BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 47, Number 4, October 2010, Pages 713–722
S 0273-0979(10)01288-7
Article electronically published on February 8, 2010

Sub-Riemannian geometry: general theory and examples, by Ovidiu Calin and Der-Chen Chang, Encyclopedia of Mathematics and Its Applications, vol. 126, Cambridge University Press, New York, 2009, xi+367 pp., hardcover, US$99.00, ISBN 978-0521897303

Take an n-dimensional manifold M. Endow it with a distribution, by which I mean a smooth linear subbundle $D \subset TM$ of its tangent bundle TM. So, for $x \in M$, we have a k-plane $D_x \subset T_x M$, and by letting x vary we obtain a smoothly varying family of k-planes on M. Put a smoothly varying family g of inner products on each k-plane. The data (M, D, g) is, by definition, a sub-Riemannian geometry. Take the viewpoint that we can explore M only by traveling along paths tangent to D. Call such paths *horizontal.* Since g measures lengths of horizontal (D) vectors, we can measure lengths of horizontal paths and formulate the sub-Riemannian geodesic problem: to find the shortest horizontal path connecting two given points.

Gromov and his school use the term *Carnot-Carathéodory geometry* for what we call sub-Riemannian geometry. The sub-Riemannian analogues of Euclidean space are a class of Lie groups endowed with left-invariant sub-Riemannian metrics, christened *Carnot groups* by the Gromov school, *homogeneous groups* by Stein and his school, and their Lie algebras named *symbol algebras* by Tanaka's school.

The first nontrivial example is the 3-*dimensional Heisenberg group* and is the simplest non-Euclidean Carnot group. Its sub-Riemannian geodesic problem is essentially the isoperimetric problem. Take M to be standard 3-space $\mathbb{R}^3$. Define D by the vanishing of the 1-form $\theta = dz - (1/2)(xdy - ydx)$; in other words $D_{(x,y,z)}$ is the 2-plane $\{(v_1, v_2, v_3) : v_3 - (1/2)(xv_2 - yv_1) = 0\}$. D can be visualized as a kind of continuous spiral staircase. See the accompanying Figure 1, inspired by the description on p. 40 of [1]. Along the z-axis, the 2-planes of D are parallel to the xy-plane. As we move out radially from the axis along any orthogonal ray, the 2-plane spins about the axis of the ray, rotating monotonically so that by the time we "reach" infinity, they have just rotated by 90 degrees, turning vertical. We have just described the standard contact structure on 3-space. The projection along the z-axis maps each plane linearly onto the xy-plane. Declaring these restricted projections to be isometries defines the family g of inner products on D. Consider a horizontal path γ connecting the origin to a point A units up along the z-axis. The projection c of γ to the xy-plane is closed, and by definition of g the sub-Riemannian length of γ equals the standard Euclidean length of c. An application of Stokes theorem to the equation $A = (1/2)\int_c (xdy - ydx)$ shows that the (signed) area inside c equals the height A. Thus, the sub-Riemannian geodesic problem with endpoint constraints $\gamma(0) = (0,0,0)$, $\gamma(1) = (0,0,A)$ is equivalent to the dual of the isoperimetric problem: among all closed curves c in the plane enclosing a fixed area A, find the one of minimum length.

Gauge theory and falling cats. My introduction to sub-Riemannian geometry started with a science fiction story [12], or at least what sounded like science fiction at the time. Jair Koiller (now working in Rio de Janiero) had told me how

2000 *Mathematics Subject Classification.* Primary 53C17; Secondary 37J99, 53C60, 58E10, 70H05, 35H10, 53C20, 22E40.

