In this formalism, the **classical action** S_{cl} is the central player which is defined as the integral of the Lagrangian L along the solution to the Euler-Lagrange equation. It is found by taking the classical path $(q_{cl}(t), \dot{q}_{cl}(t))$ between two given points $q(0) = q_0$ and $q(t) = q_1$ and evaluating:

$$S_{cl}(q_0, q_1; t) := \int_0^t L(q_{cl}(\tau), \dot{q}_{cl}(\tau), \tau)\, d\tau$$

Alternatively, this classical action can be found as the solution to the so-called **Hamilton-Jacobi equation**

$$\frac{\partial}{\partial t} S_{cl} + H\left(q, \frac{\partial S_{cl}}{\partial q}\right) = 0$$

which is a nonlinear partial differential equation and therefore in general may not have unique solutions. Note that in the notation of the Hamilton-Jacobi equation, we have substituted the momentum variable in the Hamiltonina with the gradient of the classical action $p = \partial_q S_{cl} = \nabla S_{cl}$

6.3.1. **Conservative systems and operators with potentials.** For a conservative system, *i.e.,* when $H = E$ is the total energy along the geodesics, the Hamilton-Jacobi equation simplies to:

$$\partial_t S_{cl} = -E$$

Furthermore, if we consider again a one-dimensional system with Hamiltonian of the form $H(p, x) = \frac{1}{2}p^2 + U(x) = E$, we can solve for $\dot{x} = \sqrt{2E - 2U(x)}$. Then, since along the geodesics $p = \partial_x S_{cl}$ and $\dot{x} = p$, we obtain:

$$\partial_x S_{cl} = \dot{x} = \sqrt{2E - 2U(x)}$$

In other words, we can simply solve for the classical action as:

$$S_{cl} = \int_{x_0}^{x_1} \sqrt{2E - 2U(x_{cl})}\, dx$$

7. Path Integrals and Van Vleck's Formula

In the previous section we reviewed some of the standard approaches to studying classical mechanical systems. The reason for this was clear as we can study partial differential equations through the principal symbol of their differential operators which we define to be the Hamiltonian associated to the differential operator. However, beyond the Hamilton formalism being useful for finding the geodesics or solutions to Hamilton's equations, why would we need to the Hamilton-Jacobi formalism? Why would we be interested in solving for the classical action using the Hamilton-Jacobi equation?

The answer is that we can express the heat kernel of an operator $\mathbb{L}$ using the path integral approach through the following formula which gives the heat kernel as simply the propagator in terms of Feynman's path integral:

$$K(x_0,x;t)=\int_{\mathcal{B}_{x_0,x;t}} e^{-S(\phi,t)}d\mathfrak{m}(\phi)=e^{-S_{cl}(x_0,x;t)}\int_{\mathcal{B}_{0,0;t}} e^{-S(\psi,t)}d\mathfrak{m}(\psi) \tag{7.1}$$

where $\mathcal{B}_{x_0,x;t}:=\{\phi:[0,t]\to\mathbb{R}^n:\phi(0)=x_0,\phi(t)=x\}$, *i.e.,* the space of horizontal curves ϕ between the points x_0 and x parametrised on the interval $[0,t]$. The classical action is then $S(\phi,t)$ evaluated along the horizontal curve $\phi\in\mathcal{B}_{x_0,x;t}$. The measure $d\mathfrak{m}(\phi)$ is the Wiener measure along the horizontal distributions, *i.e.,* on the space $\mathcal{B}_{x_0,x;t}$, and is obtained by a limit process $\lim_{n\to\infty}d\mathfrak{m}(\phi_n)$, where $d\mathfrak{m}(\phi_n)$ is given by a product of $d\phi_n$'s. For the derivation of this formula using the Trotter formula in addition to discussion on the existence of such a measure, see chapter 7 in [5]. Note that our definition of the path integral has been Wick rotated, *i.e.,* we have replaced $t\to it$ in the exponent of the path integral integrand.

It turns out that in many cases, such a second-order elliptic operators with constant, linear, and quadratic potential, the second integral $\int_{\mathcal{B}_{0,0;t}} e^{-S(\psi,t)}d\mathfrak{m}(\psi)$ is a function of t only, *i.e.,*

$$\int_{\mathcal{B}_{0,0;t}} e^{-S(\psi,t)}d\mathfrak{m}(\psi)=V(t).$$

We use V to denote this function in representation of its name **volume function** or **volume element**. It tells us how the geodesics spread away from the initial point x_0 as we move along the parameter of the geodesics. Accordingly, it satisfies a transport equation, usually of the form $V'(t)+V(t)\mathbb{L}(S_{cl})=0$. More precisely, the square of $V(t)$ gives us the density of the geodesic flow, *i.e.,* how geodesics spread away from a given geodesics. This square is given in many cases by the determinant of the rate of change of the action, *i.e.,* the **van Vleck-Morette determinant**:

$$V^2(t)=\det\left(-\frac{1}{2\pi}\frac{\partial^2 S_{cl}(x_0,x)}{\partial x_0\partial x}\right) \tag{7.2}$$

This is called **van Vleck's formula** and can be used to write the heat kernel/propagator (7.1) in an even simpler form:

$$\begin{aligned} K(x_0,x;t)&=\int_{\mathcal{B}_{x_0,x;t}} e^{-S(\phi,t)}d\mathfrak{m}(\phi)=e^{-S_{cl}(x_0,x;t)}\int_{\mathcal{B}_{0,0;t}} e^{-S(\psi,t)}d\mathfrak{m}(\psi)\\ &=\sqrt{\det\left(-\frac{1}{2\pi}\frac{\partial^2 S_{cl}(x_0,x)}{\partial x_0\partial x}\right)}e^{-S_{cl}(x_0,x,t)}=V(t)e^{-S_{cl}} \end{aligned} \tag{7.3}$$

Notice therefore that we have evaluated the path integral to be essentially some prefactor $V(t)$, which is independent of x_0 and x and is given by some determinant, times a typical soliton $e^{-S_{cl}}$ solution. We can therefore think of the Van Vleck formula as a semiclassical approximation to the path integral, see Section 16 in [27]. For more on this formula, see section 7.9 in [5].

Note: See Theorem 3.1.1 in [5], Theorem 7.3.1, part 7 in [5], or Theorem 7.3.2, part 7 in [5].

What is the underlying intuition of van Vleck's formula when applied to PDE operator theory? The general idea is that each operator $\mathbb{L}$, as we have discussed previously, defines or is associated to an underlying geometry (the principal symbol gives the metric). Investigating how the geodesics move, *i.e.,* the geodesics flow, in this geometry can be done with the help of the heat kernel of the operator, as they are fundamental solutions which give us a way of finding all the solutions to the PDE defined by the given operator. In fact, the heat kernel is *the* fundamental solution to the associated heat operator $(\partial_t - \mathbb{L})$. The heat kernel is amply named as you can think of the heat kernel as giving us a sort of *heat flow* which propagates along the geodesics, the groves and edges in the underlying geometry/geometric structure. The density of this heat flow is described by the volume function/element which satisfies a transport equation. This transport equation one can think of loosely as an analogue to a continuity equation describing fluid flow but for the density of heat flow, see Section 3.2 in [5]. This volume element is therefore given precisely by the van Vleck formula in the path integral approach which tells us the density of the geodesics paths in the underlying geometry.