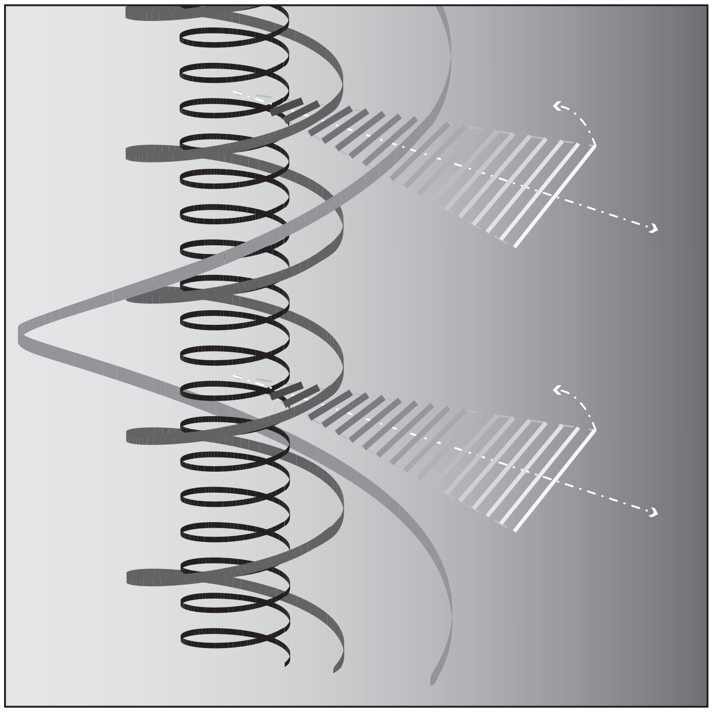

FIGURE 1. The Heisenberg group and a few of its geodesics.

physicists Shapere and Wilczek were using gauge theory to understand the physical mechanism behind how some microorganisms swim. The story was true. I learned the basics of their paper and became engrossed with a simplified model of their idea, the problem of how a cat, dropped upside down, rights itself in midair so as to land on its feet.

A frozen cat will drop like a stone and lands on its back. If the living cat is to land on its feet, it must change its shape in midair. To understand how this works, first ask, What is a shape? A configuration of the cat is to be viewed as an actual cat, floating in space. When we choose a model of the cat, such as N-point masses

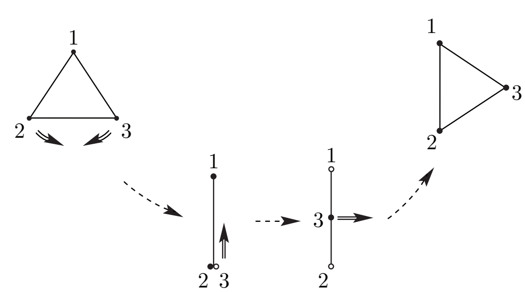

FIGURE 2. A cat consisting of three mass points rotates 60 degrees, its angular momentum zero at every instant.

or a continuum model, the set of all configurations becomes a manifold M. Two configurations of the cat represent "the same shape" if there exists an orientation-preserving isometry of space taking one to the other. Thus, the space of shapes is the quotient space M/G, where G is the group of orientation-preserving isometries of space. To rotate and land on its feet, the cat would love to just move purely in the G-direction. But it cannot, as the frozen cat illustrates. The reason such motions are impossible is conservation of its angular momentum J. The cat's initial angular momentum J is zero, and it remains constant throughout freefall. J is a vector-valued function on TM which is linear in the fiber variables—the velocities. Thus the assertion $J = 0$ defines a distribution D on M. The angular momentum zero constraint forces the cat to move horizontally. The cat's mass distribution induces a physically natural Riemannian metric on M, the *kinetic energy metric*, and relative to this metric, D is orthogonal to the G-orbits: to rotate, the cat must move orthogonally to rotations. Figure 2, modified from Figure 8 of Chenciner [4], is a cartoon of how a cat consisting of three equal mass points might rotate by 60 degrees. If we restrict this metric (or any other) to D we get a g as above, and the cat's problem has become a case of the sub-Riemannian geodesic problem.

I will divide sub-Riemannian geometry into three streams of research, according to the methods and motivations used. The first stream arises out of optimal control and surrounds the sub-Riemannian geodesic problem. I am most familiar with this stream. The second stream is partial differential equations (PDEs), exemplified by a theorem of Hormander. The third stream concerns sub-Riemannian geometries arising out of limits of Riemannian or other geometries.

Optimal control. The deepest results on the sub-Riemannian geodesic problem have been obtained by mathematicians trained in optimal control. The basic problem in nonlinear control theory is to construct paths $\gamma(t)$ connecting two points q_0, q_1 of a manifold M, subject to path constraints of the form $\dot\gamma(t) \in V(\gamma(t))$, where $V(x) \subset T_xM$ are subsets varying with x. The sets $V(x)$ are represented as images of a map $u \mapsto f(x;u)$ where the "control" u varies over some auxiliary set, usually a subset of a Euclidean space. In applications u is fuel, steering angle, temperature, etc. To relate this setup to sub-Riemannian geometry, set $V(x) = D_x$. If we throw in a function $L(x,u)$ and try to minimize $\int L(\gamma(t), u(t))dt$ over the space of all controls $u(t)$ which "steer" from q_0 to q_1 (thus $\dot\gamma = f(\gamma, u(t))$, $\gamma(0) = q_0$, $\gamma(1) = q_1$), we have formulated the basic problem in optimal control. By taking the $V(x) = D_x$ and expressing L using g, we have represented the sub-Riemannian geodesic problem as a problem in optimal control.

The first question that arises when facing the sub-Riemannian geodesic problem is *accessibility*: Is there any horizontal path at all which connects the two given points? If D is involutive—meaning Lie brackets of horizontal vector fields remain horizontal—then, by the Frobenius integrability theorem, the answer is no for generic pairs of points. The Frobenius theorem asserts that the set of points connected to q_0 by a horizontal path form an immersed submanifold of dimension k $(= \text{rank}(D))$ called the *leaf* of D through q_0. A typical q_1 will not lie on this leaf and so cannot be horizontally connected to q_0. The first big theorem in the subject, the Chow-Rashevskii theorem [5], [11], asserts that for D at the opposite extreme of involutive, the answer is yes, locally. The technical term for *at the opposite extreme of involutive* is *bracket generating*. A distribution D is called bracket generating if its horizontal vector fields, $X_1, X_2, \ldots$, together with their iterated Lie brackets, $[X_i, X_j], [X_i, [X_j, X_k]], \ldots$, span the tangent bundle to M.

Subelliptic PDEs. Chow-Rashevskii's bracket generating condition reappeared 40 odd years later in Hörmander [8]. Form a "sub-Laplacian" $\Delta = \sum_{i=1}^{k} X_i^2$, where the X_i are a (local) orthonormal frame for D. Does Δ "act" like the usual Laplacian, meaning are there close analogues of the usual elliptic regularity theorems? Hörmander's theorem asserts that the answer is yes, provided that D is bracket generating. In honor of his theorem, the bracket generating condition is often referred to as *Hormander's condition* in the PDE literature.

Most problems in Riemannian geometric analysis involve, in one way or another, the Riemannian Laplacian. So it is not a surprise that sub-Riemannian Laplacians are central to sub-Riemannian geometric analysis. Over the last ten years there has been enormous work on the sub-Riemannian versions of the minimal surface equations, constant mean curvature equations, Yamabe's problem, etc. Most of the analysis has been on Carnot groups.

Carnot and Carathéodory. Carnot, working on heat engines, realized that certain thermodynamic states are inaccessible from other states. At the urging of his friend the physicist Max Born, Carathéodory used a local version of inaccessibilty to derive the second law of thermodynamics, and of entropy. What he established was a kind of converse to Chow-Rashevskii for co-rank 1 ($k = n - 1$) distributions D. Carathéodory showed that if every point in such an (M, D) has a nearby inaccessible point, then D is involutive. See pp. 38–42 of [1] for a beautiful rendition of this story.

Nilpotentization and Carnot groups. The infinitesimal model of a Riemannian manifold is Euclidean space. The infinitesimal model of a sub-Riemannian geometry (M, D, g) is a Carnot group: a simply connected Lie group G whose underlying Lie algebra $\mathfrak{g}$ is graded nilpotent, i.e., $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \cdots \oplus \mathfrak{g}_r$ where $[\mathfrak{g}_i, \mathfrak{g}_j] \subset \mathfrak{g}_{i+j}$ and $\mathfrak{g}_s = 0$ for $s > r$. To construct such a $\mathfrak{g}$ from the infinitesimal neighborhood of a point p of a manifold with distribution D, set $\mathfrak{g}_1(p) = D_p$. Take $\mathfrak{g}_k(p)$ to be the vector space of T_pM spanned by all k-fold iterated Lie brackets of horizontal vector fields, evaluated at p, modulo the subspace of $(k-1)$-fold iterated brackets of such vector fields. We call p a *regular point* if the dimensions of the $\mathfrak{g}_k(q)$ are constant for q near p. In this case, the Lie bracket of vector fields induces a Lie algebra structure on $\mathfrak{g}(p) = \mathfrak{g}_1(p) \oplus \cdots \oplus \mathfrak{g}_r(p)$. (An open dense subset of points of M are regular.) The bracket generating condition on D implies that $\mathfrak{g}_1(p)$ Lie generates $\mathfrak{g}(p)$. We call $\mathfrak{g}$ the *nilpotentization* of the distribution at p. Tanaka's school [14] calls this Lie algebra the *symbol algebra* of D. The nilpotentization of the Heisenberg group discussed in the beginning is the standard 3-dimensional Heisenberg algebra.