8. Elliptic Integrals and Elliptic Function Theory

In this section, we will give a brief review of the theory of elliptic integrals and functions. This is necessary as when we try to solve for the geodesics of certain elliptic operators using the energy method, *i.e.,* fancy separation of variables, we will obtain an elliptic integral which is not expressible in terms of elementary functions. Instead what we can do is express such elliptic integrals in terms of special functions including elliptic functions. In addition, some of Chang and his collaborators previous results use elliptic functions, e.g. [4], and therefore it is useful to review these elliptic functions to understand such previous results.

A good reference for the following overview is Derek F. Lawden's *Elliptic Functions and Applications* (1989) [24].

8.1. Theta functions.

Following Lawden's approach, before introducing the Jacobi's elliptic functions and discussing elliptic integrals, it is useful to obtain some intuition on how they arise. For this, we will first derive expressions for the so-called (Jacobi) **theta functions**, which can physically be interpreted as functions describing the evolution of the temperature on some domain subject to certain boundary conditions, as solutions to the heat equation on the domain $[0, \pi]$:

$$\kappa \nabla^2 \theta = \frac{\partial \theta}{\partial t}$$

Here κ is called the diffusivity and is assumed to be constant. We can simplify the heat equation by assuming that there is no variation in temperature in the x, y-directions, hence reducing the gradient term to just a derivative in the z-direction:

$$\kappa \frac{\partial^2 \theta}{\partial z^2} = \frac{\partial \theta}{\partial t}$$

If we want to solve this partial differential equation, we will need to assume some boundary conditions. Different boundary conditions will result in different solutions, *i.e.,* different θ-functions:

(1) **Dirichlet boundary conditions.** If we assume the ends of the material are held at constant temperature $\theta = 0$ for $z = 0, \pi$ for all t, *i.e.,* $\theta(0, t) = \theta(\pi, t) = 0 \; \forall t \in \mathbb{R}_{\geq 0}$, then separation of

Note: As opposed to other forms of theta functions such as Neville theta functions. They are named after Carl Gustav Jakob Jacobi who published a text about elliptic function in 1829 [21].

variables tells us that:

$$\theta(z,t) = \sum_{n=1}^{\infty} b_n e^{-n^2\kappa t} \sin nz$$
$$b_n = \frac{2}{\pi}\int_0^{\pi} f(z)\sin nz dz$$

where b_n are Fourier coefficients. If we assume $f(z) = \pi\delta(z - \frac{1}{2}\pi)$, then we get $b_n = 2\sin(n\pi/2)$ and so:

$$\theta(z,t) = 2\sum_{n=1}^{\infty}(-1)^n e^{-(2n+1)^2\kappa t}\sin[(2n+1)z]$$
$$= \boxed{2\sum_{n=0}^{\infty}(-1)^n q^{(n+1/2)^2}\sin[(2n+1)z] = \theta_1(z,q)}$$

where in the second-to-last step we used $q := e^{-4\kappa t}$. This gives us the **first theta function**. It is defined by an infinite series valid for all complex values of z and q such that $|q| < 1$. Often it is convenient to write the complex number q as $q = e^{i\pi\tau}$ where q is called the **nome** of the theta function and τ the **parameter**. Therefore, the first theta function is often also written as $\theta_1 = \theta_1(z|\tau)$. Note that there are other ways to express the first theta function, as is also the case for the other theta functions, as for example with a doubly infinite series.

(2) **Neumann boundary conditions.** If we assume assume that the boundaries are isolated, *i.e.,* $\partial\theta/\partial z = 0$ for $z = 0, \pi$, then separation of variables gives us:

$$\theta(z,t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n e^{-n^2\kappa t}\cos nz$$
$$a_n = \frac{2}{\pi}\int_0^{\pi} f(z)\cos nz dz$$

In the case of $f(z) = \pi\delta(z - \frac{1}{2}\pi)$, we get $a_n = 2\cos(n\pi/2)$ and so:

$$\theta = \boxed{\theta(z,q) = 1 + 2\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos[2nz] =: \theta_4(z,q)}$$

This gives us **fourth theta function**. Again, it is defined by an infinite series.

Similarly, we can define the second and third theta functions θ_2 and θ_3 as translations of θ_1 and θ_4 by $\pi/2$ in the z-direction:

$$\boxed{\theta_3(z,q) := 2\sum_{n=0}^{\infty} q^{(n+1/2)^2}\cos[(2n+1)z]}$$

and

$$\boxed{\theta_3(z,q) := 1 + 2\sum_{n=1}^{\infty} q^{n^2}\cos[2nz]}$$

There are many useful and interesting properties of these theta functions which for the interested reader are summarised in Sections 1.3-1.6 of Chapter 1 in [24].

For visual intuition, the graphs of the four theta functions with $q = 0.7$ and $z \in [0, 12]$ can be seen in Figure 1.3 in [5], reproduced below.

Note: See Equation (1.2.1) on page 4 in [24].

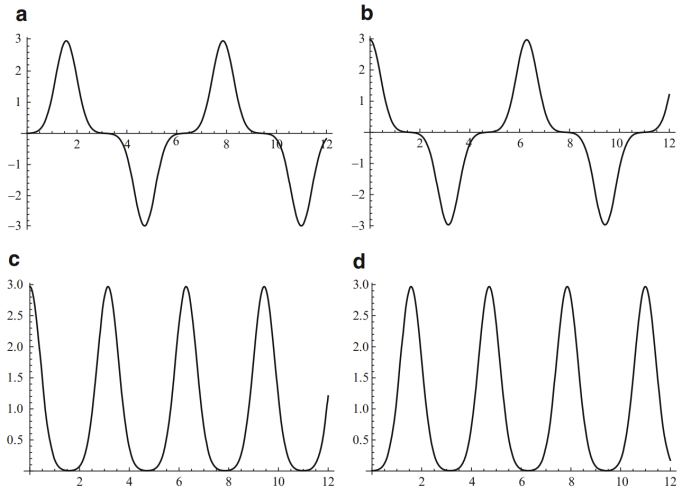

FIGURE 1. Graphs of the four theta functions with **(a)** $\theta_1(z, q = 0.7)$, **(b)** $\theta_2(z, q = 0.7)$, **(c)** $\theta_3(z, q = 0.7)$, **(d)** $\theta_4(z, q = 0.7)$. Source: Figure 1.3 on Page 9 in [5]

8.2. **Jacobi's elliptic functions.** From the four theta functions defined in the previous section, we can introduce our first elliptic functions, called **Jacobi's Elliptic functions**. The Jacobi Ellitpic functions are a set of twelve elliptic functions. The three main Jacobi's elliptic functions are defined as the ratios of theta functions:

$$\begin{aligned}
\mathrm{sn}(u) &:= \frac{\theta_3(0)}{\theta_2(0)}\frac{\theta_1(z)}{\theta_2(z)} \\
\mathrm{cn}(u) &:= \frac{\theta_4(0)}{\theta_2(0)}\frac{\theta_2(z)}{\theta_4(z)} \\
\mathrm{dn}(u) &:= \frac{\theta_4(0)}{\theta_3(0)}\frac{\theta_3(z)}{\theta_4(z)}
\end{aligned}$$

These functions have a lot of interesting properties summarised in Chapter 2 in [24]. We will however only mention a few of them, such as the fact that they are doubly periodic in u, e.g. in the case of $\mathrm{sn}(u)$:

$$\begin{cases}
\mathrm{sn}(u + 2\pi\theta_3^2(0)) = \mathrm{sn}(u) \\
\mathrm{sn}(u + \pi\tau\theta_3^2(0)) = \mathrm{sn}(u)
\end{cases}$$

This allows us to define the quantities K and iK' which are called the **real quarter period** and **imaginary quarter period** of the elliptic functions:

$$\begin{cases}
K = \frac{1}{2}\pi\theta_3^2(0) \\
iK' = \frac{1}{2}\pi\tau\theta_3^2(0) = \tau K
\end{cases}$$

Note: An elliptic function is simply defined as a meromorphic function that is doubly periodic, *i.e.,* $\exists\omega_1, \omega_2 \in \mathbb{C}$ where ω_i are linearly independent vectors when viewed in $\mathbb{R}^2$ such that $f(z + \omega_i) = f(z)\ \forall z \in \mathbb{C}$.