Since $\mathfrak{g}$ is nilpotent, the exponential map yields a diffeomorphism from $\mathfrak{g}$ to G. Translate $\mathfrak{g}_1$ about G to obtain a left-invariant distribution on G. Now throw in the metric g_p on $D_p = \mathfrak{g}_1$. We get a sub-Riemannian structure on G for which left translation acts by isometry and whose nilpotenization is again $\mathfrak{g}$. Various theorems assert quantitatively how the Carnot group G approximates the sub-Riemannian manifold M near p. Unlike the Riemannian case, this approximation is generally *not* Lipshitz.

There is a striking alternative definition of a Carnot group, which we will have cause to use later. Scalar multiplication on a Euclidean space E defines a 1-parameter family of metric dilations, $x \mapsto tx = \delta_t(x)$, which are also group automorphisms when E is viewed as an Abelian group. A Carnot group is a Lie group endowed with a left-invariant metric which admits a 1-parameter family δ_t of metric dilations which are also group automorphims. On the infinitesimal level, $\delta_t(v) = (tv_1 \oplus t^2 v_2 \oplus \cdots \oplus t^r v_r)$ for $v = (v_1 \oplus v_2 \oplus \cdots \oplus v_r) \in \mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \cdots \oplus \mathfrak{g}_r$.

Metric limits. The most surprising applications of sub-Riemannian ideas can be found in Gromov's strange and beautiful book-length paper "Groups of Polynomial Growth and Expanding Maps" [7]. Given a discrete group Γ (preferably infinite) with a finite set of generators, one forms its *Cayley graph* by declaring two elements of the group to be connected by an edge if multiplying one element (on the right, say) by a generator yields the other. Setting all edge lengths to 1 yields a metric on Γ. The group is said to have polynomial growth if the *volume* (number of elements) of balls of radius R is bounded by a polynomial in R, as $R \to \infty$. A group is called *nilpotent* if all sufficiently long iterated commutators are the identity. Bass, Milnor, and others had proved that discrete nilpotent groups have polynomial growth, as did their close cousins, the *virtually nilpotent*. Gromov proved the converse: any group Γ of polynomial growth is virtually nilpotent. Here is how he did it. He shrunk all the edges of the graph to length ϵ, thus getting scaled metric spaces Γ_ϵ on the same underlying group. He constructed a metric—now known as the Gromov-Hausdorff metric—on the space of pointed metric spaces and showed that as $\epsilon \to 0$ his Γ_ϵ converged to *something* in this space of all pointed metric spaces. Then he proved that this something was a Carnot group G endowed with a sub-Riemannian geometry (more accurately, sub-Finsler). The distribution on G corresponds to the choice of generators. The original group Γ (or a finite index subgroup thereof) is embedded in G as a lattice. Lattices in Carnot groups are nilpotent, finishing the proof. (For the experts, apologies for playing a bit fast and loose with the full truth.) More on Carnot groups momentarily.

Gromov was inspired by the appearance of sub-Riemannian geometries in proofs of Mostow rigidity theorems [10]. Mostow's original rigidity theorem asserts that in dimensions $n \geq 3$ if two compact hyperbolic n-manifolds have isomorphic fundamental groups, then they are diffeomorphic. By Hadamard's theorem, such groups form lattices within the standard simply connected hyperbolic n-space, and Mostow's proof involves studying the asymptotics of these lattices. Replacing hyperbolic space by one of the other noncompact symmetric spaces, one can formulate analogous rigidity theorems. The simplest noncompact symmetric space after hyperbolic space is complex hyperbolic space, a complete Riemannian metric space whose curvatures are bounded between $-1/4$ and -1. The boundary at infinity for complex hyperbolic space is an odd-dimensional sphere which inherits, by a Riemannian limiting procedure, a contact distribution endowed with a conformal sub-Riemannian structure. Isometries of complex hyperbolic space are in one-to-one correspondence with conformal sub-Riemannian isometries of the boundary. Rigidity problems for lattices in complex hyperbolic spaces can be rephrased as delicate regularity problems concerning quasi-conformal maps of the boundary. A number of deep works in sub-Riemannian analysis and geometry have been inspired by these connections. Most of this work, as with most of the analysis in sub-Riemannian PDEs, has concentrated on the case of Carnot groups.