Note: In fact, K and iK' are special functions, which can be defined as follows:

$$K = K(m) = \int_0^{\pi/2} \frac{d\theta}{1 - m\sin^2\theta}$$

where this integral is simply an elliptic integral of the first kind with parameter m, and:

$$iK'(m) = iK(1 - m).$$

We can therefore see why these quantities are called *quarter* periods since the three main Jacobi elliptic functions are periodic in $4K$ and $4iK'$:

$\mathrm{sn}(u)$	$4K, 2iK'$	$\mathrm{sn}(u) = \mathrm{sn}(u+4K) = \mathrm{sn}(u+2iK')$
$\mathrm{cn}(u)$	$4K, 2K+2iK'$	$\mathrm{cn}(u) = \mathrm{cn}(u+4K) = \mathrm{cn}(u+2K+2iK')$
$\mathrm{dn}(u)$	$4K, 2K+2iK'$	$\mathrm{dn}(u) = \mathrm{dn}(u+4K) = \mathrm{dn}(u+2K+2iK')$

The other nine can be defined as reciprocals and quotients of these three:

$$\mathrm{ns}\,u = \frac{1}{\mathrm{sn}\,u}, \qquad \mathrm{nc}\,u = \frac{1}{\mathrm{cn}\,u}, \qquad \mathrm{nd}\,u = \frac{1}{\mathrm{dn}\,u},$$
$$\mathrm{sc}\,u = \frac{\mathrm{sn}\,u}{\mathrm{cn}\,u}, \qquad \mathrm{cd}\,u = \frac{\mathrm{cn}\,u}{\mathrm{dn}\,u}, \qquad \mathrm{ds}\,u = \frac{\mathrm{dn}\,u}{\mathrm{sn}\,u},$$
$$\mathrm{cs}\,u = \frac{\mathrm{cn}\,u}{\mathrm{sn}\,u}, \qquad \mathrm{dc}\,u = \frac{\mathrm{dn}\,u}{\mathrm{cn}\,u}, \qquad \mathrm{sd}\,u = \frac{\mathrm{sn}\,u}{\mathrm{dn}\,u}.$$

There is a lot of deep theory and notational ingenuity regarding these twelve Jacobi functions. Let us give just a brief taste of it.

The twelve Jacobi elliptic functions are all denoted in the form $\mathrm{pq}(u)$, where p and q are any of the letters c, s, n, or d. The variable u is called the argument and can in general be complex. We can therefore graph the u plane as simply the complex plane. It turns out that the Jacobi functions are meromorphic functions in this plane and have special properties regarding the distribution of zeros and poles in the u-plane, arising from the fact that they are doubly periodic in u. These zeros and poles form a repeating lattice or parallelogram with sides lengths $2K$ or $4K$ on the real axis and $2K'$ or $4K'$ on the imaginary axis, depending on the Jacobi elliptic function. This lattice is called the **period** or **fundamental parallelogram**. For simplicity, we will draw this period parallelogram as a rectangle between the origin $(0, 0)$ and the corner (K, K'), keeping in mind it may in general be a parallelogram, looking slightly different for each Jacobi elliptic function.

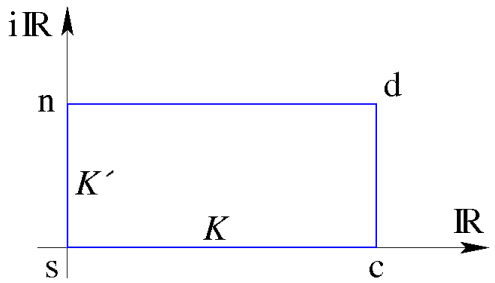

FIGURE 2. The period rectangle of Jacobi elliptic functions in the complex u-plane [13]

The rectangle's corners are labeled s, c, d, and n with the Jacobi elliptic function $\mathrm{pq}(u)$ having a zero at the corner labeled p and a simple pole at the corner q. This rectangle is repeating all throughout the complex u-plane since the function $\mathrm{pq}(u)$ turns out to be periodic in the direction pq with period $2(\mathrm{p}-\mathrm{q})$ and periodic in the other two directions pq' and pq'' with periods $4(\mathrm{p}-\mathrm{q}')$ and $4(\mathrm{p}-\mathrm{q}'')$. As a result, we can in fact view the domain of the Jacobi elliptic functions as being a (complex) torus, rather than a plane, with the circumference of the two circles forming the torus being $4K$ and $4K'$.

Functions where $p = q$, *i.e.,* $\mathrm{pp}(u)$ are simply unity.

Technically the Jacobi elliptic functions are also functions of a second variable, usually denoted by m, called the **parameter**, which may also be complex, or by k, called the **elliptic modulus**. They are related by $k^2 = m$. Sometimes the second variable is also given in terms of the **modular angle** α where $m = \sin^2\alpha$. In addition, sometimes the first variable is given by the **amplitude** denoted φ rather than the argument u.

This in fact happens when u is real and the parameter is real in $0 < m < 1$. For the more general period parallelogram cases, see pages 32 and 33, Figures 2.4 and 2.5 in [24].

This is similar to how trigonometric sine and cosine functions are defined on a circle. In the case of the Jacobi elliptic functions, there are two circles giving us a torus. There are other ways in which the Jacobi elliptic functions are related to the usual trigonometric functions, which is reflected in the notation. For example, when we send the modulus $k \to 0$, where $k = -m^2 = \theta_2^2(0)/\theta_3^2(0)$, we recover the usual trigonometric functions:

$$\begin{cases} \mathrm{sn}(u,k) \to \sin u \\ \mathrm{cn}(u,k) \to \cos u \\ \mathrm{dn}(u,k) \to 1 \\ \mathrm{sc}(u,k) \to \tan u \end{cases}$$

Furthermore, due to the numerical factors used in front of the definitions of the three main Jacobi elliptic functions, e.g. $\theta_3(0)/\theta_2(0)$ for $\mathrm{sn}(u)$, we obtain similar identities to the trigonometric case, e.g.:

$$\mathrm{sn}^2 u + \mathrm{cn}^2 u = 1$$

In addition, the numerical factor in front of $\mathrm{cn}u$ has been chosen so that much like the trig cosine, we have $\mathrm{cn}(0) = 1$. As a result, the three main Jacobi elliptic functions are also called the **elliptic sine**, the **elliptic cosine**, and the **elliptic delta**.

For more on the algebraic properties such as analogue double angle formulas (so-called addition theorems), derivatives, integrals, Taylor expansions, and discussion of the poles and other properties of the Jacobi elliptic functions, see Chapter 2 in [24]. We summarise the most important formulas in the table below. The graphs of the three main Jacobi functions are also given on page 29 in [24]. Lastly, since the inverses of the Jacobi elliptic functions are best motivated by discussing elliptic integrals, we will introduce the inverses in the next section.