In each of the research streams just described, there arise phenomena with no Riemannian counterparts. Associated with these phenomena are open problems. I discuss two.

Geodesics that do not satisfy the geodesic equations. All the geodesics for a Riemannian metric on M arise out of a single Hamiltonian system: the geodesics are the projections to M of solutions to a system of ODEs on the cotangent bundle T^*M. In sub-Riemannian geometry, there is a similar Hamiltonian system on T^*M whose solutions, projected to M, are sub-Riemannian geodesics. For some time it had been believed that all sub-Riemannian geodesics arise in this way. Alleged theorems, published in the literature, claimed that "all sub-Riemannian geodesics satisfy the geodesic problem".

I came up with a counterexample to these alleged theorems in the case $(k, n) = (2, 3)$. The distribution D for my example is a *Martinet distribution*—a generic degeneration of a rank 2 contact distribution. Liu and Sussmann generalized my counterexample to the case of $(2, n)$, for general n. The resulting *singular minimizers*—geodesics that do not satisfy the geodesic equations—are stable: they cannot be perturbed away by perturbing the sub-Riemannian structure.

The key error in purported proofs of these alleged theorems was a misunderstanding of the fundamental result of optimal control theory, the maximum principle of Boltjanski-Gramkrelidze-Pontrjagin. (See the book by L.C. Young [15] for clear expositions around this principle.) The maximum principle is a fancy version of the method of Lagrange multipliers. Recall this method for the problem of minimizing a function F subject to a constraint $G = \text{const}$. Write $\mathbf{x}$ for the vector variable of F and G. Introduce the multiplier λ. Solve the pair of equations $dF(\mathbf{x}) + \lambda dG(\mathbf{x}) = 0$, $G(\mathbf{x}) = \text{const}$ for $\mathbf{x}$. Sort through the resulting solutions by evaluating F to find the minimizer. If the minimizer is not a singular point of the constraint surface $G = \text{const}$, then the method will find it. But if the constraint surface has singular points, one of these may be the minimizer, and the method will most likely miss it. In the sub-Riemannian application, the variable $\mathbf{x}$ is a horizontal path, varying over the space (a Banach manifold) of paths issuing from the initial point $q_0 = \gamma(0)$. F is the length of the path. The constraint map G is the endpoint map, $G(\gamma) = \gamma(1)$, so that $G = \text{const}$ imposes the final endpoint condition $\gamma(1) = q_1$. The minimizers caught by the method are the solutions to the sub-Riemannian geodesic equations as described by Hamilton's equations. The ones missed—minimizers that are singular points of the endpoint map—are the *singular* or *abnormal extremals* appearing in the maximum principle. These curves embody phenomena not present in the Riemannian case where the corresponding endpoint map is always regular.

How rare are singular extremals, the singular curves for the endpoint map? Sard's theorem, fundamental in differential topology, asserts that the set of critcal *values* of a smooth mapping between finite-dimensional manifolds has measure zero. The endpoint mapping G for a distribution is a map between an infinite-dimensional path space (a Banach manifold) and a finite-dimensional manifold, so we cannot just quote Sard. Does an analogous Sard theorem hold for the endpoint map? In other words, how "rare" are endpoints of singular extremals? This basic question in the theory remains open.

The dimension is bigger than the dimension. The notions of Hausdorff dimension and measure make sense for a general metric space. The Hausdorff dimension of a Riemannian n-manifold is n and its Hausdorff measure is the Riemannian volume form, up to a constant scale factor.

Recall that a point p of a sub-Riemannian manifold is called *regular* if the dimensions of the vector spaces $\mathfrak{g}_i$ used in constructing the nilpotentization are constant near p. John Mitchell proved that in a neighborhood of a regular point the Hausdorff dimension N of the sub-Riemannian manifold is given by $N = \Sigma i \dim(\mathfrak{g}_i)$. Since $n = \Sigma \dim(\mathfrak{g}_i)$, we have $N > n$: the metric dimension is larger than the manifold dimension. For example, 4 is the Hausdorff dimension of the 3-dimensional Heisenberg group.