Derivatives	Integrals	Taylor Expansions around $u = 0$
$\frac{d}{du}\,\mathrm{sn}(u) = \mathrm{cn}(u)\,\mathrm{dn}(u)$	$\int \mathrm{sn}(u)\,du = \frac{1}{k}\ln(\mathrm{dn}(u) - k\mathrm{cn}(u))$	$\mathrm{sn}(u) = u - \frac{1}{3!}(1+k^2)u^3 + \frac{1}{5!}(1+14k^2+k^4)u^5 - \dots$
$\frac{d}{du}\,\mathrm{cn}(u) = -\mathrm{sn}(u)\,\mathrm{dn}(u)$	$\int \mathrm{cn}(u)\,du = \frac{1}{k}\sin^{-1}(k\mathrm{sn}(u))$	$\mathrm{cn}(u) = 1 - \frac{1}{2!}u^2 + \frac{1}{4!}(1+4k^2)u^4 - \dots$
$\frac{d}{du}\,\mathrm{dn}(u) = -k^2\mathrm{sn}(u)\,\mathrm{cn}(u)$	$\int \mathrm{dn}(u)\,du = \sin^{-1}(\mathrm{sn}(u))$	$\mathrm{dn}(u) = 1 - \frac{1}{2!}k^2u^2 + \frac{1}{4!}(4k^2+k^4)u^4 - \dots$
$\frac{d}{du}\,\mathrm{sc}(u) = \mathrm{dc}(u)\,\mathrm{sn}(u)$	$\int \mathrm{sc}(u)\,du = -\ln(\mathrm{ns}(u) + d\mathrm{su})$	$\mathrm{sc}(u) = \frac{\mathrm{sn}(u)}{\mathrm{cn}(u)}$
$\frac{d}{du}\,\mathrm{ds}(u) = -\mathrm{cs}(u)\,\mathrm{ns}(u)$	$\int \mathrm{ds}(u)\,du = \ln(\mathrm{ns}(u) - \mathrm{cs}(u))$	$\mathrm{ds}(u) = \frac{\mathrm{sn}(u)}{\mathrm{dn}(u)}$
$\frac{d}{du}\,\mathrm{cd}(u) = -k^2\mathrm{sd}(u)\,\mathrm{nd}(u)$	$\int \mathrm{cd}(u)\,du = \ln(\mathrm{nc}(u) + \mathrm{sc}(u))$	$\mathrm{cd}(u) = 1 - \frac{1}{2!}k^2u^2 + \dots$
$\frac{d}{du}\,\mathrm{sd}(u) = \mathrm{cd}(u)\,\mathrm{nd}(u)$	$\int \mathrm{sd}(u)\,du = \ln(\mathrm{cn}(u) + \mathrm{sc}(u))$	$\mathrm{sd}(u) = \frac{\mathrm{cn}(u)}{\mathrm{sn}(u)}$
$\frac{d}{du}\,\mathrm{ns}(u) = -\mathrm{cs}(u)\,\mathrm{ds}(u)$	$\int \mathrm{ns}(u)\,du = \ln(\mathrm{ds}(u) + \mathrm{cs}(u))$	$\mathrm{ns}(u) = \frac{1}{\mathrm{sn}(u)}$
$\frac{d}{du}\,\mathrm{nd}(u) = k^2\mathrm{sd}(u)\,\mathrm{cd}(u)$	$\int \mathrm{nd}(u)\,du = \sin^{-1}(\mathrm{cd}(u))$	$\mathrm{nd}(u) = \frac{\mathrm{sn}(u)}{\mathrm{cn}(u)}$
$\frac{d}{du}\,\mathrm{dc}(u) = -k^2\mathrm{sd}(u)\,\mathrm{nd}(u)$	$\int \mathrm{dc}(u)\,du = \ln(\mathrm{nc}(u) + \mathrm{sc}(u))$	$\mathrm{dc}(u) = \frac{1}{\mathrm{cn}(u)}$
$\frac{d}{du}\,\mathrm{dc}(u) = k^2\mathrm{sc}(u)\,\mathrm{nc}(u)$	$\int \mathrm{dc}(u)\,du = \ln(\mathrm{nc}(u) + \mathrm{sc}(u))$	$\mathrm{dc}(u) = \frac{1}{\mathrm{sc}(u)}$

8.3. **Elliptic integrals.** Elliptic integrals will repeated appear when computing the energy integrals of the form (6.4). Therefore, it is important for us to have an understanding of how to express such integrals using special functions such as elliptic functions.

Definition 8.1. Let R be a rational function of its two argument and $P(t)$ a cubic or quartic polynomial with no repeated roots (all zeros are different), then an integral of the form:

$$\int R\left(t, \sqrt{P(t)}\right) dt \tag{8.1}$$

is called a general elliptic integral. Given boundary conditions with c some constant, we can call a function $f(x)$ an elliptic integral if we can express it in the form:

$$f(x) = \int_c^x R\left(t, \sqrt{P(t)}\right) dt \tag{8.2}$$

In general, elliptic integrals cannot be expressed in terms of elementary functions. We can think of this as telling us that the space of functions which we call elementary does not span all possible functions which we may write down and hence we need to introduce more basis functions in order to be able to express elliptic integrals in terms of some set of basis functions. It turns out, as is shown in Section 3.3 in [24], that all integrals of the form (8.1) can be expressed in terms of integrals over rational functions which we can express in terms of elementary functions and a set of three elliptic integral basis functions called **Legendre canonical forms**. The three Legendre canonical forms are also called the elliptic integrals of first, second, and third kind. They are obtained by setting the rational integrand $R(t, y) = P(t, y)/Q(t, y)$, where we have set $y^2 = P$, and finding that we have express R as:

$$R(t, y) = y^{-1}R_1(t) + R_2(t)$$

The integral over R_2 can be expressed in terms of elementary function, while the integral over $y^{-1}R_1$ can be expressed as:

$$\int y^{-1}R_1(t)\, dt = \int R_3(t)[(A_1t^2 + B_1)(A_2t^2 + B_2)]^{-1/2}\, dt = \int R_3(t)X^{-1/2}\, dt \tag{8.3}$$

where $X := (A_1t^2 + B_1)(A_2t^2 + B_2)$ and R_3 is again a rational function and can be expressed as in terms of even and odd components:

$$R_3(t) = \frac{P(t)}{Q(t)} = R_4(t^2) + tR_5(t)$$

with R_4, R_5 rational. Plugging this separation for $R_3(t)$ into the integral (8.3) above, we find it separated into two integrals:

$$\int R_3(t)X^{-1/2}\, dt = \int R_4(t^2)X^{-1/2}\, dt + \int tR_5(t^2)X^{-1/2}\, dt$$

The second integral can be transformed using $t^2 = u$ and can then be written in terms of elementary functions. The first integral can be rewritten using partial fractions for $R_4(t^2)$ which gives us two terms: one involving powers of t^{2m} for some $m \in \mathbb{Z}$ and one involving powers of $(t^2 + \gamma)^{-n}$ for some $n \in \mathbb{Z}_+$ and $\gamma \in \mathbb{C}$, giving us:

$$\int R_4(t^2)X^{-1/2}\, dt = \int t^{2m}X^{-1/2}dt + \int (t^2 + \gamma)^{-n}X^{-1/2}\, dt =: I_{2m} + J_n$$

Using certain reduction formulae (Equation 3.3.33 in [24]), we can show that any I_{2m} can be written as a linear combination of I_0 and I_2. The second integral J_m can similarly be studied using certain reduction formulae (Equations 3.3.35 and 3.3.36 in [24]), and it can be shown that any J_m be expressed