For the case of a Carnot group we have a simple explanation for this dimensional discrepancy. The Lebesgue measure dvol on an n-dimensional Euclidean vector space scales to λ^ndvol under metric dilation by a factor λ. A Carnot group of topological dimension n is diffeomorphic to n-dimensional Euclidean space, and under the diffeomorphism dvol becomes the Haar measure on G which is also its Hausdorff measure (up to scale). But under the Carnot metric dilation δ_λ by a factor λ, as described above, we compute that dvol $\mapsto \lambda^N$dvol with N as above. Thus, the Hausdorff dimension N of G is the usual scaling dimension of G, viewed as a metric space, as one finds it discussed in fractal textbooks.

In analysis and geometry, a basic and very useful fact is the Euclidean isoperimetric inequality: there is a universal constant C such that for any bounded domain Ω in the n-dimensional Euclidean space with rectifiable boundary $\partial\Omega$, we have $\text{area}(\partial\Omega)^n \leq \text{Cvol}(\Omega)^{n-1}$. Moreover, the optimal (smallest) constant is found by taking Ω to be the round ball and replacing the inequality by equality. The best constant is directly related to best constants in Sobolev inequalities.

Pansu uncovered the analogous assertion for the Heisenberg group. To understand it, recall that the central fact about the distribution on the Heisenberg group is its nonintegrability: there are no surfaces everywhere tangent to the distribution. Related to this fact is the fact that the Hausdorff dimension of any surface in the Heisenberg group is 3 rather than 2. One has a natural notion of the sub-Riemannian *volume* of such a surface S, which I will still write as area(S), although it is not a constant multiple of the Euclidean area. Pansu's Heisenberg isoperimetric inequality asserts that for domains in the Heisenberg group with sub-Riemannian rectifiable boundaries, we have $\text{area}(\partial\Omega)^4 \leq \text{Cvol}(\Omega)^3$. It is an open problem to find the best constant C in Pansu's inequality, which is to say, the Heisenberg isoperimetric replacement for the ball. There is a candidate for this replacement—the bubble sets as described in [2].

The zoo of distributions. Forget the fiber inner product g on D. The local study of distributions D is extremely rich, perhaps too rich. This fact came as a huge surprise when I started working in the field.

I was brought up only knowing about contact distributions and involutive distributions. A contact distribution is one with $k = n-1$, n even, and whose nilpotentization is the Heisenberg algebra on k generators. (We continue using k for rank(D) and $n = \dim(M)$.) The Darboux theorem asserts that every contact distribution in dimension n is locally diffeomorphic to its nilpotentization. All contact distributions are locally isomorphic. There is no local theory. This fact is essential for the burgeoning field of contact topology. Similarly, the Frobenius theorem asserts that all involutive distributions (with fixed (k,n)) are locally diffeomorphic, allowing for the existence of the field of the topology of foliations on manifolds. For both contact and involutive distributions, the local symmetry algebra of the distribution D, meaning the Lie algebra of vector fields whose (local) flows map D to itself, is infinite dimensional.

Now the contact condition is stable: if we perturb a contact distribution by a small amount, then the resulting distribution remains contact, and hence by Darboux, locally diffeomorphic to the original. On the other hand, if (k, n) are in the range $1 < k < n - 1$, $(k, n) \neq (2, 4)$, then we can perturb the distribution by an arbitrarily small amount so as to obtain a new distribution not locally diffeomorphic to the original. The first occuring case is $(k, n) = (2, 5)$ and is the subject of a famously difficult paper by Cartan [3], his "5 variables paper". A generic $(2, 5)$ distribution has nilpotentization (at a generic point) equal to the free graded three-step Lie algebra $\mathfrak{g}$ on two generators X, Y. Thus $\mathfrak{g}$ has a basis X, Y, Z, U, V for which the nonzero bracket relations are $[X, Y] = Z, [X, Z] = U, [Y, Z] = V$. Cartan established the existence of curvature-type tensor invariants for such D's, invariants which establish the existence of moduli of inequivalent distributions and provide quantitative means for establishing when two such distributions can be locally diffeomorphic. These Cartan *curvatures* vanish if and only if D is locally diffeomorphic to its nilpotentization. The local symmetry algebra of such a *flat* D is the first occuring exceptional simple Lie algebra: the 14-dimensional Lie algebra $\mathfrak{g}_2$ (in its split real form). If the distribution's curvatures are nonzero, then the symmetry algebra has dimension less than 14, so in particular the symmetry algebra of such a distribution is *never* infinite dimensional. (Generically, the symmetry algebra is trivial: there are no symmetries.)