If $P(t)$ had two identical/repeated roots α, then we might be able to factor out a term $(x - \alpha)^2$ term in $P(t)$, allowing us to integrate $R(t, \sqrt{P(t)})$ in terms of elementary functions. See Section 3.3 in [24].

in terms of J_1, J_0, and J_{-1}. It turns out that J_0 can be expressed in terms of I_0 and J_{-1} can be expressed as a linear combination of I_0 and I_2. Hence, given a general elliptic integral, we may express it as:

$$\begin{aligned}
\int R(t,y)\,dt &= \int y^{-1}R_1(t)\,dt + \int R_2(t)\,dt \\
&= \int R_3(t)X^{-1/2}(t)\,dt + \int R_2(t)\,dt \\
&= \int R_4(t^2)X^{-1/2}\,dt + \int tR_5(t^2)X^{-1/2}\,dt + \int R_2(t)\,dt \\
&= I_{2m} + J_n + \int tR_5(t^2)X^{-1/2}\,dt + \int R_2(t)\,dt \\
&= (aI_0 + bI_2) + (cJ_1 + dJ_0 + eJ_{-1}) + \int tR_5(t^2)X^{-1/2}\,dt + \int R_2(t)\,dt \\
&= (aI_0 + bI_2) + (cJ_1 + dI_0 + e(fI_0 + gI_2)) + \int tR_5(t^2)X^{-1/2}\,dt + \int R_2(t)\,dt \\
&= c_1I_0 + c_2I_2 + c_3J_1 + \underbrace{\int tR_5(t^2)X^{-1/2}\,dt + \int R_2(t)\,dt}_{\text{Expressible in terms of elementary functions}}
\end{aligned}$$

Hence, to be able to express general elliptic integrals, we only need to introduce three new functions into our basis set of function: I_0, I_2, J_1.

$$I_0 = \int_0^x \frac{dt}{\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}} \tag{8.4}$$

$$I_2 = \int \frac{t^2}{\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}}\,dt \tag{8.5}$$

$$J_1 = \int \frac{dt}{(t^2 - \gamma)\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}} \tag{8.6}$$

This motivates us to introduce respectively the elliptic integrals of the first, second, and third kind, defined as follows.

Definition 8.2. We define the incomplete elliptic integral of the first kind as:

$$\operatorname{arcsn}(x,k) = \operatorname{sn}^{-1}(x,k) = u = F(x,k) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}, \quad 0 \le x \le 1$$

This called **Jacobi's algebraic form**. Using the substitution $t = \sin\theta$ and $x = \sin\phi$, we can obtain **Legendre's trigonometric form**:

$$F(\phi,k) = F(\sin\phi;k) = \operatorname{sn}^{-1}(\sin\phi,k) = \int_0^\phi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}$$

where ϕ is called the amplitude, k the elliptic modulus, and $m = k^2$ the parameter. Thus, the elliptic integral of the first kind can be motivated as the inverse of the Jacobi elliptic function $\operatorname{sn}(x,k)$.

Rather than using the Legendre canonical forms, we could also introduce other basis functions such as Carlson's symmetric forms R_F, R_J, R_G, but these are expressible in terms of the Legendre forms and so are simply a different set of basis vectors/functions. We should also note that some authors, such as [24], actually define elliptic integrals of the second and third kind as (8.5) and (8.6). Looking at the different values of A_1, B_1, A_2, B_2 lead to different examples of elliptic integrals of the respective kinds. However, each of these integrals turns out to be reducible/expressible in terms of the elliptic integrals of the form expressed below in Definition (8.2).

The incomplete elliptic integral of the second kind is given in the Jacobi algebraic and Legendre trigonometric form as (again using $t = \sin\theta, x = \sin\phi$):

$$E(x,k) = \int_0^x \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,dt = \int_0^\phi \sqrt{1-k^2\sin^2\theta}\,d\theta = E(\phi,k) = E(\sin\phi;k)$$

If we let $x = \mathrm{sn}(u)$, we call $\mathcal{E}(u,k)$ Jacobi's epsilon function.

The incomplete elliptic integral of the third kind is defined as:

$$\Pi(n;\phi|m) = \int_0^{\sin\phi} \frac{1}{1-nt^2}\frac{dt}{\sqrt{(1-mt^2)(1-t^2)}}$$

where n is called the characteristic.

We call an incomplete integral is complete when the amplitude is $\phi = \pi/2$, *i.e.,* $x = \sin\phi = 1$. As a result, while incomplete elliptic integrals are functions of two variables, e.g. (x,k) or (ϕ,k), complete elliptic integrals are only functions of one variable k or m. We thus have the following defintions and notational conventions.

Definition 8.3. We define the complete elliptic integral of the first kind is:

$$K(k) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \int_0^1 \frac{dt}{\sqrt{(1-t^2)\,(1-k^2t^2)}} = F(\pi/2,k) = F(1,k)$$

The notation K is justified as $\mathrm{sn}(u = K) = 1$. As a result $\mathrm{sn}^{-1}(x = 1) = K = F(1,k)$. We say this function is complete since the inverses of Jacobi elliptic functions are defined on the range $(0,K)$ due to the Jacobi Elliptic functions ceasing to be bijective in their argument u. Hence, once the inverse function reaches its value K, it is complete.

The complete elliptic integral of the second kind is defined as:

$$E(k) = \int_0^{\frac{\pi}{2}} \sqrt{1-k^2\sin^2\theta}\,d\theta = \int_0^1 \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,dt = E(\pi/2,k) = E(1,k)$$

It is complete in the sense that $E(x = \mathrm{sn}(u,k),k) = \mathcal{E}(u,k) = \int_0^{\mathrm{sn}(u)} \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}dt = \int_0^u \mathrm{dn}^2(w,k)dw$ which evaluating at $u = K$ gives $E(k) = \mathcal{E}(K,k)$.

The complete elliptic integral of the third kind is defined as:

$$\Pi(n,k) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\left(1-n\sin^2\theta\right)\sqrt{1-k^2\sin^2\theta}}.$$

It is complete in the sense that $\Pi(n;\phi = \mathrm{am}(u,k)\,|\,m = k^2) = \int_0^u \frac{dw}{1-n\,\mathrm{sn}^2(w,k)}$, with $u = K$ giving us the complete integral.

We can graph the complete integrals as functions of k as follows.

Other authors use a different canonical form as well, e.g. in [24], the canonical form of the incomplete integral of the third kind is given as Equation (3.7.5):

$$\Pi(u,a,k) = \int_0^u \frac{k^2\,\mathrm{sn}(a)\,\mathrm{cn}(a)\,\mathrm{dn}(a)\,\mathrm{sn}^2(v)}{1-k^2\,\mathrm{sn}^2\,a\,\mathrm{sn}^2\,v}\,dv$$

This canonical form is related to our canonical form, which Lawden denotes $\Lambda(u,\alpha,k)$ via Equation (3.7.11).

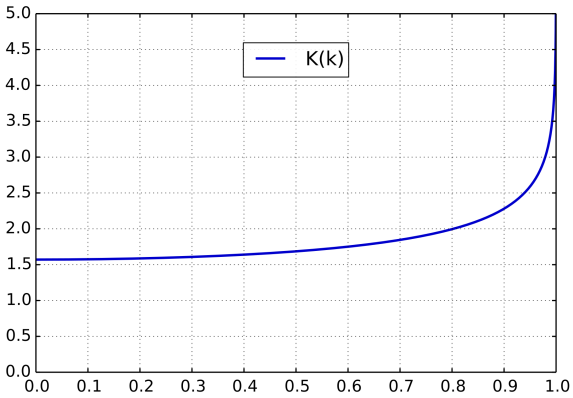

FIGURE 3. Graph of $K(k)$, the complete elliptic integral of first kind [15]

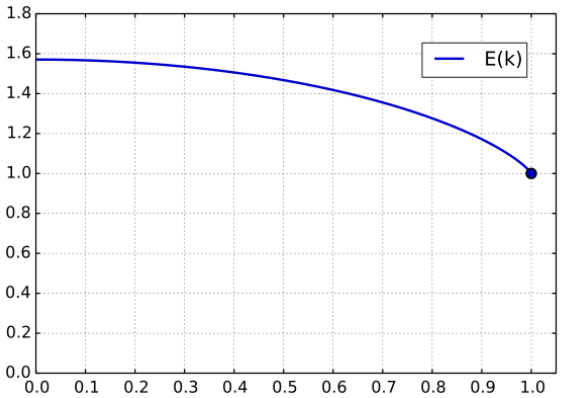

FIGURE 4. Graph of $E(k)$, the complete elliptic integral of second kind [14]

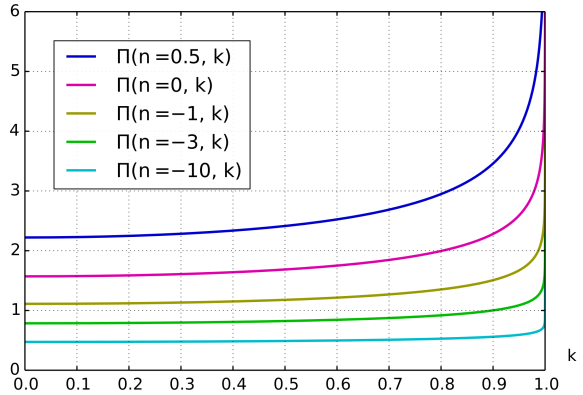

FIGURE 5. Graph of $\Pi(n, k)$, the complete elliptic integral of third kind, for different fixed $n \in \{-10, -3, -1, 0, 0.5\}$ [16]

There are many interesting properties of the incomplete and complete elliptic integrals, such as finding their Taylor series, see Chapter 3 in [24]. However, the last relation we will highlight is that the

complete integrals can be expressed in terms of Gauss's hypergeometric function we will meet shortly as:

$$K(k) = \tfrac{\pi}{2}\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;k^2\right)$$
$$E(k) = \tfrac{\pi}{2}\,{}_2F_1\left(\tfrac{1}{2},-\tfrac{1}{2};1;k^2\right)$$

This will be important as it turns out there are many equivalent ways to express general elliptic integrals in terms of special functions. Some expressions are easier to work with than others depending on the context. Therefore, converting between them is important.

8.3.1. **Inverses of Jacobi elliptic functions.** As already seen above, the inverses of Jacobi elliptic functions can be expressed in terms of elliptic integrals. Recall that the inverse of a function, e.g. $\mathrm{sn}(u,k)$, in the argument u is defined as solving the equation $x = \mathrm{sn}(u,k)$ for u, *i.e.,*:

$$u = \mathrm{arcsn}(x,k) = \mathrm{sn}^{-1}(x,k)$$

We can obtain the elliptic integral expression for this inverse by noticing if $x = \mathrm{sn}(u,k)$, then:

$$\frac{dx}{du} = \frac{d}{du}\,\mathrm{sn}(u,k) = \mathrm{cn}(u,k)\,\mathrm{dn}(u,k) = \sqrt{1-\mathrm{sn}^2(u,k)}\sqrt{1-k^2\,\mathrm{sn}^2(u,k)} = \sqrt{(1-x^2)(1-k^2x^2)}$$

Now integrating, we obtain:

$$\mathrm{sn}^{-1}(x,k) = u = \int du = \int \frac{du}{dx}dx$$

Hence:

$$\boxed{\mathrm{sn}^{-1}(x,k) = \int_0^x \frac{dt}{\sqrt{(1-x^2)(1-k^2x^2)}}, \quad 0 \le x \le 1}$$

Similiarly, we can obtain an expression for $u = \mathrm{cn}^{-1}(x,k)$ by setting $x = \mathrm{cn}(u,k)$ and finding:

$$\frac{dx}{du} = -\,\mathrm{sn}(u)\,\mathrm{dn}(u) = -\sqrt{(1-x^2)(k'^2+k^2x^2)} = -\sqrt{(1-x^2)(1-k^2+k^2x^2)}$$

Integrating over the range $(0,u)$ for u and $(1,x)$ for x gives:

$$\boxed{cn^{-1}(x,k) = \int_x^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-k^2+k^2t^2)}}, \quad 0 \le x \le 1}$$

Lastly, we can also obtain the result that:

$$\boxed{\mathrm{dn}^{-1}(x,k) = \int_x^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(t^2+k^2-1)}}}$$

We can continue to use this method to solve for the inverses of the remaining nine elliptic functions, see Equations (3.2.6)-(3.2.15) in [24]. An important results however, as seen in Section 3.2 in [24], is that all of these integrals are reducible or expressible in terms of $\mathrm{sn}^{-1}(x,k)$, *i.e.,* they are all incomplete elliptic integrals of the first kind.

8.3.2. **Beta function.** There are two **Euler integrals**:

(1) Euler integral of the first kind, called the **Beta function**:

$$\mathrm{B}(z_1,z_2) = \int_0^1 t^{z_1-1}(1-t)^{z_2-1}\,dt = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)}$$

with complex arguments z_1, z_2 such that $\mathrm{Re}(z_1), \mathrm{Re}(z_2) > 0$.

(2) Euler integral of the second kind, called the **Gamma function**:

$$\Gamma(z) = \int_0^\infty t^{z-1}\,\mathrm{e}^{-t}\,dt$$

for complex argument z with $\mathrm{Re}(z) > 0$.

Both of these functions can be generalised to the **incomplete Beta function** and the (lower and upper) incomplete Gamma functions. We will only be interested in the incomplete Beta function, defined as:

$$\boxed{\mathrm{B}_z(a,b) = \mathrm{B}(z;\,a,b) = \int_0^z t^{a-1}\,(1-t)^{b-1}\,dt} \tag{8.7}$$

It is generally assumed that $a, b > 0$ and $0 \leq z \leq 1$, however it is straightforward to analytically continue a, b, z to other values, including complex values. As a result, setting $z = 1$ recovers the usual (complete) Beta function, $B_1(a,b) = B(a,b)$. We can also Taylor expand the incomplete Beta function in z, giving us:

$$B_z(a,b) = z^a \sum_{n=0}^{\infty} \frac{(1-b)_n}{n!(a+n)} z^n \tag{8.8}$$

where $(x)_n$ is the Pochhammer symbol we define in the next subsection.