We have just described the intricacies of the first occuring *nonstable* cases of (k, n). As we increase k, n and $n - k$, the situation gets much more complicated. Quickly moduli of inequivalent nilpotentizations arise. Even if we fix the nilpotentization type, Cartan's curvature invariants become prohibitive even to write down. The situation appears beyond hope of systematic classification. The huge zoo of distributions in the range $1 \ll k < n-1$ has stymied the subject. Researchers tend to hide out in the contact case or the Carnot case. Only rarely do we go into situations in which we have to face the diversity of the zoo of inequivalent distributions. Exceptions include detailed work on the $(2, n)$ case by Bryant and Hsu, Liu and Sussmann, Zelenko, Agrachev, and others.

The book. I'll begin with the good news. The book contains a wealth of worked examples of sub-Riemannian metrics, geodesics, solutions to the Hamilton-Jacobi equations and to the sub-Laplacian. Some appear potentially useful to students. They describe and compute Martinet geodesics in terms of elliptic functions, first uncovered by Kupka, in detail. I learned in Chapter 3 a new proof of the Chow-Rashevskii theorem, based on a theorem they attribute to one "Teleman" whom I did not know. Their section on the motivational thermodynamical work of Carathéodory is also worth spending some time on, although I prefer Born's treatment [1], referenced earlier. Toward the end there is a chapter not found in other books on sub-Riemannian geometry. This chapter concerns Grushin metrics on $\mathbb{R}^n$. Grushin metrics are not sub-Riemannian geometries in our sense, because the *distribution* drops rank at the origin, but they are an interesting class of geometric structures sharing many properties of sub-Riemannian geometries which are susceptible to explicit computation. There is also an interesting appendix on nonsolvable linear subelliptic PDEs. I enjoyed puzzling my way through the example in Chapter 3 which is that of the rank 2 distribution defined by the Pfaffian system $dx - xydz = 0$ on $\mathbb{R}^3$. Any distribution decomposes its space up into accessible sets—the horizontally connected components. In this example those sets are two half-spaces and the plane $x = 0$ dividing them. The distribution necessarily fails to be bracket generating along this plane.

Now for the bad news. This text contains at least two serious errors, errors that would confound probably anyone trying to learn the subject. Theorem 4.2.2, on p. 88, the book's first theorem on sub-Riemannian geodesics, falsely asserts that if we are handed a solution S to the Hamilton-Jacobi equation, then a curve is a geodesic γ if and only if it satisfies $d\gamma/dt = \nabla S(\gamma)$. (The ∇ is the horizontal gradient, defined by $dS_x(v) = g(\nabla S(x), v)$ for any horizontal vector v at x. The Hamilton-Jacobi equation is the equation $\frac{\partial S}{\partial t} = H(dS(x;t))$, where H is the sub-Riemannian kinetic energy, above.) It is difficult to list the number of ways this *theorem* is wrong. To give one, since S is fixed in the theorem, if the *theorem* were true it would imply that all geodesics passing through a point at a fixed time must point in the same direction: ∇S. (I have posted some other ways the theorem is wrong, and a version of how the theorem can be salvaged, on my web page.)

The other serious error I stumbled across is equation 7.1.2 on p. 138. That equation asserts that a certain differential expression which depends upon the choice of an orthonormal frame for D, is in fact independent of that choice. If true, the expression would yield a well-defined connection associated to any sub-Riemannian geometry, a fact unknown to previous researchers, and one which would be central to the theory: a great result if only it were true! The error occurs in the first line of the proof, also the first line on p. 139. It is an error most students of differential geometry have made when learning index notation, especially of the Einstein variety: there is a double sum, but the authors use the same dummy index, here "ℓ", for both, inadvertently setting the two summation indices equal. The error is tantamount to insisting that the $(dg)g^{-1}$ term arising in the transformation formula for connections under changes of gauge (local section) is never present, i.e., always zero. For the authors, g is a function with values in the orthogonal group $O(k)$. They do successfully prove just before the offending line, the well-known fact that the diagonal entries of $(dg)g^{-1}$ are zero, but they miss the essential presence of the off-diagonal entries in their alleged derivation. This error destroys much of the sense of their Chapter 7.