Importantly, the derivative and indefinite integral are given by (the derivative rule trivially follows from the Leibniz rule):

$$\boxed{\frac{d}{dz} B_z(a,b) = (1-z)^{b-1} z^{a-1}} \tag{8.9}$$

$$\boxed{\int B_z(a,b)\,dz = zB_z(a,b) - B_z(a+1,b)} \tag{8.10}$$

We are now interested in inverting the incomplete Beta function in its argument z, giving us the **inverse incomplete Beta function** $B_z^{-1}(a,b)$. Recall that the inverse $f^{-1}(z)$ of a function $f(z)$ can be equivalently defined as solving $f(f^{-1}(z)) = z$. Hence, the inverse incomplete Beta function $B_z^{-1}(a,b)$ should satisfy:

$$x = B_{B_z^{-1}(a,b)}(a,b) = \int_0^{B_z^{-1}(a,b)} t^{a-1}(1-t)^{b-1}\,dt$$

From here, we can easily find the derivative of the inverse incomplete Beta function using Leibniz's rule.

$$\boxed{\frac{d}{dz} B_z^{-1}(a,b) = [B_z^{-1}(a,b)]^{1-a}[1 - B_z^{-1}(a,b)]^{1-b}} \tag{8.11}$$

This rule also trivially follows from the formula for the derivative of the inverse of a function. What justification do we have for the existence of the inverse incomplete Beta function? Since the function $z \mapsto B_z(a,b)$ is strictly increasing in z on $[0,1]$, we can define its inverse $z(x) := B_x^{-1}(a,b)$ to be the solution to the equation $B_z(a,b) = x$.

We can also derive a formula for the (lower) indefinite integral of the inverse incomplete beta function. Here, "lower" refers to the fact that we are integrating over the 'lower' input x of the Beta function as opposed to the inputs a, b. Letting $I(x) = \int B_x^{-1}(a,b)\,dx$ and setting $z = B_x^{-1}(a,b)$, we know that:

$$x = B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\,dt$$

Applying the Leibniz integral rule, we find:

$$\frac{d}{dz}x = \frac{d}{dz}\int_0^z t^{a-1}(1-t)^{b-1}\,dt = z^{a-1}(1-z)^{b-1}$$

As a result, $dx = z^{a-1}(1-z)^{b-1}\,dz$ and we can now rewrite the integral $I(x)$ of the inverse incomplete Beta function, using the fact that $z = B_x^{-1}(a,b)$, as:

$$\begin{aligned} I(x) &= \int B_x^{-1}(a,b)\,dx = \int z\,dx = \int z\,z^{a-1}(1-z)^{b-1}\,dz = \int z^a(1-z)^{b-1}\,dz \\ &= B_z(a+1,b) + c = B_{B_x^{-1}(a,b)}(a+1,b) + c \end{aligned}$$

Hence, we have the integral formula:

$$\boxed{I(x) = \int B_x^{-1}(a,b)\,dx = B_{B_x^{-1}(a,b)}(a+1,b) + c} \tag{8.12}$$

Lastly, we should note that we can express the incomplete Beta function in terms of the hypergeometric function, defined below, as:

$$B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}dt = \frac{z^a}{a}\,{}_2F_1(a,1-b;a+1;z) \tag{8.13}$$

As a result, we can express the integral (8.12) also as:

$$I(x) = B_{B_x^{-1}(a,b)}(a+1,b) + c = \frac{[B_x^{-1}(a,b)]^{a+1}}{a+1}\,{}_2F_1(a+1,1-b,a+2;B_x^{-1}(a,b))$$

In terms of graphing the incomplete and inverse incomplete Beta function, the behaviour depends highly on the values of a, b. The interested reader may consult the following link to obtain a visual intuition for the graph of $B_x(a,b)$ for different values of a, b. The inverse incomplete Beta function would then simply be reflected across the line $y = x$.

8.3.3. **Gauss's hypergeometric function.** Gauss's hypergeometric function is a special function defined for $|z| < 1$ by the hypergeometric series:

$${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!} = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots$$

where $(q)_n$ is the (rising) Pochhammer symbol:

$$(q)_n = \begin{cases} 1 & n = 0 \\ q(q+1)\cdots(q+n-1) & n > 0 \end{cases}$$

The function ${}_2F_1$ is simply one possible generalised hypergeometric series of the form ${}_pF_q$, where $p = 2$ refers to the fact that we have two separate variables a, b in the numerator while $q = 1$ just refers to having only one variable c in the denominator. We call these series hyper*geometric* since setting $a = 1$ and $b = c$, we recover the geometric series:

$${}_2F_1\left(1,b;b;z\right) = {}_1F_0\left(1;;z\right) = 1 + z + z^2 + z^3 + z^4 + \cdots$$

We should also note that while ${}_2F_1$ is formally only defined for $|z| < 1$, for complex $|z| \geq 1$, we can analytically continue along any path in $\mathbb{C}$, as long as we avoid the branch cuts of ${}_2F_1$ at 1 and infinity.

The derivative of ${}_2F_1$ is given by:

$$\frac{d}{dz}\,{}_2F_1(a,b;c;z) = \frac{ab}{c}\,{}_2F_1(a+1,b+1;c+1;z) \tag{8.14}$$

which generalises to:

$$\frac{d^n}{dz^n}\,{}_2F_1(a,b;c;z) = \frac{(a)_n(b)_n}{(c)_n}\,{}_2F_1(a+n,b+n;c+n;z) \tag{8.15}$$

The hypergeometric function can also be related to other special functions. We have seen above how it relates to the Beta function. For certain values of a, b, c, we can relate ${}_2F_1$ to the complete elliptic integrals $K(k), E(k)$ and $D(k)$ by:

$$K(k) = \tfrac{\pi}{2}\,{}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2\right) \tag{8.16}$$

$$E(k) = \tfrac{\pi}{2}\,{}_2F_1\left(-\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2\right) \tag{8.17}$$

$$D(k) = \tfrac{\pi}{4}\,{}_2F_1\left(\tfrac{1}{2}, \tfrac{3}{2}; 2; k^2\right) \tag{8.18}$$

For more, including graphs of the hypergeometric function, see [24].

9. Beta Function Formalism

Having reviewed all the fundamentals necessary to understand the mathematics and the motivation to study subelliptic operators such as the Grushin operators, let us give an overview of the algorithm or geometric method we will use to study them. In this method, the use of the Beta function will naturally appear, as seen in the subsequent chapters, lending us to call this approach the **Beta Function Formalism**.

Given a partial differential operator $L(x, D) = \sum_{|\alpha|\leq m} a_\alpha(x) D^\alpha$, the first step in our algorithm is to find the principal symbol: $P_m(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha$. We then set the Hamiltonian H associated to the differential operator equal to the principal symbol $H(x, \xi) = P_m(x, \xi)$, interpreting the ξ's as conjugate momenta. We can then set up the Hamiltonian system for the operator by finding for each pair of conjugate momenta $x_i \leftrightarrow \partial_{x_i} \leftrightarrow \xi_{x_i}$ Hamilton's equations $\dot{x}_i = H_{\xi_{x_i}}$ and $\dot{\xi}_{x_i} = -H_{x_i}$. The solutions of this Hamiltonian system give us the normal geodesics.

In the case we will study, the Grushin operator, the solutions of the Hamiltonian system can be found using an Energy Ansatz, meaning the solutions will be expressible in terms of an energy integral: $t = \int_{x_0}^{x_1} \frac{dx}{\sqrt{2E-2U(x)}}$. In the Beta function formalism, we will express this integral in terms of (incomplete) Beta functions $B_z(a, b)$. We then invert this expression to solve for the x-parametrisations of the geodesics. The y-parametrisations can then be found by solving a similar integral which we will express in terms of Beta functions. This will give us the geodesics in Beta function form.

We can then choose two arbitrary points (x_i, y_i) and (x_f, y_f) and count the number of geodesics connecting the two points. This number will be given by the intersections between the x- and y-parametrisations.