By this point, I lost patience. I suspect that the authors have worked out carefully and explicitly many interesting and wonderful low-dimensional examples. As a reference text, it is infuriatingly inconsistent. I urge prospective readers to go carefully through results before using them. The book also lacks from leaving out all post-Carathéodory motivation for bothering with the subject. Gromov's and Mostow's application of sub-Riemannian ideas are not touched upon. Cartan's "zoo" is not mentioned.

Other books. I know of three books in the field. Gromov wrote [6], a book contained within a book. Acres of research can probably still be mined out of his ideas. I wrote [9] which focuses on sub-Riemannian geodesics and outgrowths of the falling cat problem. Finally, the multiauthored book [2] focuses on geometric analysis on the Heisenberg group, with a grand finale of various proofs of Pansu's Heisenberg isoperimetric inequality. It is well written in a detailed yet friendly manner. As an added bonus it explains a proposed application of Heisenberg geometry toward understanding the workings of a certain part of the visual cortex. I recommend any of these books over the one being reviewed. For the reader who wants just a taste more of the subject without diving into a book, I recommend [13], with its beautiful exposition and unique way of putting the subject in context.

Acknowledgments

I would like to thank Alain Chenciner and Rick Moeckel for useful comments on a draft of this book review. I would also like to thank Gregoire Vion for redrawing Figure 1.

References

[1] M. Born, **Natural Philosophy of Cause and Chance**, Oxford Press, 1949.

[2] L. Capogna, D. Danielli, Donatella, S. Pauls, and J. Tyson, **An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem.** Progress in Mathematics, 259. Birkhäuser Verlag, Basel, 2007. MR2312336 (2009a:53053)

[3] E. Cartan, *Les systèmes de Pfaff à cinque variables et lès équations aux dérivées partielles du second ordre*, Ann. Sci. Ècole Normale, no. 3, **27**, 109-192, (1910). MR1509120

[4] A. Chenciner, *De l'espace des triangles au probleme des* 3 *corps,* web page: http://www.imcce.fr/Equipes/ASD/person/chenciner/chen_preprint.php, under the year 2004. MR1704972 (2000h:70023)

[5] W.L. Chow, *Uber Systeme van Linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann., vol. 117, 98-105, (1939). MR0001880 (1:313d)

[6] M. Gromov, *Carnot-Caratheodory spaces seen from within*, pp. 79-323 in **Sub-Riemannian Geometry**, A. Bellaiche and J-J. Risler, eds., Birhauser, Basel, Switzerland, (1996). MR1421823 (2000f:53034)

[7] M. Gromov, **Groups of Polynomial Growth and Expanding Maps**, Publ. Math. IHES, **53**, 53-73, (1981). MR623534 (83b:53041)

[8] L. Hörmander, *Hypoelliptic second order differential operators*, Acta. Math., **119**, 147-171, (1967). MR0222474 (36:5526)

[9] R. Montgomery, **A tour of sub-Riemannian geometries, their geodesics and applications**, Math. Surveys Monogr., vol. 91, American Mathematical Society, Providence, RI, 2002. MR1867362 (2002m:53045)

[10] G.D. Mostow, **Strong Rigidity of Locally Symmetric Spaces**, Ann. of Math. Stud., vol. 78, Princeton Univ. Press, (1973). MR0385004 (52:5874)

[11] P. K. Rashevskii, *About connecting two points of complete nonholonomic space by admissible curve*, Uch. Zapiski Ped. Inst. Libknexta, no. 2, 83-94, (1938) [Russian].

[12] A. Shapere and F. Wilczek, *Self-propulsion at low Reynolds number*, Physical Review Letters, vol. 58, 2051-2054, (1987).

[13] R. Strichartz, *Realms of mathematics: elliptic, hyperbolic, parabolic, sub-elliptic*, Math. Intelligencer, **9**, no. 3, 56–64 (1987). MR895772 (88h:35002)

[14] N. Tanaka, *On the equivalence problem associated with simple graded Lie algebras*, Hokkaido Math. Journal, v. 8, 23-84, 1979. MR533089 (80h:53034)

[15] L.C. Young, **Lectures on the Calculus of Variations and Optimal Control Theory**, Chelsea, 1980.

Richard Montgomery
UC Santa Cruz
E-mail address: `rmont@math.ucsc.edu`