The last goal of our method is to find the heat kernel $v(x, t)$ of the given operator L. We will do this by writing down the associated heat equation $\partial_t u = L[u]$ and partial Fourier transforming the equation in one of the variables x to obtain a simpler equation in which one of the derivatives has become an algebraic "momentum" variable. We can then use an exponential Ansatz to solve the associated heat equation, with solutions of the form:

$$u \sim e^{t\mathbb{L}}$$

where $\mathbb{L}$ is partial Fourier transformed operator of L. We thus want to find the fundamental solution K of $\mathbb{L}$, which we can do using the path integral approach and van Vleck's formula. To apply van Vleck's formula, we will need to find the classical action S_{cl} associated to $\mathbb{L}$. We can do this by finding again the geodesics $x_{cl}(t)$ of $\mathbb{L}$ and expressing them in terms of (incomplete) Beta functions. The usual approach would then be to find the Lagrangian $\mathcal{L}$ associated to the Hamiltonian H of $\mathbb{L}$ by inverse Legendre transforming. Then, integrating $\mathcal{L}$ evaluated along $x_{cl}(t)$ over t gives the classical action S_{cl}, which we can plug into van Vleck's formula:

$$K \sim \sqrt{-\frac{1}{2\pi}\frac{\partial^2}{\partial x_0 \partial x} S_{cl}}\; e^{-S_{cl}}$$

From here, we can then inverse partial Fourier transform our exponential Ansatz to obtain the heat kernel v of our given operator L:

$$v \sim \frac{1}{2\pi} \int e^{iy\xi} K d\xi$$

where y is the variable in which we partial Fourier transformed at the beginning.

Statements and Declarations

The authors declare that they have no conflict of interest, and the manuscript has no associated data.

Acknowledgments

The first author is partially supported by NSF grant DMS-1408839 and a McDevitt Endowment Fund at Georgetown University and the second author is grateful for the financial support from the Goldwater foundation and ARCS foundation during his study at Georgetown University.

References

[1] V. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, New York, 1975.
[2] C. J. Bruijnen. Bracket-Generating Distributions and Filtered Structures, 2020. https://studenttheses.uu.nl/bitstream/handle/ 20.500.12932/37131/Master.
[3] O. Calin and D.-C. Chang. *Geometric Mechanics on Riemannian Manifolds: Applications to Partial Differential Equations*. Birkhäuser, Boston, MA, 2005.
[4] O. Calin and D.-C. Chang. *Sub-Riemannian Geometry: General Theory and Examples*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2009.
[5] O. Calin, D.-C. Chang, K. Furutani, and C. Iwasaki. *Heat Kernels for Elliptic and Sub-elliptic Operators: Methods and Techniques*. Birkhäuser, Boston, MA, 2011.
[6] D.-C. Chang and Y. Li. Asymptotics for heat kernels of subLaplacian on Heisenberg sub-Laplacian and the Grushin operator. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471:Article ID 20140943, 2015.
[7] D.-C. Chang and J. Tie. *The SubLaplacian Operators of Some Model Domains*. Walter de Gruyter GmbH, Berlin, Boston, 2022.
[8] C.-H. Chang, D.-C. Chang, B. Gaveau, and P. Geiner. Geometric Analysis on a step 2 Grushin Operator. *Bulletin of the Institute of Mathematics, Academia Sinica, New Series*, 4:119–188, 2009.
[9] W.-L. Chow. Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung. *Mathematische Annalen*, 117:98–105, 1940.
[10] M. Christ. Hypoellipticity in the infinitely degenerate regime. In J. D. McNeal, editor, *Complex Analysis and Geometry: Proceedings of a Conference at The Ohio State University*, pages 59-84, June 3–6, USA, 1999. De Gruyter, Berlin, Boston, 2001.
[11] M. Christ. A Rremark on sums of squares of complex vector fields. *ArXiv*:0503506, 2005.
[12] C. Fefferman and D.H. Phong. Subelliptic eigenvalue problems. *Conference on harmonic analysis in honor of Antoni Zygmund*, Pages 590–606, Chicago, 1981. Wadsworth International Group, Belmont, California, 1983.
[13] The fundamental rectangle in the complex plane of u, Wikipedia, the free encyclopedia, 2006. https://en.wikipedia.org/wiki/File:JacobiFunctionAbstract.png.
[14] Plot of the complete elliptic integral of second kind $E(k)$, Wikipedia, the free encyclopedia, 2015. https://commons.wikimedia.org/wiki/File: Mplwp-complete-ellipticEk-svg.
[15] Plot of the complete elliptic integral of first kind $K(k)$, Wikipedia, the free encyclopedia, 2015. https://commons.wikimedia.org/wiki/File: Mplwp-complete-ellipticKk-svg.
[16] Plot of the complete elliptic integral of third kind $\Pi(n, k)$, Wikipedia, the free encyclopedia, 2015. https://commons.wikimedia.org/wiki/File: Mplwp-complete-ellipticPi-nfixed-k.svg.
[17] S. Donaldson. Elliptic operators: lecture notes: London School of Geometry and Number Theory course, 2020. https://www.ma.imperial.ac.uk/skdona/elliptic.
[18] L. Hörmander. Subelliptic operators, Seminar on singularities of solutions of linear partial differential equations. *Princeton University Press*, 91:127–208, 1979.
[19] L. Hörmander: *The Analysis of Linear Partial Differential Operators I*. Springer-Verlag, Heidelberg, 1983.
[20] L. Hörmander. *The Analysis of Linear Partial Differential Operators II*. Springer-Verlag, Heidelberg, 2003.

[21] C. Jacobi. *Fundamenta Nova Theoriae Functionum Ellipticarum*. Reprinted by Cambridge University Press in 2012, 1829.
[22] J. J. Kohn. Hypoellipticity and loss of derivatives. *Annals of Mathematics*, 163:943–986, 2005.
[23] J. J. Kohn and L. Nirenberg. Non-coercive boundary value problems. *Communication on Pure and Applied Mathematics*, 18:443–492, 1965.
[24] D. Lawden. *Elliptic Functions and Applications*. Springer-Verlag, New York, 1969.
[25] E. Le Donne. *Lecture Notes on Sub-Riemannian Geometry*, 2011. http://www.math.ethz.ch/ ledonnee/sub-Riem-notes.pdf.
[26] W. Liu and H. Sussmann. Shortest paths for sub-Riemannian metrics on rank-two distributions. *Memoirs of the American Mathematical Society*, 118:1–104, 1995.
[27] R. Littlejohn. The Propagator and the Path Integral, 2021. https://bohr.physics.berkeley.edu/classes/221/notes/pathint.pdf.
[28] F. Maiale and A. Tarsia. *Elliptic Equations*. Lecture Notes, Department of Mathematics, Pisa University, 2018. https://poisson.phc.dm.unipi.it/ fpmaiale/notes/EE.pdf.
[29] J. Rauch. *Partial Differential Equations*. Springer-Verlag, New York, 1991.
[30] B. Street. What is . . . Hypoellipticity? *Notices of the American Mathematical Society*, 65:418–423, 2018.
[31] B. Street. What else about . . . Hypoellipticity? *Notices of the American Mathematical Society*, 65:424–425, 2018.
[32] D. Tartakoff: *Nonelliptic Partial Differential Equations*. Springer-Verlag, New York, 2011.
[33] Z. Yang. *What is Hypoellipticity?* Mathematisches Institut, University of Gottingen, Germany, 2020. https://www.yzpmath.com/post/2020-2/2020-2.pdf.